

State of California Department of Transportation

Transportation System Data Business Plan



RFO# TSI DPA-0003
September 29, 2011



CAMBRIA
SOLUTIONS

Table of Contents

Executive Summary	4
Chapter 1 Caltrans Data Governance	7
1.1 <i>Mission, Goals and Success Measures.....</i>	7
1.2 <i>Roles and Responsibilities.....</i>	9
1.3 <i>Data Governance Charter Overview</i>	13
1.4 <i>Impact of Prior Efforts</i>	13
1.5 <i>Funding and Resources.....</i>	14
Chapter 2 Processes to Identify Business Needs	15
2.1 <i>Process Maps and Narratives Overview</i>	15
2.2 <i>Data Governance Processes.....</i>	16
2.3 <i>Data Governance Forms</i>	24
Chapter 3 Data Products and Data Assessment	26
3.1 <i>Data Products</i>	26
3.2 <i>Transportation System Corporate Data Sets.....</i>	27
3.3 <i>Data Quality Assessment Process and Tools</i>	32
3.4 <i>Assessing Caltrans Corporate Data Sets.....</i>	38
Chapter 4 Data Architecture	49
4.1 <i>Data Architecture</i>	49
4.2 <i>Data Element Catalog.....</i>	55
4.3 <i>Data Set Catalog</i>	58
Chapter 5 Data Governance Program Management and Implementation Plan.....	62
5.1 <i>Data Governance, a Program Management Perspective.....</i>	62
5.2 <i>Data Governance Change Management Plan.....</i>	63
5.3 <i>Data Governance Website Concept Overview.....</i>	65
5.4 <i>Risk Management Strategies and Controls.....</i>	66
5.5 <i>Data Governance Program of Projects.....</i>	69
5.6 <i>Near-term Data Governance Projects</i>	73
5.7 <i>Implementation Schedule</i>	79
5.8 <i>Project Summary Form.....</i>	83
Appendix A Glossary of Terms.....	84
Appendix B Data Governance Board Charter	87
Appendix C Data Issue Form	92

Appendix D	Project Summary Form	93
Appendix E	Proposed Project Prioritization Form	94
Appendix F	Quarterly Project Progress Form	96
Appendix G	Data Governance Policy Form.....	97
Appendix H	Data Element Metadata Form.....	98
Appendix I	Corporate Data Set Metadata Form.....	99
Appendix J	Data Quality Assessment Worksheet	100
Appendix K	Data Element Analysis Worksheet.....	101
Appendix L	Risk Management Worksheet	102
Appendix M	Corporate Data Sets – Complete List.....	103
Appendix N	Vehicle Miles Traveled Data Sources Fact Sheet	109
Appendix O	Project Summary – Vehicle Miles Traveled (VMT) Pilot.....	112
Appendix P	Project Summary – County/Route/Postmile- LRS Pilot	114
Appendix Q	Project Summary – Annual Average Daily Traffic (AADT) Pilot.....	116
Appendix R	Project Summary – Create the Technical Advisory Committee (TAC).....	118
Appendix S	Project Summary – Data Governance Website Mockups	120
Appendix T	Data Governance Change Management/Communication Plan Content	130
Appendix U	Data Governance Process Maps & Narratives	135
Appendix V	Data Quality Assessment Process Map & Narrative	158

Executive Summary

The Transportation System Data Business Plan (Data Business Plan) is necessary for Caltrans to address transportation system data problems that impact the ability of Caltrans business users to efficiently and effectively perform important business duties. The Data Business Plan identifies the key issues facing Caltrans and develops an approach to resolving these issues, along with the organization needed to initiate and manage the plan.

The Data Business Plan is based upon a data governance approach that achieves an orchestration of people and processes to enable an organization to leverage data as an enterprise asset. Data “governance” aims to break down organizational barriers and use data to serve the needs of the business. Because data governance is a strategic initiative involving multiple functions across the organization, a Data Governance Program should include:

- A governing body (steering committee or council);
- An agreed upon common set of procedures; and,
- A plan to communicate and execute the data governance procedures.

The Data Business Plan defines governance roles and responsibilities, develops the data related processes and architecture, and provides a timeline and an approach for implementing data governance.

At the outset of the data governance effort, a Data Governance Steering Committee (Steering Committee) was established to oversee the development of the Data Business Plan. The Steering Committee and data governance effort was led by the Division of Transportation System Information (TSI) and supported through strong representation from divisions and offices throughout Caltrans, including: TSI, Environmental, Mass Transportation, Design, Transportation Planning, Maintenance, Right-of-Way, Aeronautics, Traffic Operations, Rail, and Information Technology.

The Steering Committee established the initial direction for the data governance project and provided a strong foundation for a successful corporate effort to improve transportation system

data. The direction established for data governance is reflected in this Data Business Plan and the chapters summarized below:



Chapter 1 – Caltrans Data Governance: This chapter includes the strategic direction of data governance at Caltrans, roles and responsibilities of the data governance structure, policies, available resources, and other items. This chapter documents the establishment of the Transportation System Data Governance Board (Board) and is organized in the following sections:

- **Mission, Goals and Success Measures** – This section summarizes the main themes of the Caltrans Data Governance Plan and includes the mission statement, goals, governance metrics, and success metrics.
- **Roles, Responsibilities, and Accountabilities** – This section summarizes the roles that will be involved in data governance, as well as, the responsibilities assigned to each role. Roles include: the Board, a Technical Advisory Committee (TAC), Data Custodians, Business Owners, and Communities of Interest.
- **Data Governance Charter Overview** – This section provides an introduction to the elements of the Charter. The

detailed Charter can be found in Appendix B.

- **Impact of Prior Efforts** – This section provides an overview of prior projects and initiatives that relate to transportation system data issues.
- **Funding and Resources** – This section summarizes the impact of limited resources on the initial Caltrans data governance structure.

Chapter 2 – Processes to Identify Business Needs:

This chapter provides the core processes for the successful operation of data governance at Caltrans and the guidance and forms to support performing those processes. This chapter is organized in the following sections:

- **Process Maps and Narratives Overview** – This section provides a high-level introduction of the methodology used in documenting the data governance process maps and narratives.
- **Data Governance Processes** – This section provides diagrams and descriptions of the data governance processes.
- **Data Governance Forms** – This section provides templates and descriptions of the forms that will be used to assist in performing these processes.

Chapter 3 – Data Products and Data Assessment:

This chapter is the first of two chapters describing ‘what’ is being governed by data governance. It presents data products as data sets and corporate data sets from a business perspective. Data assessment criteria are outlined and applied to three corporate data sets in order to identify data issues. These data issues will serve as the foundation for one or more project concepts outlined below:

- **Caltrans Data Products** – This section provides an overview of Caltrans data products and their relationship to the next chapter (*Data Architecture*).
- **Caltrans Corporate Data Sets** – This section defines and presents a list of twelve of the corporate data sets

identified by the Board (see Appendix M for a complete list of corporate data sets).

- **Data Quality Assessment Process and Tools** – This section presents a data assessment process and associated data assessment criteria for evaluating data products and corporate data sets.
- **Assessing Caltrans Corporate Data Sets** – The project team conducted interviews and applied the data quality assessment criteria to three high priority corporate data sets that the Caltrans Data Governance Board selected. This section provides the findings from those interviews, illustrates the types of data quality issues facing Caltrans, and lays the foundation for future data improvement projects.

Chapter 4 – Data Architecture:

This chapter presents the initial elements of data architecture at Caltrans and is the second of two chapters that outline ‘what’ is being governed by data governance at Caltrans. The next chapter (*Data Governance Program Management and Implementation Plan*) will present an action plan for fully implementing the Data Business Plan, including the Data Element Catalog and Data Set Catalog presented in this chapter. This chapter is organized in the following sections:

- **Data Architecture** – This section defines data architecture, discusses the attributes and benefits of metadata, the characteristics of quality metadata, Caltrans’ data architecture, and the relationship between the Data Element Catalog and Data Set Catalog to the overall Caltrans data architecture.
- **Data Element Catalog** – This section presents the metadata associated with data elements and a Data Element Catalog.
- **Data Set Catalog** – This section presents the metadata associated with data sets and a Data Set Catalog.

Chapter 5 – Data Governance Program Management and Implementation Plan: This chapter describes the path forward from a program management perspective and includes near-term and ongoing long-term actions and projects. The actions or projects will include “standing up” the data governance structure, initiating ongoing data governance activities, finalizing detailed standards, populating the Data Element Catalog and Data Set Catalog, and initiating projects to address data issues previously outlined. This chapter includes the following sections:

- **Data Governance, A Program Management Perspective** – This section describes how Caltrans’ data governance requires a program management approach to ensure ongoing data governance and how data related projects support Data Governance Program objectives. This section also includes near-term actions and projects that are required to ‘stand up’ and support the Board and TAC.
- **Data Governance Change Management Plan** – This section provides a series of recommended activities to assist in the adoption of the Data Governance Program, including activities to:
 - Promote Caltrans’ awareness of the Data Governance Program
 - Promote Caltrans’ desire to support and participate in the Data Governance Program
 - Introduce the tools and processes that support the Data Governance Program
 - Provide access to the results and decisions for the data governance activities
 - Reinforce the value of the Data Governance Program
- **Data Governance Website Concept Overview** – This section presents f a concept for a website that will function as a critical tool to:
 - Support ongoing communications about data governance progress
 - Serve as Caltrans’ central portal for all data governance products

- **Risk Management Strategies and Controls** – This section summarizes the likely risks and mitigation strategies, as well as recommendations for a Risk Management Methodology, which includes processes and tools.
- **Data Governance Program of Projects** – This section is devoted to four projects that initiate data governance oversight and communications.
- **Near-term Data Governance Projects** – This section identifies a series of pilot projects to be undertaken by the project team to address known data issues and achieve the goals and objectives of the Data Business Plan in the near-term.
- **Implementation Schedule** – This section provides a three year schedule for implementing the Data Governance Program, including a Gantt chart displaying the timeline for program management activities and projects, and data governance pilot projects. This section also contains a PERT chart that graphically illustrates the dependencies between activities and projects.
- **Project Summary Form** – This section includes an overview of the Project Summary Form to be used for the pilot projects outlined in the “Near-term Data Governance Projects” section and for future projects. The Project Summary Form is included in Appendix D.

Appendices – Glossary, Charter, Forms and Worksheets, and Project Summaries:

The Data Business Plan concludes with appendices designed to support the Board and data governance at Caltrans. The Charter outlines the Board’s roles and responsibilities and the Forms will be used by the Board and TAC to manage data governance projects and to establish data governance standards. Worksheets will generally be used by the TAC and data governance project teams for analysis activities. Several appendices support near term data governance projects and reflect preliminary analysis of three corporate data sets. The Data Business Plan, including the documents in the Appendix, support practical, near term action.

Chapter 1 Caltrans Data Governance

1.1 Mission, Goals and Success Measures

Establishing a strategic direction is essential to the functioning of any organization to define its overall purpose and goals. For data governance at Caltrans, a mission statement and high-level goals provide the strategic direction for the data governance effort.

1.1.1 Caltrans Data Governance Mission

The Caltrans data governance structure was created with the following mission:

The Caltrans Transportation System Data Governance Board (Board) ensures that Caltrans creates and maintains reliable transportation system data that is accessible to Caltrans and its partners.

This mission statement reflects the intent to focus on the business needs of Caltrans first and foremost. The statement emphasizes the following:

- The Board will have the authority to implement change and require accountability of others
- Data governance and the Data Business Plan are focused on transportation system data and the needs of Communities of Interest who rely on these data assets
- There will be accountability for Business Owners to develop and maintain data that can be relied upon for decision-making
- Reliable data will be accessible to both Caltrans and external users

This mission statement provides the Board and the overall data governance effort at Caltrans with a clear purpose and a reliable reference point to base future decisions.

1.1.2 Caltrans Data Governance Goals and Objectives

High-level strategic goals are developed to address the issues confronting an organization. Caltrans' data governance goals are designed to address data issues and challenges existing in the organization, which hinder good business decision-making. These issues are at the heart of the improvements in data management that Caltrans desires.

Table 1: Key Data Issues

#	Data Issue
1.	Difficulty finding consistent data for business needs and partner/customer inquiries
2.	Real and perceived data quality issues
3.	Difficulty integrating data across divisions, offices, and districts
4.	Challenges in identifying data that is valuable and needed to drive decisions
5.	Lack of a standardized approach to address new data needs
6.	Difficulty finding staff who are knowledgeable and responsible for the data
7.	Difficulty accessing needed data

With key data issues in mind, the Steering Committee (a precursor to the Board) established five strategic goals for the data governance effort. These goals guide the Board's data governance activities, the implementation of data governance, and the selection of data improvement projects. The five strategic data governance goals are shown in the table below with their related objectives.

Table 2: Goals and Objectives

#	Goal Title	Goal Description	Objectives
1.	Leadership	Champion data solutions to ensure accountability and increase the value of data assets	<ul style="list-style-type: none"> Promote data governance within Caltrans Communicate data-related changes to all Communities of Interest Monitor progress and ensure accountability of data governance tasks and projects
2.	Quality	Oversee efforts to provide high quality data that is accurate, clear, and easy to access	<ul style="list-style-type: none"> Establish a Data Quality Assurance Program Increase the accuracy and clarity of data Improve accessibility of data
3.	Prioritization	Prioritize efforts to address data gaps and needs	<ul style="list-style-type: none"> Establish clear priorities to address data gaps and needs Communicate priorities to Caltrans business units
4.	Cooperation	Facilitate cross-organizational collaboration, data sharing, and integration (break-down barriers between business units, reduce data silos)	<ul style="list-style-type: none"> Increase opportunities for data sharing Eliminate data silos and other barriers Ensure business units know the identity of Business Owners and Data Custodians Ensure Business Owners know the identity of the Communities of Interest
5.	Flexibility	Encourage creative and innovative solutions to data needs	<ul style="list-style-type: none"> Identify innovative data solutions throughout Caltrans Communicate innovative solutions to Business Owners and Communities of Interest

1.1.3 Caltrans Success Measures

The Board will be responsible for prioritizing the various objectives and developing projects and processes to achieve them. These projects and processes may be assigned to a team made up of various Subject Matter Experts (SMEs) or to individuals with the capability to complete the project or process. Projects will have timelines and expectations for decisions, recommendations, and/or deliverables designed to facilitate project completion and further the achievement of the objective. Processes will support the implementation of data governance as well as guide long-term consistency in how data is created and maintained within Caltrans.

To determine whether or not progress is being made toward achieving the objectives, the Board will establish performance measures related to each objective. These performance measures will ensure the accountability of a team or individual to produce the results assigned by the Board. The following table contains a list of objectives and sample performance measures:

Table 3: Sample List of Objectives and Performance Measures

#	Objective	Performance Measure
1.	Ensure business units know the identity of Business Owners and Data Custodians	Increase # and/or % of data elements with identified owners/custodians that are available to business units
2.	Increase the accuracy and clarity of the data	Reduction in # and/or % of inaccurate data elements and data elements without clear definitions
3.	Communicate data-related changes to all Communities of Interest	Development and implementation of a data communication plan approved by the Board

#	Objective	Performance Measure
4.	Eliminate data silos and other barriers	# of data silos identified, and # of solutions implemented to address data silos

The list above contains samples only for the purpose of demonstrating potential performance measures. The Board will develop performance measures to reflect its decisions regarding the prioritization of data governance projects and tasks.

The mission and goals of data governance at Caltrans discussed above are relevant to a variety of organization types; from small to large, from simple to complex, staffed by a permanent staff or staff loaned from other parts of Caltrans. The amount of resources devoted to data governance is a primary determinant of the types of roles and responsibilities that are established for data governance. The Caltrans data governance structure discussed in the next section (*1.2 - Roles, Responsibilities, and Accountabilities*) is a reflection of resource limitations and other factors influencing the organization.

1.2 Roles and Responsibilities

There are a variety of roles and responsibilities that may be used in developing a data governance structure. Based upon the specific structure selected, different organizations choose roles and responsibilities that best meet their needs for the chosen structure. In general, data governance requires the following:

- A governing body with the authority to set policy and make decisions regarding the management of data
- Technical staff dedicated to the data governance effort (if resources are available for dedicated staff)
- SMEs who represent the data interests and needs of business units within the organization
- Management of the business units that own specific data sets
- Communities of Interest who use data or have an interest in the data assets of others

1.2.1 Common Data Governance Roles and Responsibilities

There are common data governance roles that have been developed and used for data governance by those who have implemented a Data Governance Program in their organization. The table below presents some typical roles and the definition of the responsibility of each role:

Table 4: Typical Data Governance Roles and Responsibilities

#	Role	Responsibilities
1.	Data Governance Board (Board)	A decision-making body with primary responsibility for developing and implementing policy related to data assets.
2.	Technical Advisory Committee (TAC)	An advisory body of technical data experts who review detailed information and provide advice and recommendations to the decision-making body.
3.	Data Steward	The owner of the Data Business Plan and the person responsible for initiating changes to established data stewardship and management processes.
4.	Data Custodian	The SME and point of contact for their assigned data products. Represents the Business Owner and maintains communication with the Community of Interest.

#	Role	Responsibilities
5.	Business Owner	The manager responsible for ensuring that the data product is being well managed and is providing value commensurate with the level of required investment for ongoing upkeep.
6.	Business Data Architect	The person responsible for ensuring data consistency, integration, and efficient access to support business functions. The Business Data Architect creates and maintains formal descriptions of data entities and their relationship to business processes.
7.	Data Coordinator	The person responsible for the master catalogs of data products, and makes sure all of the elements are complete and up to date. Monitors development of new data products and ensures data services are being carried out.
8.	Communities of Interest	An individual or group (internal or external to the organization) that could be affected by the data that is in-scope for the Data Governance Program. They include groups who have an interest in transportation system data in general, create data, those who use data, and those who set rules and requirements for data. This group can be a data user, owner, recipient or other person/group that has an interest in data.

Not all of the roles listed above appear in every organization. Many of these roles are present only in mature, well-resourced Data Governance Programs. Often, data governance is established with a minimum structure and grows only as necessary to accommodate increased responsibilities with increased support. The specific preliminary Caltrans data governance roles and responsibilities are described in *Section 1.2.3 – Caltrans Initial Data Governance Structure* and can be seen in *Figure 1: Caltrans Initial Data Governance Structure*, with a description of each role afterwards. The following section describes the factors that were considered in developing the initial Caltrans data governance structure.

1.2.2 Caltrans Data Governance Organization – Factors Considered

The type of data governance structure that an organization develops is influenced by various factors. These include:

- **Authority** – The authority level of a governance body will help determine if it will be a decision-making body or if it will be advisory in nature.
- **Resource Availability** – An organization is limited to the amount of resources (staff and funds) that are allocated to the organization.
- **Mission and Goals** – The strategic direction of a governance body will help guide the development of that governance body.
- **Size of organization and representation** – The size of the data governance organization is a careful balance of necessary Data Custodians, Business Owners, and Communities of Interest, while ensuring that the organization is appropriately sized for flexibility and decision-making.
- **Scope** – The scope of the data governance tasks and projects selected will indicate the type and size of organization that will be needed.
- **Initial data governance structure** – The roles selected are determined by the factors listed above and determine the final content of the data governance structure.

A Steering Committee was formed to direct the development of the plan and establish the foundation of the data governance effort. After considering the existing data issues, mission, goals and objectives of data governance at Caltrans, and the factors above, the Steering committee made the following decisions regarding the data governance organization:

Authority: The Steering Committee considered the issues, goals and objectives of data governance, and recommended creation of a governance board that:

- Has the authority to make needed decisions on pressing data issues and rally resources
- Is not an organizational body such as a Division, Office, or Branch
- Will consist of members who have authority to rally resources and direct staff to participate in collaborative efforts across Divisions, Districts, and Offices (ideally Division Chiefs and appropriate district representatives)

Resource Availability: The Steering Committee recommended that resources (staff and funds) be determined when pilot projects are selected, and that Divisions and Districts will collaborate to address resource requirements as needed. No permanent budget or staffing allocation was recommended at this time.

Mission and Goals: The Steering Committee developed the mission and goals for data governance when considering the organization (mission and goals are described in the previous section, *1.1 - Mission, Goals, and Success Measures*).

Size and Representation: The Steering Committee recognized that many Divisions and all Districts are both creators and customers of transportation system data and represent key Communities of Interest in this data governance process. They also recognized that a smaller governing body is most practical for decision-making. The Steering Committee recommended that:

- Districts be represented on the Board
- Not all Districts would be represented at all times, utilizing a potential rotating membership for broader representation
- Key Divisions will be represented on the Board with possible rotating membership
- The Board will be kept to a practical size

The Steering Committee elected to act as the Interim Board for until membership is finalized and processes are initiated.

Scope: The Steering Committee recommended a “go slow” approach to the initial scope of the data governance effort and project selection. This reflects a concern and two recommendations:

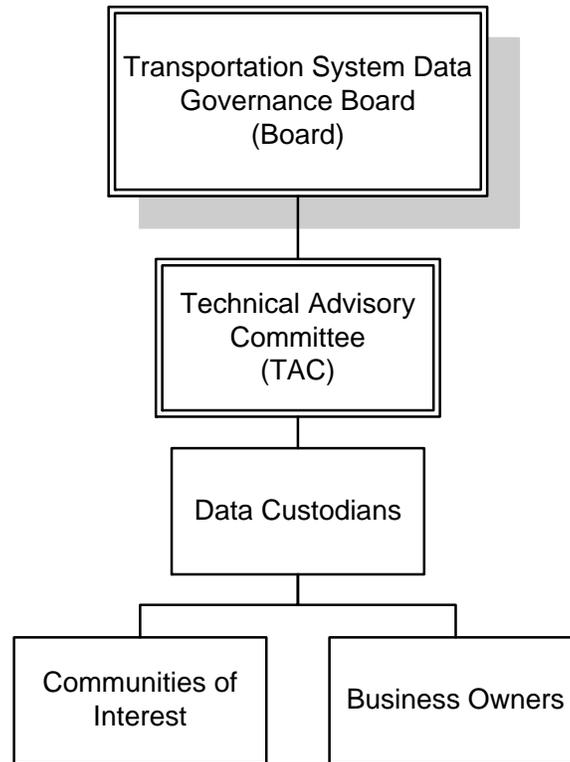
- Resources would be limited, particularly in the short-term
- Data governance at Caltrans should evolve over time and begin operations with a limited agenda that is attainable and provides early successes upon which to build
- A limited number of pilot projects should be selected to initiate the data governance process

The result is the desire for a Caltrans data governance structure with decision-making authority able to solve problems; but, with limited scope, and thus limited size and participation (at least initially). These decisions regarding level of authority and scope, along with limited resources, and the Mission and Goals, provide an overall context for the development of a data governance structure that best meets the needs of Caltrans.

1.2.3 Caltrans Initial Data Governance Structure

Based on the above considerations and a discussion of the potential roles and responsibilities of typical data governance efforts, the Steering Committee recommended an initial data governance structure shown below. Although the future may require expanding data governance capabilities, given the limits and influences described above, the figure below represents the initial data governance structure at Caltrans for the first year of activity (descriptions of this structure immediately follow).

Figure 1: Caltrans Initial Data Governance Structure



This structure satisfies the initial requirements of the Caltrans data governance structure in the following ways:

- The Transportation System Data Governance Board (Board) is made up of Caltrans Division-level and District-level leadership and has the authority to decide and implement most of the decisions necessary for effective data governance. (Note: Some decisions may have to be referred to the Directorate – See Appendix A for a definition of “Directorate”)
- The Technical Advisory Committee (TAC) consists of existing Caltrans staff with particular subject matter expertise in existing data and/or data requirements at Caltrans. Some members of this committee may also serve as Data Custodians (see below). This group assists the Board by reviewing technical subject matter and providing information and recommendations to the Board to facilitate their decision-making and monitoring roles. The TAC will convene intermittently so as not to overly burden the workload of the committee members.
- Data Custodians are existing Caltrans staff and are the recognized data experts for their functional areas (i.e. Division, Branch, Office, etc.). They serve on an as-needed basis on data governance project teams so as not to overly burden their ongoing workload.
- Business Owners are the managers of the business units that own data. They appoint Data Custodians, as needed, to address projects identified by the Board.
- Communities of Interest are groups of internal and external stakeholders of data that have an interest in data governance actions that impact data in which they have an interest.

There are many different Business Owners, Data Custodians, and related Communities of Interest at Caltrans, although presently they are not always explicitly identified. These groups may be identified on a project-by-project basis (as project teams are formed or individuals are selected to address specific issues).

Upward Accountability: The Board is responsible for raising highly sensitive issues with a department-wide impact, and issues that impact other agencies, to the Directorate for their final decisions on the Board's recommendations. In addition, those issues that are of concern to the SHOPP Executive Committee and/or the Financial Policy Board (see Appendix A for definitions) will be presented to these bodies for their review and input. Also, the Board will raise issues that impact Caltrans' information technology to the IT Governance Board.

In summary, Caltrans has made implementing corporate data governance a high priority. To begin operations as soon as possible, an interim Board was established from members of the Steering Committee to begin initial operations and decision-making.

The Board's operations are defined by a charter, which establishes the membership, rules of operation, decision authority, and other concerns. The next section (*1.3 - Data Governance Charter Overview*) provides an overview of the charter for the Board.

1.3 Data Governance Charter Overview

A Data Governance Charter has been established to help govern and guide the operations and decisions of the Board. The Charter reaffirms the strategic direction of the Board and Caltrans' Data Governance Program and provides the framework for the operations of the Board. The contents of the Charter include:

- The Board's Purpose, Mission, and Goals
- Performance Objectives
- Relationship to Caltrans Management
- Membership
- Meeting Protocols
- Processes
- Definitions

The latest version of the Charter is included as Appendix B.

1.4 Impact of Prior Efforts

Although prior initiatives at Caltrans have not created a formal data governance structure, there have been related efforts that have created information and tools that are beneficial to the Board. Even though these previous projects have been limited in scope, they have provided valuable information and tools that the Board will be able to build upon while improving data governance at Caltrans. Some of these prior efforts include:

- GIS Corporate Structure Value Analysis Study (ongoing)
- Geospatial Data Management Committee (GDMC - ongoing)
- GIS Management Committee (disbanded 2005)
- Enterprise Architecture Council (EAC - ongoing)
- Data Inventory and Cataloging Project (completed 2010)

These different efforts have provided valuable information related to existing data issues and problems, the identity of Business Owners and Communities of Interest, where data exists, potential Data Custodians, how to operate a governance program, and existing ideas and innovations for addressing transportation system data issues.

1.5 Funding and Resources

The availability of resources is a primary constraint on the operation of any organization. The Board is no different. As described previously, the extent of the initial data governance structure and the scope of the Board's first projects and tasks will be limited by the resources (staff and funds) that Caltrans can dedicate to data governance.

As good stewards of California's resources, Caltrans stretches existing resources as far as possible. Caltrans does not intend to dedicate new resources or redirect significant existing resources to the data governance structure at this time. Business units that see the benefit of data governance will be asked to contribute staff and funds, as needed, to the data governance effort to keep it operational, at least in the initial stages of data governance.

These resource limits are one of the primary contributing factors to the decision to limit the size of the data governance structure at Caltrans and to limit the scope of projects and tasks that the Board will initiate in the early stages of data governance. In future years, as the value of data governance becomes more widely recognized, dedicated funding may be made available by the Caltrans Directorate and/or participating divisions and districts. The opportunity to establish full-time data governance staff to manage a program with a broader scope may become feasible with sufficient resources. But ultimately, data governance will have to compete with other worthy efforts throughout Caltrans in a constrained resource environment that will likely exist for the foreseeable future. In such a circumstance, the ability of the Board to effectively prioritize the projects it undertakes will be of great importance to make sure Caltrans gets the most data governance benefit for its limited data governance resources.

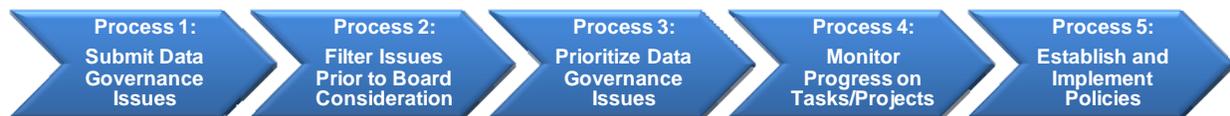
Chapter 2 Processes to Identify Business Needs

Data governance must succeed at identifying and resolving transportation system data issues, resulting in the improvement of Caltrans' business operations. The success of data governance is dependent, in part, on the implementation of standard processes to guide aspects of data governance in a clear and consistent manner. As such, data governance processes must be formalized, clearly described and provided in an easy-to-understand format so that all who participate in data governance understand what the processes are, how they are conducted, and what the desired result will be. By formalizing processes, the Board and all participants in data governance can focus on *resolving issues* versus determining how the governance process works.

The Caltrans core processes developed to support data governance are:

- Submitting data governance issues for consideration
- Filtering issues prior to Board consideration
- Prioritizing data governance issues for action
- Monitoring progress on existing tasks and projects
- Establishing and implementing new data governance policies

Figure 2: Overview of Data Governance Processes



These five processes represent the most important data governance activities to be undertaken during the implementation of data governance at Caltrans. The next section (2.1 – *Process Maps and Narratives Overview*) describes the process that was used in documenting the five data governance processes. Section 2.2 (*Data Governance Processes*) includes a representation of each of the five core data governance processes which are portrayed graphically in a more detailed “process map” (flow chart). See Appendix U for larger “process maps” and detailed “process narratives” for each of these five core data governance processes.

2.1 Process Maps and Narratives Overview

Process Maps (or flow charts) provide a graphical representation of a process and are organized so as to present each role/actor in the process in its own “swim lane.” Each swim lane contains the process steps (activities) belonging to the role/actor (the process steps are numbered consecutively). The numbering of the process steps corresponds to the numbering of the activities within each process narrative.

Process Narratives include an introductory section providing a brief description of the process, any assumptions relevant to the process, the “trigger” event that initiates the process, and the “completion indicator” signaling that the process has concluded.

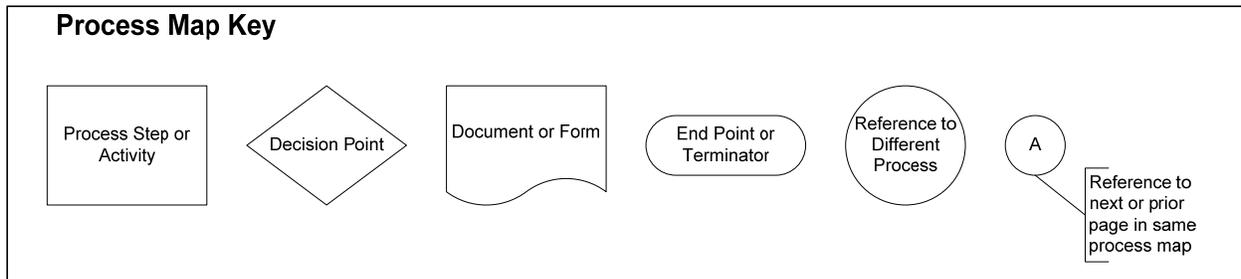
Each process narrative also includes a more detailed description of each process step (activity) in the process, including the role that performs the activity, what inputs (if any) contribute to the activity, and what outputs (if any) the activity produces. These activity descriptions also contain decision points and points at which two different processes may be connected by activities within individual processes.

Forms - Four forms have been developed to support these processes. They are displayed in the appropriate processes to indicate when the particular form is to be used and who is responsible for

completing and/or reviewing the form. The forms are explained in the section after the next (2.3 – Data Governance Forms).

The process maps consist of symbols that depict process steps/activities, decision points, documents/forms, adjoining processes, and end points. Process maps are able to show sequential and parallel activities, key players in the process, and the relationships between them. For each of the flow charts below please refer to the following key for an explanation of the symbols in the maps.

Figure 3: Process Map Key



2.2 Data Governance Processes

Process 1: Submitting Data Governance Issues for Consideration

Data issues are identified by a Data Custodian, Business Owner, or Community of Interest and gathered by the TAC to determine the proper course of action to resolve the data issue. Data issues may be identified throughout the year and submitted at any time to the TAC.

The following data governance roles contribute to this process:

- TAC
- Data Custodians
- Business Owners
- Communities of Interest

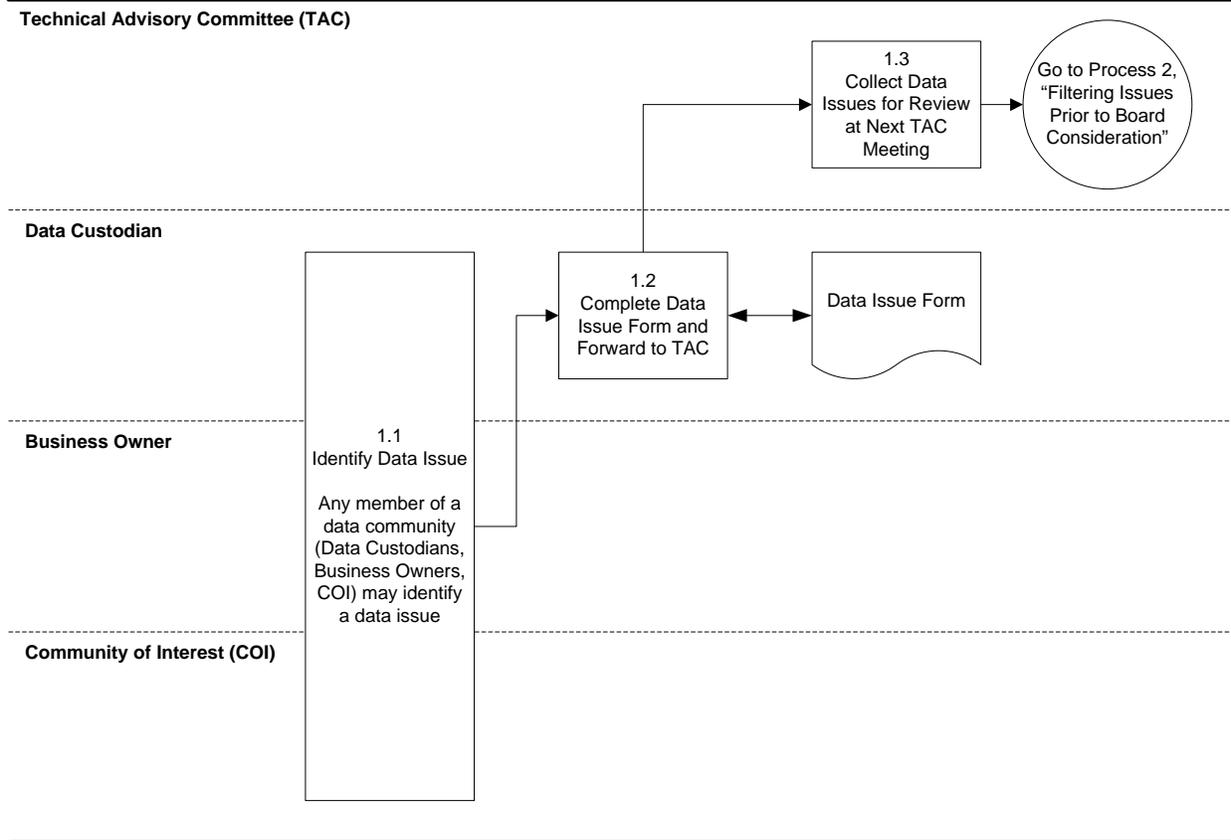
Process 1 Activity Steps

The following steps are reflected in the process map below:

1. Identify Data Issue
2. Complete Data Issue Form and Forward to TAC
3. Collect Data Issues for Review at Next TAC Meeting
(Go to Process 2, "Filtering Issues Prior to Board Consideration")

Why this process is important: The TAC can work with directly with Caltrans staff to ensure that all relevant information is collected regarding identified data issues, and hold the 'repository' of issues until the Board is ready to consider them.

Figure 4: Process 1 - Submitting Data Governance Issues for Consideration



Process 2: Filtering Issues Prior to Board Consideration

For all data issues identified in Process 1, the TAC will determine whether or not the issue should be referred to the Board, if it should or could be resolved by the TAC, or if the issue is out of scope of the data governance process. The TAC will have the delegated authority to resolve lower level data issues without bringing them forward to the Board for consideration.

The following data governance roles contribute to this process:

- TAC
- Data Custodians
- Business Owners

Process 2 Activity Steps

The following steps are reflected in the process map below:

(From Process 1, "Submitting Data Governance Issues for Consideration")

1. Review the Data Issue at Next TAC Meeting
2. Decision: Refer Issue to Board?
3. Forward Data Issue Form to Board

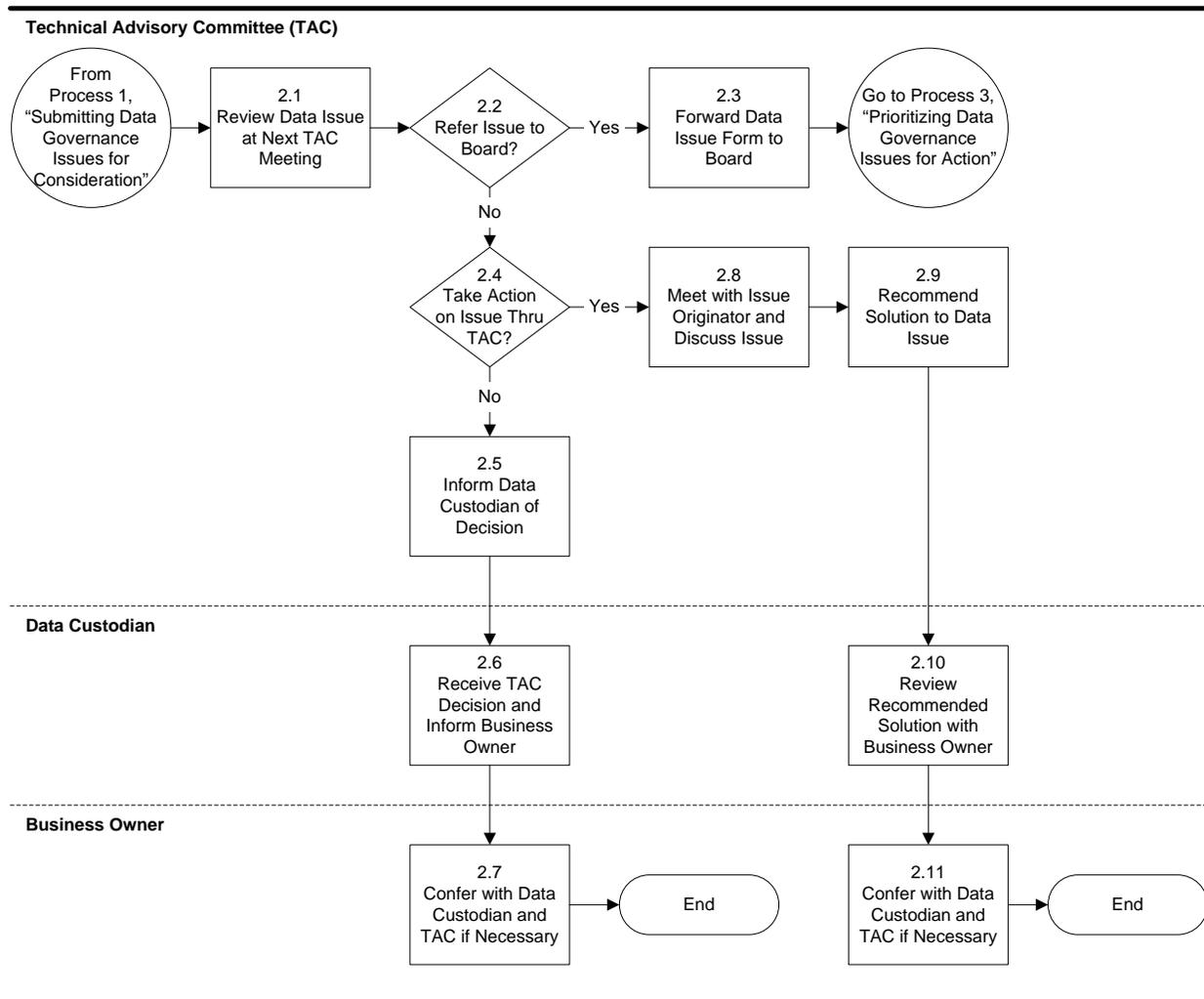
(Go to Process 3, "Prioritizing Data Governance Issues for Action")

4. Decision: Take Action on Issue thru TAC?

5. Inform Data Custodian of Decision
6. Receive TAC Decision and Inform Business Owner
7. Confer with Data Custodian and TAC if necessary
(End of Process)
8. Meet with Issue Originator and Discuss Issue
9. Recommend Solution to Data Issue
10. Review Recommended Solution and Inform Business Owner
11. Confer with Data Custodian and TAC, if necessary
(End of Process)

Why this process is important: The TAC can resolve some data issues or determine if a data issue has already been resolved (e.g. by a policy). Some issues may never need to be raised to the Board to receive data governance guidance.

Figure 5: Process 2 - Filtering Issues Prior to Board Consideration



Process 3: Prioritizing Data Governance Issues for Action

In preparation for a Board meeting, the TAC will conduct a preliminary prioritization of data issues to be presented to the Board. Prioritization can occur at each Board meeting, on an 'as needed' basis, or even one time per year, leaving the remaining quarterly Board meetings to focus on approved efforts rather than identifying new projects.

The Board will determine the final issue prioritization as well as determine which issues will become data governance projects, and direct the TAC to begin establishing project teams. The proposed project teams will be reviewed by the Board Chairperson (and any impacted Business Owners) before the project team membership will be confirmed and they may kickoff the project.

The following data governance roles contribute to this process:

- Board
- Board Chairperson
- TAC
- Business Owners

Process 3 Activity Steps

The following steps are reflected in the process map below:

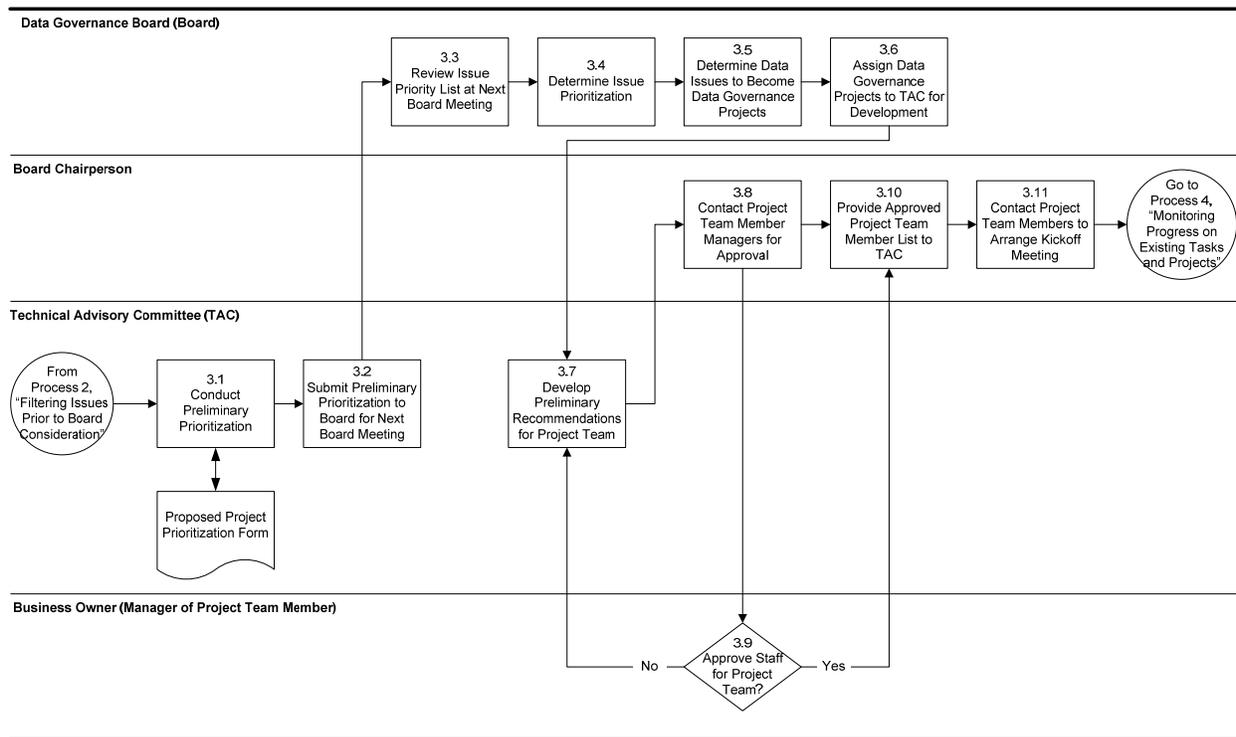
(From Process 2, "Filtering Issues Prior to Board Consideration")

1. Conduct Preliminary Prioritization
2. Submit Preliminary Prioritization to Board for Next Board meeting
3. Review Issue Priority List at Next Board Meeting
4. Determine Issue Prioritization
5. Determine Data Issues to Become Data Governance Projects
6. Assign Data Governance Projects to TAC for Development
7. Develop Preliminary Recommendations for project team
8. Contact Project Team Member Managers for Approval
9. Decision: Approve Staff for Project Team?
10. Provide Approved Project Team Member List to TAC
11. Contact Project Team Members to Arrange Kickoff Meeting

Go to Process 4, "Monitoring Progress on Existing Tasks and Projects"

Why this process is important: In the initial years of the Data Governance Program, the Board must be selective on the number of issues and projects it undertakes. As the program matures, more efforts may be addressed.

Figure 6: Process 3 - Prioritizing Data Governance Issues for Action



Process 4: Monitoring Progress on Existing Tasks and Projects

The Board may recommend that a project team be formed to resolve a difficult data issue. While actively pursuing a recommendation or solution for a data issue, the project team must periodically report progress to the data governance organization. The project team will complete their analysis and provide it to the Data Custodian, Business Owners, and Communities of Interest for review and feedback. The project team will update their recommendation and submit it to the TAC. If accepted by the TAC, the recommendation will be forwarded to the Board for consideration.

The results of the project may be to:

- Establish a new policy, procedure or guideline (see Process 5).
- Establish a new data standard or perform a data quality assessment (these will be described in detail in Chapter 3).

The following data governance roles contribute to this process:

- Board
- TAC
- Project Team
- Data Custodians
- Business Owners
- Communities of Interest

Process 4 Activity Steps

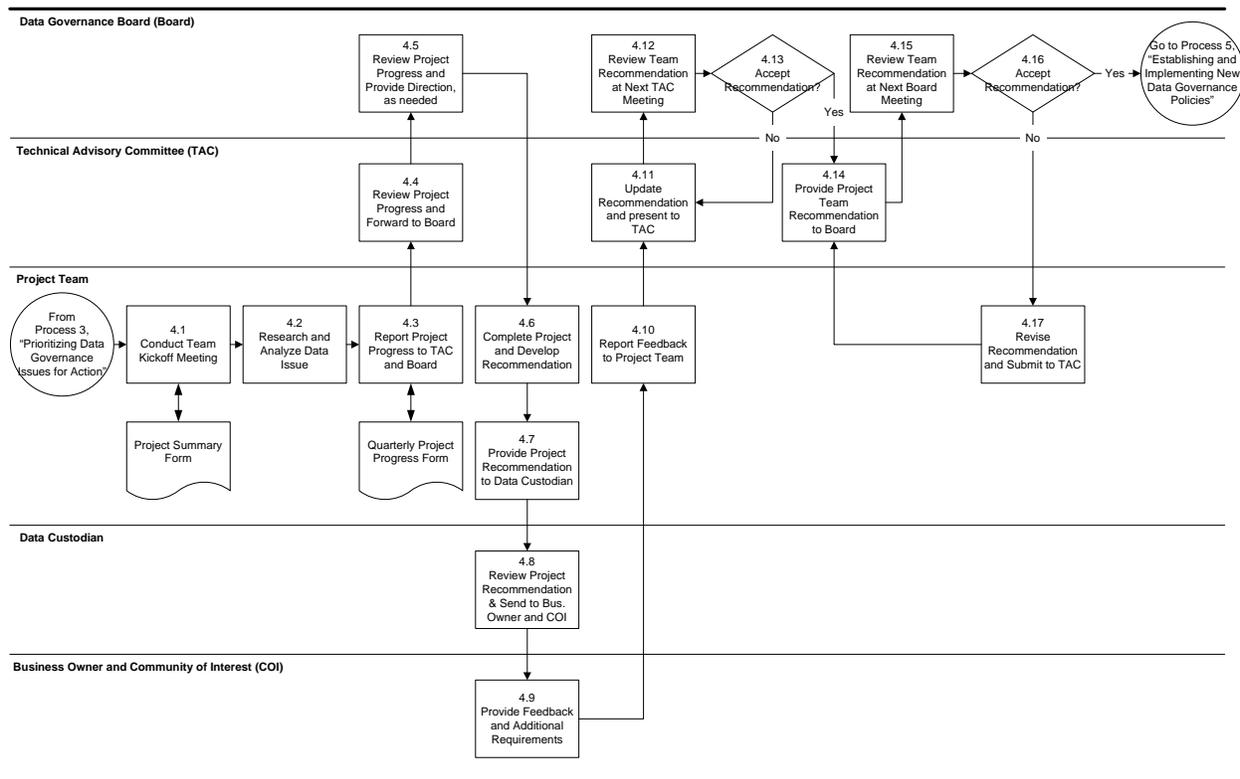
The following steps are reflected in the process map below:

(From Process 3, "Prioritizing Data Governance Issues for Action")

1. Conduct Team Kickoff Meeting
2. Research and Analyze Data Issue
3. Report Project Progress to TAC and Board
4. Review Project Progress and Forward to Board
5. Review Project Progress and Provide Direction, as needed
6. Complete Project and Develop Recommendation
7. Provide Project Recommendation to Data Custodian
8. Review Project Recommendation and Send to Business Owner and Communities of Interest
9. Provide Feedback and Additional Requirements
10. Report Feedback to Project Team
11. Update Recommendation and Present to TAC
12. Review Team Recommendation at Next TAC Meeting
13. Decision: Accept Recommendation?
14. Provide Project Team Recommendation to Board
15. Review Team Recommendation at Next Board Meeting
16. Decision: Accept recommendation?
17. Revise Recommendation and Submit to TAC

Why this process is important: Some of the Board's data governance decisions will result in long-term efforts. The Board will need to monitor ongoing progress to ensure results are aligned with the data governance goals and that expected outcomes are achieved.

Figure 7: Process 4 - Monitoring Progress on Existing Tasks and Projects



Process 5: Establishing and Implementing New Data Governance Policies

Following the establishment of a new data governance policy, procedure or guideline, there will be implementation and monitoring of the policy, procedure, or guideline, including the development of detailed instructions and assistance with implementation. Low level compliance issues will be resolved by Data Custodians and the TAC. High level compliance issues will be resolved by the Board.

The following data governance roles contribute to this process:

- Board
- TAC
- Data Custodians
- Business Owners
- Communities of Interest

Process 5 Activity Steps

The following steps are reflected in the process map below:

(From Process 4, "Monitoring Progress on Existing Tasks and Projects")

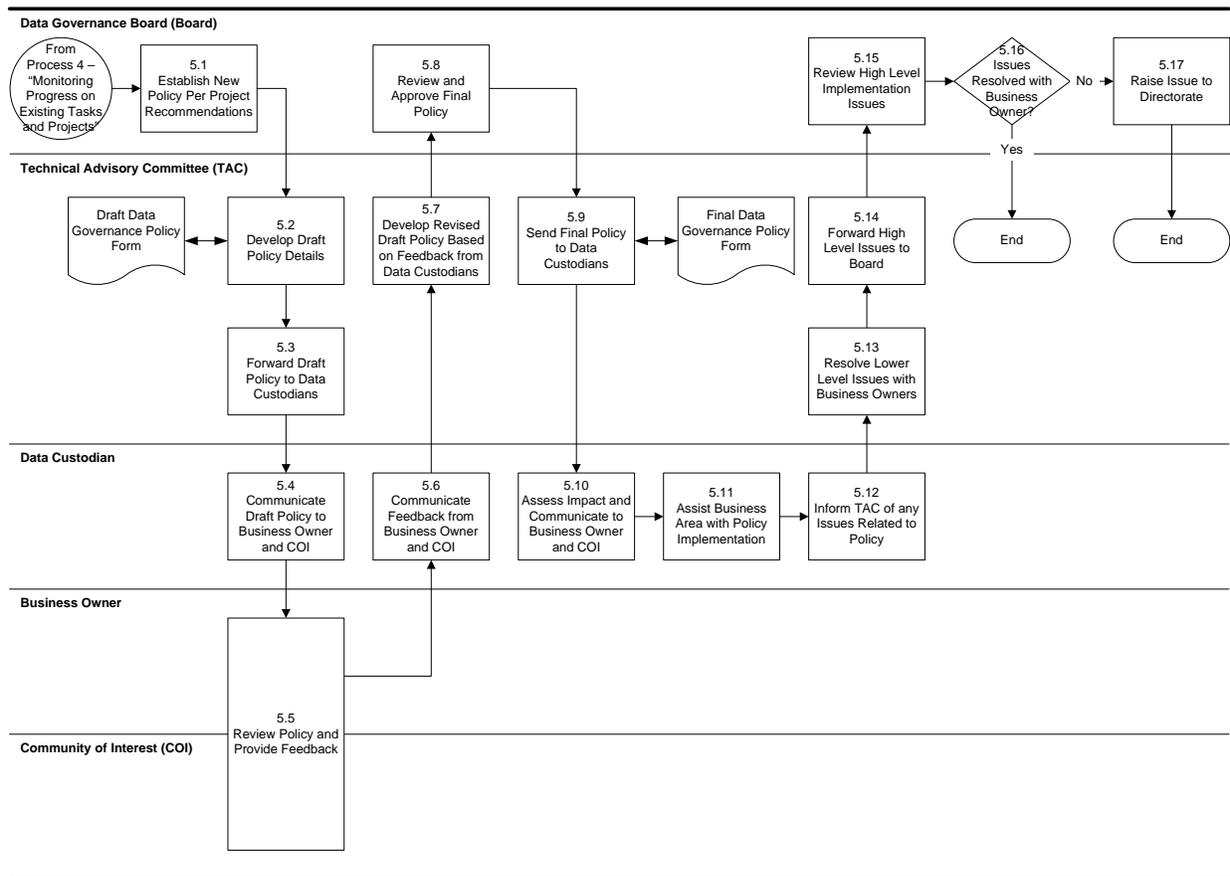
1. Establish New Policy per Project Recommendations
2. Develop Draft Policy Details
3. Forward Draft Policy to Data Custodians
4. Communicate Draft Policy to Business Owner and Communities of Interest
5. Review Policy and Provide Feedback

6. Communicate Feedback from Business Owner and Communities of Interest
7. Develop Revised Draft Policy Based on Feedback from Data Custodians
8. Review and Approve Final Policy
9. Send Final Policy to Data Custodians
10. Assess Impact and Communicate to Business Owner and Communities of Interest
11. Assist Business Area with Policy Implementation
12. Inform TAC of any Issues Related to Policy
13. Resolve Lower Level issues with Business Owners
14. Forward High Level Issues to Board
15. Review High Level Implementation Issues
16. Decision: Issues resolved with Business Owners?
17. Raise Issue to Directorate

(End of process)

Why this process is important: While the data governance organization is not an enforcement group, it will be important to ensure that data policies and data standards are followed, and to ensure that data governance objectives are met.

Figure 8: Process 5 - Establishing and Implementing New Data Governance Policies



2.3 Data Governance Forms

The processes described in the section above (2.2 – *Data Governance Processes*) include the use of various forms to assist in completing the processes. These forms are:

- Data Issue Form (Appendix C)
- Project Summary Form (Appendix D)
- Proposed Project Prioritization Form (Appendix E)
- Quarterly Project Progress Form (Appendix F)
- Data Governance Policy Form (Appendix G)

These five forms are described below, and each of the forms is included as Appendices C through G.

2.3.1 Data Issue Form

This form is used by those submitting data issues to the TAC. Data issues can come from any part of the data community (i.e. Data Custodians, Business Owners, and Communities of Interest). The Data Issue Form contains project overview information, a listing of impacted data and systems, a project description, and a description of the business impact for Caltrans. The Data Issue Form (See Appendix C) is utilized in the following processes:

- Process 1 – Submitting data governance issues for consideration
- Process 6 – Data quality assessment

2.3.2 Project Summary Form

This form is used by Project Teams to provide a description of a data governance project that has been initiated by the Board. The Project Summary Form contains a project description, risks, benefits, relationship to other projects, roles, schedule, and estimated cost. The Project Summary Form (See Appendix D) is utilized in the following process:

- Process 4 – Monitoring progress on existing tasks and projects

2.3.3 Proposed Project Prioritization Form

This form is used by the TAC when developing their preliminary recommendations regarding issue/project prioritization to the Board for review and modification. The Proposed Project Prioritization Form includes a first sheet with a complete listing of proposed data issues/projects, including the project title and the proposing business unit. Additional sheets will contain a summary overview of each issue/project listed on the first sheet to provide reference information for the Board while discussing the prioritization of issues/projects. The Proposed Project Prioritization Form (See Appendix E) is utilized in the following process:

- Process 3 – Prioritizing data governance issues for action

2.3.4 Quarterly Project Progress Form

This form is used when a project team is reporting project status to the TAC and the Board. This form will only be used by project teams established by the Board and TAC and will be submitted on a quarterly basis to be reviewed by the TAC and the Board at the quarterly Board meetings. This form contains project overview information, progress status, accomplishments, upcoming tasks and milestones, and any risks and mitigation strategies. The Quarterly Project Progress Form (See Appendix F) is utilized in the following process:

- Process 4 – Monitoring progress on existing tasks and projects

2.3.5 Data Governance Policy Form

This form is to be used when documenting formal data governance policies following the adoption of a new policy by the Board. This will normally occur following the completion of a data governance project that contained recommendations for a new policy that was approved by the Board. This form contains sections for a description of the new policy, background information (including related or superseded policies), parties responsible for implementing the policy and their roles and responsibilities, and the scope of applicability within Caltrans. The Data Governance Policy Form (See Appendix G) is utilized in the following process:

- Process 5 – Establishing and implementing new data governance policies

All data governance forms and tools will be available for downloading on the Caltrans data governance website at www.dot.ca.gov/hq/tsip after the website is developed and available.

Chapter 3 Data Products and Data Assessment

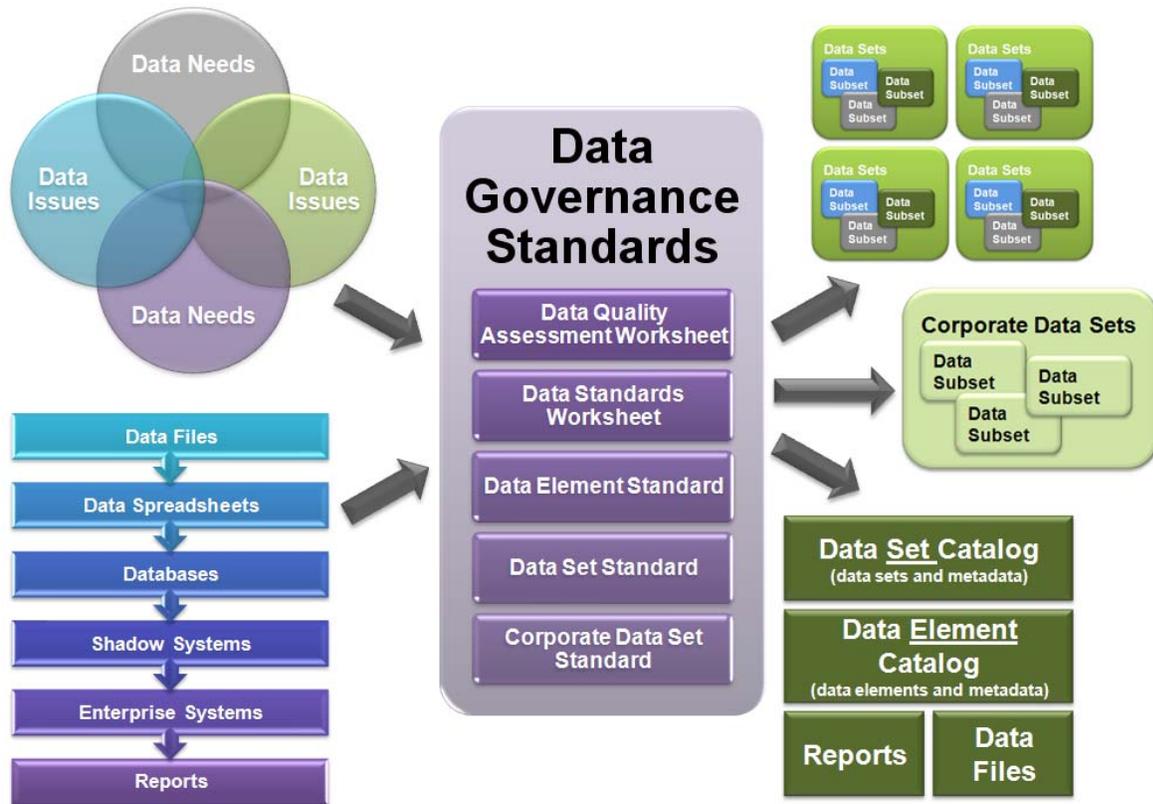
3.1 Data Products

Within the context of Caltrans data governance, *data products are defined from a business perspective; not from an information technology or technical perspective.* Data products are simply collections of data that Communities of Interest require in order to manage the transportation system. Data products may exist as raw data in a database table, data in an Excel Workbook, data in a corporate computer system, or data contained in a report. The 'data product' term connotes a product of value to customers internal or external to Caltrans.

This chapter focuses on a particular type of data product called "data sets," and high priority data sets, known as "corporate data sets." The Data Architecture chapter will define additional data products, including a data element catalog, a data set catalog, and metadata.

As displayed in the figure below, data products are part of a larger universe of data needs, issues, sources, standards, data sets, and data catalogs. Data governance requires that Business Owners consider data needs and data quality issues in terms of available data (available from a variety of files, spreadsheets and systems) and data governance standards. The data governance standards lead to recognized data sets (a collection of data and accompanying documentation which relate to a specific theme), higher priority corporate data sets, catalogs of data elements and data sets, and data files and reports.

Figure 9: Data Needs, Sources, Standards, and Products



In addition to the data products contained in this chapter, numerous items have been developed to support data governance. These data standards and data products fall into three broad categories.

- Data standards and forms that support data governance (from Chapter 2 – “Processes to Identify Business Needs”):
 - Data Issue Form (Appendix C)
 - Project Summary Form (Appendix D)
 - Proposed Project Prioritization Form (Appendix E)
 - Quarterly Project Progress Form (Appendix F)
 - Data Governance Policy Form (Appendix G)

- Data standards and forms to ensure Data Quality and Standardization (from Chapter 2 – “Processes to Identify Business Needs”, Chapter 3 (this chapter) – “Data Products and Data Assessment,” and Chapter 4 – “Data Architecture”):
 - Data Quality Assessment Worksheet (Appendix J)
 - Data Element Analysis Worksheet (Appendix K)
 - Data Element Metadata Form (Appendix H)
 - Corporate Data Set Metadata Form (Appendix I)

- Data standards and products for Communities of Interest (from Chapter 3 (this chapter) – “Data Products and Data Assessment,” and Chapter 4 – “Data Architecture”):
 - Data Element Catalog
 - Data Set Catalog
 - Metadata (e.g. data definitions, data specifications)
 - Corporate data sets
 - Databases and/or data files
 - Reports

The remaining sections of this chapter further define data sets and corporate data sets, and demonstrate how to apply data quality standards to these data sets.

3.2 Transportation System Corporate Data Sets

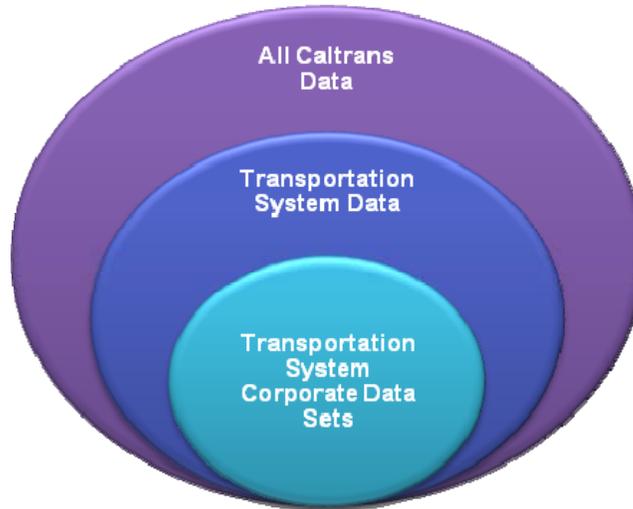
Transportation system data does not encompass all Caltrans data, but rather data directly related to the transportation system. The following is the definition of transportation system data from the Board Charter:

Transportation system data is data related to the definition, condition, or performance of the physical transportation system, including all transportation modes. It does NOT include administrative data, such as financial system data, human resources data, procurement data, and data related to the operations of administrative functions.

Transportation system data can broadly impact multiple business functions within Caltrans, or can exist within a “silo” environment that narrowly impacts a single or relatively few business functions. Transportation system data sets relevant to multiple functions across Caltrans are generally the highest priority and are defined as corporate data sets. The figure below shows the relationship between

Caltrans “Transportation System Corporate Data Sets,” “Transportation System Data,” and “All Caltrans Data”:

Figure 10: Caltrans Data Relationships



3.2.1 All Caltrans Data

This data includes a collection of all data managed by Caltrans. It includes transportation system data (see below), as well as administrative, non-transportation system data, such as financial data, human resources data, and other administrative data.

3.2.2 Transportation System Data

Transportation system data is data related to the transportation system managed by Caltrans, including the definition, condition, or performance of all modes of the physical transportation system. This data is distinguished from all of Caltrans data by what it does not include; particularly financial data, human resources data, and other data used in administrative functions.

3.2.3 Transportation System Corporate Data Sets

A transportation system corporate data set is an important subset of transportation system data. Corporate data sets span multiple business functions within Caltrans and therefore have an increased importance across the corporate structure of Caltrans. Corporate data sets are typically of great value to both internal Caltrans users and external users (e.g. the Legislature, the Governor’s Office, local agencies, etc.). Corporate data sets are priorities for data governance because of the broad and visible impact improvements in corporate data sets have across Caltrans. This span of impact is not the only defining aspect of corporate data sets. The list of criteria for identifying corporate data sets at Caltrans is shown in the table below.

Table 5: Corporate Data Set Criteria

#	Criteria	Description
1.	Impacts multiple divisions or programs	Corporate data sets have a broad impact across Caltrans, spanning divisions and/or programs.
2.	Answers important internal and external questions for Caltrans	Corporate data sets contain data that is used frequently by Communities of Interest to answer important questions posed by both internal and external sources.
3.	Required or requested by control	Corporate data sets contain data that is important to control

#	Criteria	Description
	agencies and/or partners	agencies (e.g. Department of Finance, Legislative Analysts, or Business, Transportation and Housing Agency) and/or Partners (e.g. local agencies, regional transportation agencies, other state agencies, or the federal government).
4.	Data discrepancies cause difficult issues across Caltrans	When data discrepancies occur in corporate data sets, the result is often highly problematic and creates issues that are difficult for Caltrans to resolve.
5.	High visibility Caltrans data	Corporate data sets are highly visible within Caltrans, to external partners, and to the public. Problems with corporate data sets often become very visible and create complex problems within and outside of Caltrans.
6.	Supports Performance Management and Fiscal Management	Corporate data sets contain data that is often used by Caltrans for key performance management projects and tasks, and impacts Caltrans' fiscal management processes.

The Board's focus on corporate data sets reflects the desire to get the "biggest bang for the buck" from data governance efforts as Caltrans' data governance matures. This requires that the subjects of early data governance efforts have a broad impact and address issues that will cause Communities of Interest to take notice of data governance success stories.

The Board has reviewed a candidate list of Caltrans transportation system data sets (developed from interviews of subject matter experts and other research) to determine which satisfy the criteria to be corporate data sets. The table below contains a partial list of Caltrans' corporate data sets, including a description of each data set, the impacted business function(s) or program(s), and sample systems that use or house the corporate data set. A longer list, including all potential corporate data sets, is included in Appendix M. While the table below contains high level information (definition, business functions, systems) for each corporate data set, Chapter 4 presents a more complete set of identifying information or metadata.

Table 6: Caltrans Corporate Data Sets

#	Corporate Data Set	Definition	Business Functions	Systems
1.	Collision data	Collision data is recorded from the California Highway Patrol's (CHP) traffic collision reports that include all reported incidents on the state highway system. The data from these traffic collision reports is entered into TSN (the Transportation System Network – this data was previously entered into TASAS) from hard copies of the traffic collision reports. An estimated County/Route/Postmile location is also included and derived from the location description provided by the CHP officer.	Operations	TSN
2.	AADT (Annual Average Daily Traffic)	AADT data is derived from formulas applied to traffic census data that measures the level of traffic on the state highway at individual locations for specific dates and times. Traffic census data is gathered either through PeMS (Performance Measurement System)	Operations Planning Design	TSN PeMS

#	Corporate Data Set	Definition	Business Functions	Systems
		detector systems built into the travel way, or through other traditional traffic census methods. AADT calculations are contained in TSN (previously contained in TASAS) and are used for various operations, planning, and design purposes throughout Caltrans.		
3.	VMT (Vehicle Miles Traveled)	VMT data is derived from AADT data and measures the total number of miles traveled by vehicles, over a specific length of highway, for a specific time period. VMT is currently reported in both TSN and the Highway Performance Monitoring System (HPMS), although the VMT amounts in each system differ because of the use of different highway segments in calculating VMT.	Operations Planning Design	TSN HPMS
4.	Truck Volumes	Truck volume data is similar to AADT except that it captures only the level of traffic traveled specifically by trucks. Trucks are divided into various classifications based on the number of axles on the truck. Some truck data is gathered by Weigh-in-Motion (WIM) systems on the highways, or by other traditional traffic census methods.	Operations Planning Design Maintenance	TSN HPMS
5.	Performance Measurement System (PeMS)	<p>PeMS is both a data set and a system in that it is a system that gathers unique, real-time traffic volume data that is captured by automated counting devices (i.e. loop detectors) built into the travel way of the highway system. PeMS is currently only located in urban areas of the state. Data from the detectors is transmitted to the PeMS system, which is used as the basis for communicating real-time traffic information and developing traffic census data for specified locations.</p> <p>In addition to PeMS, two additional systems are gathering real-time traffic volume. One is for local arterial roads (APeMS) and the other is for transit ridership (TPeMS).</p>	Operations Planning	PeMS TSN
6.	Pavement Condition	Pavement condition information is derived from the "Pavement Condition Survey" which is an in-person assessment of the pavement's condition on the entire state highway system. The results from the in-person assessments are entered into the Pavement Management System (PMS) for use in determining future needs, planning projects, measuring the effectiveness of past pavement treatments, and other purposes.	TSI Planning Design Construction Maintenance Operations	PMS HPMS

#	Corporate Data Set	Definition	Business Functions	Systems
7.	Level of Service (LOS)	<p>Two types of LOS are used in Caltrans. One is used to measure the performance of highway maintenance and a second is used to measure traffic performance.</p> <ul style="list-style-type: none"> • <u>Maintenance LOS</u> is based on field reviews conducted each year, consisting of 10% of the highway system's center-line miles. Reviewers determine whether or not the inventory (i.e. guardrail, striping, fencing, etc.) of a particular postmile passes or fails based on established maintenance thresholds. This data is entered into the Maintenance LOS system and is used to calculate LOS scores for districts and regions based on accumulated review results. • <u>Traffic LOS</u> is a calculated measure of effectiveness for various types of highway facilities (freeways, ramps, highways, intersections, etc.) based on related performance measures (density, delay, speed, etc.). This LOS measure is used in Traffic Impact Studies, Route Concept Reports, etc. 	Maintenance Operations Environmental Planning	Maintenance LOS System
8.	County/Route/ Postmile	County/Route/Postmile data provides essential location data along the state highway system. This data is used in many systems but the primary system of record for County/Route/Postmile is TSN. Caltrans has also built a GIS-based LRS tool that links County/Route/Postmile location data to geospatial coordinate data.	Multiple	TSN HPMS State Roadway Network (GIS)
9.	Functional Classification	Functional classification provides a method for classifying highways and local roads according to the type of service they provide. It is used in the determination of federal aid eligibility, transportation planning processes, and project design, among other processes. Functional classification data is provided in the HPMS system for federal reporting purposes and is also recorded in California Road System (CRS) maps.	Multiple	HPMS CRS maps
10.	System Designations	System designations data refers to the designation of state highways and highway segments according to Federal Aid System designations (i.e. the Interstate System, the National Highway System, and other related subsystems).	Multiple	TSN HPMS
11.	City/County/District	City/County/District data refers to city and	Multiple	TSN

#	Corporate Data Set	Definition	Business Functions	Systems
		county governmental designations and Caltrans District designations, which are used in multiple systems to provide location and jurisdiction information.		HPMS State Roadway Network (GIS)
12	Geographical Information System (GIS) Data	GIS data and geospatial data are developed in a variety of systems in Caltrans. The primary system is the State Roadway Network maintained in the Division of Transportation System Information. Other systems are maintained in the Division of Maintenance, and in some districts. In addition, the Division of Design has resources dedicated to GIS to manage the capital project related GIS (and CADD) data used in project development.	Multiple	State Roadway Network (GIS)

These corporate data sets are foundation data products that may be further refined and developed by the data governance efforts undertaken by Caltrans and managed by the Board. The next section describes the process by which these corporate data sets and all transportation system data at Caltrans will be assessed for quality issues.

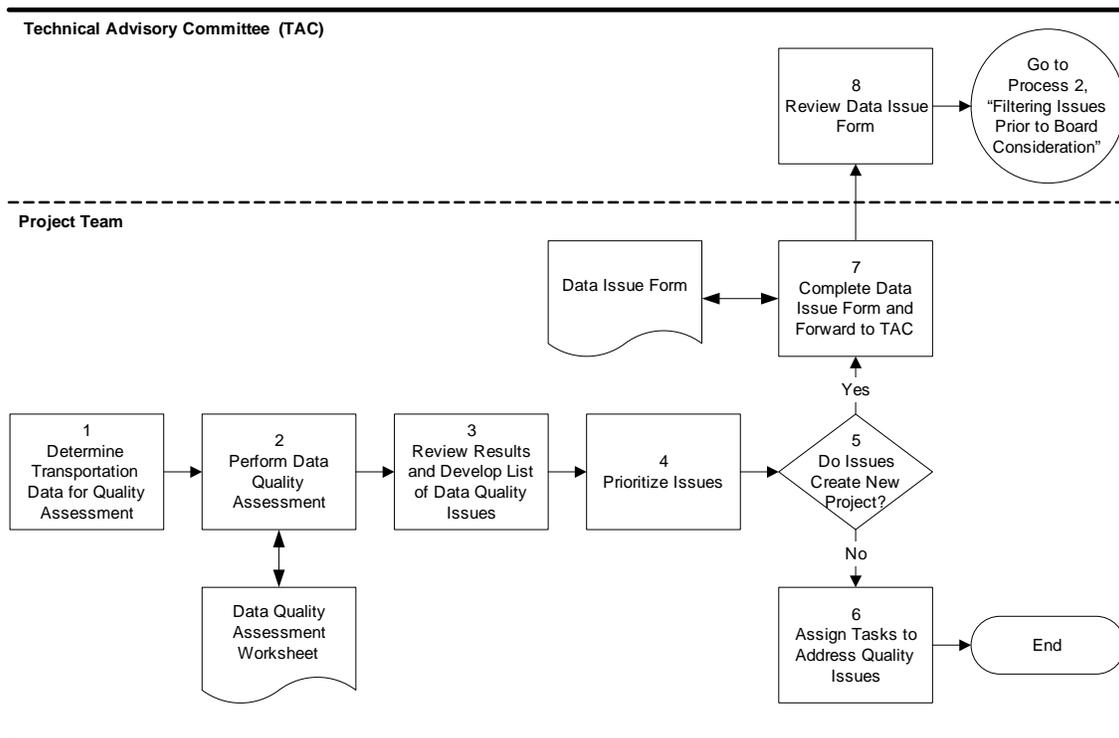
3.3 Data Quality Assessment Process and Tools

Data quality assessment is a process designed to identify the most important business needs and associated data problems that impact a particular data set. During the course of a data governance project, the project team may need to perform a data quality assessment to determine what data quality issues exist and address those issues in priority order, based on an issues impact on the business users and their processes. If a data quality assessment is necessary, the project team will determine which data will be assessed to meet the most urgent business needs of the business functions involved with the project and will then perform the assessment. The assessment will result in a list of data quality issues that the project team may have to address. Sometimes these issues may be complex or sensitive, and will require the establishment of one or more projects. The figure below shows the relationship between the “Data Assessment Criteria,” the “Data Set” and “Issues” that may arise, and potential projects to address the issues.

Figure 11: Data Assessment Process Sample

The process to perform a data quality assessment consists of defining the relevant business needs and addressing a series of questions that reflect common data quality issues found in data sets. The process map below provides a high-level overview of the process surrounding a data quality assessment, including what triggers the assessment and what occurs after the assessment is completed. A larger process map and the process narrative, which explains the process steps in more detail, may be found in Appendix V.

Figure 12: Data Quality Assessment



The questions in the table below are included in the “Data Quality Assessment Worksheet” included in Appendix J. This form can serve as a guide to project teams that undertake data governance projects and begin to address the data quality issues involved with the projects.

Table 7: Questions for a Data Quality Assessment

#	Question	Rationale
1.	Does data exist?	<p>The first consideration when assessing data is whether or not the data exists. Often, data quality issues arise because a user has a need for data that is not available or is unknown. If the data does not exist, those assessing the data quality issue must determine the following:</p> <ul style="list-style-type: none"> The essential business need that will be met by developing new data Resources required to gather the desired data Value added to Caltrans for having the data Estimated Return on Investment (ROI) for gathering the data <p>For every data set or element, there is a cost of gathering and maintaining data, which requires that a known business need will be served by the data before a resources are devoted to gathering the data. While a precise cost and benefit of having the data may not be known, it must be estimated so that a reasonable ROI analysis can be performed.</p> <p>For example, if the estimated cost of gathering very precise location data is prohibitive, perhaps a less precise version of the data can be gathered at a lesser cost and still provide value to the user. Performing this cost/benefit analysis and determining ROI for data gathering is an essential first step before gathering new data.</p>
2.	Is the data duplicated elsewhere?	<p>Communities of Interest from different areas of Caltrans sometimes independently gather data for their own specific needs, which results in duplicate sources of the data. This can lead to confusion amongst Communities of Interest as to which source is the “correct” or “official” version of the data. It also creates duplicated efforts in gathering, maintaining, and disseminating the data. This leads to extra costs that, with the confusion of duplicate sources, create an unnecessary burden for Caltrans.</p> <p>Often duplicated data is not exactly the same but instead overlaps, which causes confusion among Communities of Interest who find it difficult to differentiate between the two sources of data. This leads to confusing data definitions (see next question) and misuse or misreporting of data.</p> <p>Those performing data quality assessments must identify duplicate data or overlapping data and meet with the owners of these different data sources to reach a consensus about which data source is the most appropriate to use as the official source. This includes eliminating, to the extent possible, duplicate data gathering, maintenance, and dissemination efforts. As changes are made in the approved source of data, this information must be communicated to the Communities of Interest to ensure they are informed of the changes.</p>
3.	Does the data have a single, clear definition?	<p>Related to the data duplication and overlap issues discussed above, there are often issues associated with having a single data definition. Poorly defined data may lead to it being used or reported inappropriately, or, if users cannot understand a data definition, it may lead users to create a new version of the data if they are confused as to whether the desired data already exists.</p>

#	Question	Rationale
		<p>When performing a data quality assessment, it is necessary to find <u>all</u> data definitions (including <u>related</u> data definitions) for the data being assessed. SMEs and Business Owners of the data in question must work together closely to reach a consensus on the proper definition of each data element, and in turn, perform the following tasks:</p> <ul style="list-style-type: none"> • Correcting inaccurate definitions • Redefining overlapping definitions to remove ambiguity • Identifying duplicated data • Coming to an agreement on a single, clear definition for each data element • Determining if new data is needed to complete a data set that was previously filled by a poorly defined data element <p>As new or clarified data definitions are developed and agreed upon, these definitions will have to be communicated to all impacted Communities of Interest. When the centralized databases for data elements and data sets are established, new or improved data definitions may be available through these databases via the data governance website.</p>
4.	Is the data accurate and valid?	<p><u>Data Accuracy</u></p> <p>Data that is gathered is often not used because of its questionable accuracy and unreliability. Data accuracy must be measured against a reliable, objective source to determine whether or not data accuracy issues exist.</p> <p>When performing a data accuracy quality assessment, it is first necessary to determine the opinions of Business Owners, Communities of Interest, and other SMEs. This is done to estimate the extent of any data accuracy issues and to determine the availability of objective information to compare the data to and measure its accuracy. For example, if the “subject data” shows that 2,000 feet of guardrail was repaired for a specific location, but the “inventory data” for that same location shows that no guardrail exists, then there is a problem with the accuracy of the “repair data.” Extensive comparisons with other objective information may be necessary to determine the extent of the data accuracy issue and where it occurs.</p> <p><u>Data Validity</u></p> <p>A related issue is “data validity,” which is the degree to which data falls within an acceptable range of expected values. For instance, if data is intended to report <u>tons</u> of asphalt used per lane mile of highway, but data is entered as <u>pounds</u> of asphalt, the result will be outside of the expected value range for that specific data. This can often be the result of misunderstood data entry rules or defective formulas used in quantifying the data.</p> <p>Once the existence of a data accuracy or data validity issue is determined, it will be necessary for the assessment to uncover the cause of the problem. Some causes of the issue may include a programming issue requiring an IT solution, or a business process/business rule issue that will require business process review, training, and communication.</p>
5.	Is data complete?	<p>“Data completeness” refers to whether or not the data is entered into all of the fields in which the data should exist. If a data element is expected to be entered into a database on a monthly basis, but there are blank fields for months in which no data is entered, then there is an issue of completeness. Communities of Interest, expecting</p>

#	Question	Rationale
		<p>complete data, will find the data unreliable or will mistakenly report the data under the assumption that the data is complete.</p> <p>Incomplete data can be the result of poor data entry or maintenance habits, poor communication of data (so that it is not available for entry when anticipated), or lack of understanding of data that does not meet the anticipated parameters (i.e. it's not available each month; it's not recorded for every county and route, etc.).</p> <p>In assessing data for completeness, it will be necessary to conduct a thorough review of the data to ensure that all required fields contain expected data. If it is found that the data is incomplete, Communities of Interest and subject matter experts will be required to investigate the processes by which the data is entered and maintained to determine where the process issues exists. When the issues are located, the business processes will be revised and communicated to all staff, and those impacted will be trained in the new processes.</p>
6.	Is the data maintained?	<p>Data must be maintained periodically by updating it, and refreshing old data with new data. This may occur on a consistent basis through ongoing data entry or on a periodic basis by updating an entire data set at once. Data that is not maintained will become obsolete and unreliable, and Communities of Interest will stop using the data and perhaps even create their own data in its place (leading to "data duplication" – see above).</p> <p>When assessing data maintenance, it is necessary to know how frequently the data needs to be maintained to meet the needs of the Communities of Interest. If data is needed monthly, but is only updated quarterly, it will become increasingly less useful to Communities of Interest as each quarterly update becomes more out-of-date.</p> <p>Also, it is necessary to determine whether or not the data is being updated as frequently as planned. Sometimes issues may arise that lead to data not being updated; for example, lack of resources, poor communication, or other issues. If a data set is meant to be updated monthly, but is only updated twice a year, the needs of Communities of Interest are likely not being met.</p> <p>In addition, data maintenance is a costly activity. If it is determined that it costs more to update data than the benefit obtained by Caltrans, then the ROI of data maintenance may indicate the need to stop supporting the maintenance of the data set.</p> <p>Communities of Interest and subject matter experts will need to examine the data maintenance processes to determine whether or not the needs of Communities of Interest are being met or if an ongoing business need exists to justify the ongoing cost of maintaining the data. If the business need still exists, the owner of the data maintains the data adequately, and, if there is a resource issue related to the cost of maintaining the data, an ROI analysis will be needed to justify the continued maintenance of the data.</p>
7.	Is the data accessible?	<p>Data may not always be accessible to Communities of Interest who could benefit from having access to the data. Data may not be accessible due to technology limitations, lack of communication, and lack of access to siloed systems. This lack of access may be discovered when a business user requests the creation of a new data set, only to find that the data already exists but was unknown and/or inaccessible to the business user. This often happens when needed data is locked up in siloed systems that cannot</p>

#	Question	Rationale
		<p>be accessed directly by potential Communities of Interest, requiring the assistance of Business Owners to access and provide the data.</p> <p>When assessing data accessibility, the project team should determine who needs access to the data. Subject matter experts may be able to develop a list of potential Communities of Interest that need access to the data or this information may be gathered through a survey of potential users.</p> <p>For those who have access to the data, it will be important to determine whether or not they are satisfied with their access, or if there are barriers in accessing the data, making it more difficult for the business user to utilize the data.</p> <p>Lastly, if barriers exist, the project team should identify these barriers and remove or reduce them in order to provide greater access to the data. An example of a barrier may be the result of the data being sensitive (see below) and intentionally not being available to a wide group of Communities of Interest. In other instances, system issues beyond the scope of data governance will need to be addressed. In many cases, it may only require that Communities of Interest receive proper training on how to access the data, or requiring Business Owners being more responsive in providing data. In these instances, data governance solutions will be able to address the above issues.</p>
8.	Is the data timely?	<p>Timely data is vital to many business functions. One of the issues confronting Communities of Interest is the ability to use data when it is needed. Having to resort to obsolete data provides little value added to the end-product and can be embarrassing when providing information to other agencies or the public.</p> <p>The project team should determine the timeliness needs of Communities of Interest and whether or not they're being satisfied. Often, there will be constraints imposed by systems, data gathering obstacles, or other agencies that limit the frequency that data can be provided. In some cases, it may be a lack of communication between the needs of the Communities of Interest and Business Owners of the data regarding when the data is needed. In other instances, a review of business processes will indicate the ability to provide data in a more timely fashion to meet the timeliness needs of the Communities of Interest.</p>
9.	Is the data sensitive?	<p>Some data at Caltrans is confidential due to policy, legal, or personnel considerations. As such, although there may be a legitimate need for the data by some Communities of Interest, access to and use of the data will be restricted. Examples of confidential data would be data related to traffic collisions, lawsuits involving Caltrans, labor disputes, etc.</p> <p>The project team should first ensure that the data in question is not confidential before providing recommendations to change how the data is managed or made available. If it is not known whether or not the data is confidential, the Legal Division can be contacted to provide assistance and advice.</p>

Section 3.2, "Caltrans Corporate Data Sets," described the hierarchy of Caltrans data from a data governance perspective. Data governance focuses on the highest value corporate data sets, evaluates the available data to identify data quality issues, and then initiates projects to address the most pressing data quality issues. The next section (3.4 - *Assessing Caltrans Corporate Data Sets*) describes the initial review and high level assessment of three corporate data sets to identify corporate data set pilot projects.

3.4 Assessing Caltrans Corporate Data Sets

An important objective of the Board is to improve the way the Caltrans transportation system data meets the business needs of Caltrans business functions, and to increase the ability of Caltrans to make improved business decisions based on reliable data. Earlier sections presented corporate data sets followed by processes and tools to assess data quality. This section presents three data sets and an initial assessment of their quality. The Board has chosen to initially focus on three high-value corporate data sets as the subject of the first data governance efforts and the first efforts to produce data products at Caltrans. The three “pilot” corporate data sets are:

- Linear Reference System (LRS) and County/Route/Postmile (C-R-PM)
- Vehicle Miles Traveled (VMT)
- Annual Average Daily Traffic (AADT)

The Board chose these three corporate data sets from the larger list of corporate data sets provided in Appendix M based the criteria shown in the table below:

Table 8: Data Governance Pilot Selection Criteria

#	Criteria	Description
1.	Resolve one or more difficult and high visibility data quality issues	The selected corporate data sets must include at least one high visibility data quality issue that the pilot data governance effort can resolve to not only improve the data, but also improve the visibility of data governance at Caltrans.
2.	Demonstrate the value of the Data Governance Program	The pilot effort must be able to provide a means to demonstrate the value of data governance to both the participants in the data governance effort and to Caltrans as a whole. This will increase the demand for data governance at Caltrans and improve the ability to succeed in addressing other corporate data quality issues.
3.	Be a first effort of many leading to a mature Data Governance Program	Data governance at Caltrans is in its beginning stages. The Board selected pilot corporate data sets to serve as initial efforts to begin the process of maturing the Data Governance Program. These pilot efforts will develop and implement processes and tools that will further increase the capacity to address data quality issues in the future and build support to allow data governance to mature.
4.	Have a limited and achievable scope	For a data governance pilot to succeed, it must be “doable.” This means that it must have a scope that is limited and achievable within the limits of the resources currently available to data governance at Caltrans. These pilot efforts will provide initial successes to support building data governance efforts in the future.
5.	Focus on a few high priority corporate data sets	The Board does not want to take on too much, too soon. Therefore, the decision was made to narrow the focus of the initial efforts for data governance on a limited number of corporate data sets to increase the chances for success in each effort. These early successes will help kick off data governance and increase the chances for continued success over time.

Following the selection of the corporate data sets (for the initial data governance efforts) a series of interviews were conducted with SMEs related to each corporate data set. These interviews focused on identifying the most important data quality issues that impact the business needs of the units that own and use the corporate data sets.

In total, eight Caltrans SMEs were interviewed in six interviews, with each interview lasting one to two hours. The interviewees were:

- Sarah Chesebro (Division of Transportation System Information)
- Mark Samuelson (Division of Transportation System Information)
- Nick Compin (Division of Traffic Operations)
- Harold Feinberg and Greg White (Division of Transportation System Information)
- Jeff Duket (Division of Maintenance) and Fay Meeks (Division of Transportation Planning)
- Jane Berner (Division of Traffic Operations)

The questions for each interview were developed from the general subject matter covered in the Data Quality Assessment Worksheet (Appendix J), and from background information gathered during many interviews conducted in an earlier phase of the Data Business Plan project.

The results of these interviews are shown below, along with a listing of the data quality issues identified during the discussions.

3.4.1 County-Route-Postmile Background & Overview

The Caltrans County-Route-Postmile (C-R-PM) system is Caltrans' original sequential listing of landmarks assigned a postmile value increasing in miles from south to north and west to east. Previously, the Traffic Accident Surveillance and Analysis System (TASAS) contained Caltrans' official C-R-PM data for the state highway system. Since the transition of the TASAS data into the Transportation System Network (TSN) system, the C-R-PM data has been maintained within TSN.

The TASAS system was created for traffic safety analysis purposes. As it developed over time, it became the "gold standard" for the maintenance of C-R-PM data in Caltrans. Because of its primary purpose of tracking traffic safety data, the maintenance and updating of data previously within TASAS (now TSN, and including the latest revisions to C-R-PM data), is dependent on following detailed business rules required for traffic safety data. Often, this requires that changes in the C-R-PM data cannot be updated within TSN until other data (including traffic collision data and AADT data) is developed and entered along with the C-R-PM updates. In addition, TSN data must maintain the chronological record of changes so that the precise historical circumstances, relative to traffic accidents, can be recorded and maintained relative to the location on the highway at the time of the accident, and is not lost when C-R-PM updates occur.

C-R-PM data within Caltrans represents locations on the highway, not odometer measurements. This is a critical distinction that is important for the future development of C-R-PM and GIS LRS in the future. The C-R-PM data structure is based on the county code, route number, and postmile. Postmiles start at "0.0" at the beginning of one county line along the route, and end at the end of the route within the same county (usually, but not always at a different county line). Because of route realignments (due to construction, relinquishments, route adoptions, etc.), postmile numbers change over time to reflect the new alignment. Instead of changing the postmiles along the entire county-route to reflect a new odometer-based measurement, only those sections subject to change receive new postmiles. This is often achieved using a postmile prefix to indicate that the postmile is the result of realignment. Therefore, the postmile numbering structure is used to provide locations that do not change and do not provide a guide for measuring mileage along the route. In addition, historical postmiles are maintained so that prior route alignments continue to have their unique location postmile indicators. This allows information (i.e. traffic accidents) to be related to a unique location at a particular point in time.

3.4.2 Geographic Information System (GIS) Linear Referencing System (LRS) Background & Overview

The LRS at Caltrans today is based on GIS capabilities that use the C-R-PM data in TSN as a framework. GIS longitude and latitude data is calibrated to the C-R-PM data to show the location of events (e.g., traffic accidents) or features of the state highway system using GIS mapping capabilities and C-R-PM data.

GIS at Caltrans was originally developed using ESRI ArcInfo software and GIS data from the United States Geological Survey, called Digital Line Graph (DLG). In the past, the data was available to state agencies through the Teale Data Center, which provided for a high level of consistency of GIS data in use by the state (although the basic GIS line data was not as accurate as later GIS data). The DLG data, operated by the older ArcInfo software, is still the basis of GIS functions operating in the Division of Maintenance and the Division of Transportation Planning, and is used as the LRS for these two divisions. Although this older data and software is not as spatially accurate as later GIS data, it has a more accurate linear representation (i.e. the GIS data is more consistent with the latest available C-R-PM data) because it has been maintained. Because most of the GIS analysis performed within Maintenance and Planning is “macro-level analysis,” the consistency of the LRS with C-R-PM data is more important than the micro-level spatial accuracy of later software and GIS data.

The GIS office within the Division of Transportation System Information (TSI) is using a newer GIS software version called ArcView. In addition, TSI now uses more recent and more accurate GIS data. After first purchasing private GIS data (called “TANA”) as the GIS line data to form the basis of the Caltrans’ GIS functions, TSI now uses United States Census Bureau’s “TIGER” (Topologically Integrated Geographic Encoding and Referencing system) GIS line data (also known as “LIGER,” a Caltrans modification of TIGER that references Caltrans LRS). The TIGER GIS line data is a better spatial representation of the state highway system and provides for better spatial accuracy than the older DLG data used by Maintenance and Planning. But because of a backlog of highway alignment changes, the TSI LRS does not contain the linear accuracy of the Maintenance and Planning tools, and therefore does not meet the needs of users in Maintenance and Planning because of the lack of consistency with the latest C-R-PM data.

A basic inconsistency between the TIGER GIS data and the C-R-PM data is the odometer measurement orientation of the TIGER GIS data and the location orientation of the C-R-PM data. In developing the TIGER GIS data for use by Caltrans, TSI calibrates the GIS data so the odometer measurements of mileage along a county-route are correlated to the corresponding C-R-PM. This creates a LRS that is based on C-R-PM data and GIS data that can then be accessed by Caltrans users.

TSI has also created various tools for Caltrans users so that the LRS can be accessed via a common search format, rather than requiring the knowledge of complicated queries to create usable GIS data. One of these tools is a web-service postmile tool that allows users to request GIS data for a single or multiple postmile locations and receives a GIS map of the locations requested. This web-service postmile tool has been upgraded and is often used for finding the postmile location of traffic accidents on state highways based on the physical description of the accident, as described in the traffic collision report. Currently, both the original and newly updated web-service postmile tools are available to be used, although there are plans to retire the older version of the tool in the near future.

3.4.3 County-Route-Postmile and LRS Assessment

The results from the C-R-PM and LRS interviews are presented in terms of the data assessment guidelines provided in “*Table 7: Questions for a Data Quality Assessment*” and summarized in Appendix J. The table below provides an overview of the issues and a summary statement of the issues that need to be addressed. As noted earlier, these high level assessment findings will likely lead to one or more projects to address the data quality issues.

Table 9: County-Route-Postmile/LRS Data Quality Issues

#	Data Quality Issues	Description
1.	<p>Insufficient process for Addressing Data Changes/Updates</p> <ul style="list-style-type: none"> There is a need for a thorough and detailed Caltrans LRS maintenance process that captures the most recent alignment changes in the C-R-PM data and updates the GIS odometer data calibration to the C-R-PM data. 	<p>Odometer mileage measures vs. C-R-PM locations:</p> <p>Postmiles are highway location data, not mileage measures. GIS data more easily conforms to an odometer-based measure of a highway. Highway alignments change for a variety of reasons (e.g. the adoption of a local road into the highway system, the relinquishment of a state highway to a local agency, the construction of a bypass around a business district, the straightening of a highway alignment to eliminate curves, etc.) which changes the odometer measurement of the highway; however, postmile locations on either side of the impacted highway segment <u>do not</u> change to maintain the consistency of location indicators provided by the postmile system. This consistency is necessary over time to provide the temporal, linear, and spatial reference required of the Caltrans LRS. In particular, traffic safety data must be able to be linked to the specific location and characteristics of a highway at the time an event occurs on the highway (often a traffic collision). After each highway iteration, the LRS cannot be recalculated to accommodate new odometer measurements without preserving the unique location characteristics of its development. As a result, GIS odometer measurements must be recalculated from the beginning to the end of the county-route and also recalibrated to the new postmile locations on the highway following a change in a highway alignment.</p>
2.	<p>Duplicate Data</p> <ul style="list-style-type: none"> Multiple and siloed LRS systems exist, each requiring independent and duplicate maintenance / updating efforts. <p>Inconsistent Data</p> <ul style="list-style-type: none"> GIS LRS data between the separate LRS systems within Caltrans is inconsistent, leading to confusion among users and potential inconsistent reporting of GIS data internally and externally. 	<p>Multiple LRS systems are in use in Caltrans:</p> <p>There are many justifiable reasons why data quality issues occur. First, in addition to the Caltrans LRS, developed by TSI, the Division of Maintenance built and maintains their own LRS because of LRS design differences between the TSI LRS and the Division of Maintenance LRS. The Division of Maintenance, Division of Transportation Planning and Division of Environmental Analysis feel this LRS is more linearly accurate and consistent. Second, certain districts have GIS data in their LRS systems that are unique to their districts. Lastly, there is GIS staff in districts and divisions throughout Caltrans that use different LRS systems.</p> <p>In the past, Caltrans attempted to establish committees to provide coordinated direction to Caltrans' GIS efforts (including the primary GIS functions), without success. Most recently, a GIS Value Analysis Study has made a series of recommendations, including recommendations regarding the establishment of an authoritative body, to develop coordinated GIS policies and procedures (this function has been undertaken by the Board).</p>
3.	<p>Data is not adequately maintained</p> <ul style="list-style-type: none"> Highway realignment data to be entered into TSN is consistently backlogged. C-R-PM data in TSN is out-of-date because of 	<p>TSN C-R-PM data is out-of-date because a backlog of highway inventory alignment changes need to be updated into TSN:</p> <p>TSN's C-R-PM data is based on prior TASAS data and processes. TASAS data was originally developed and maintained to record and analyze traffic collision data on the state highway system. The C-R-PM data is carefully maintained with a detailed chronological record so that collision data, traffic volume data, and highway characteristic data can be precisely located at any time. The necessary precision and care for the maintenance of this data creates delays in the update of</p>

#	Data Quality Issues	Description
	<p>the backlog of highway realignment data that is waiting to be entered into TSN.</p> <ul style="list-style-type: none"> Chronological requirements for data entry into TSN create additional difficulties and delays in updating C-R-PM data Any LRS based on TSN C-R-PM data will also be out-of-date to the degree that TSN C-R-PM data is not updated. 	<p>the changes to the C-R-PM data. The result is that the LRS is based on the out-of-date C-R-PM data, and is therefore not an up-to-date reflection of the current state highway system. Some of the business requirements to update C-R-PM data include:</p> <ul style="list-style-type: none"> Changes to highway alignment must be entered in chronological order so that the historical characteristics of the alignment are captured and associated data can be entered with the correct chronological reference. Traffic volume data must be available and able to be entered with the updated route alignments before the updated route alignment can be accessed by users. Any traffic collision data that is associated with the updated route alignment must be identified and entered with the updated route alignment to ensure the accuracy of the traffic collision data in the system and the accuracy of any analysis that resulted from the data. <p>Because Caltrans must maintain the integrity of the traffic accident data and update processes that are inherent to the maintenance of this data, a backlog exists in updating the C-R-PM data in TSN, resulting in out-of-date data that reflects recent changes to the highway system.</p>
4.	<p>Data is not timely</p> <ul style="list-style-type: none"> Existing policy mandates that update to TSN C-R-PM data does not occur until at least six months following the opening of the realigned route to traffic. Encroachment permit realignments are not reported to TSI on a regular basis. Local agency project realignments are not reported to TSI on a consistent basis. 	<p>Some TSN C-R-PM data is not available for update on a timely basis:</p> <p>Highway alignment changes that are provided to TSI, for input into TSN, primarily come from highway project information. Caltrans policy requires waiting six months following the opening of the new highway alignment to traffic before new alignment data can be input into TSN. This creates, at a minimum, a six month delay to input highway alignment changes into TSN.</p> <p>In addition, the project information provided to TSI does not include all of the alignment changes that impact the highway system. In particular, alignment changes that result from encroachment permits and local agency projects are not communicated to TSI unless district staff provides the information after the fact. Currently, there is not a process in place to ensure that all of these changes are reported to TSI.</p>
5.	<p>Chronological Data Updates are complex and lengthy</p> <ul style="list-style-type: none"> Lengthy, manual processes are required to maintain accurate historical highway alignment calibrations between GIS geospatial data and C-R-PM data. This increases the 	<p>Calibration of GIS data to C-R-PM data is complex due to chronological update requirements:</p> <p>The latest GIS geospatial data in TSN is periodically calibrated to the latest C-R-PM data to provide a GIS-based LRS that ties geospatial data to the C-R-PM location data. However, updates to capture changes to highway alignments within the GIS LRS can become problematic and time-consuming. This is because, currently, manual processes are necessary to maintain the chronological accuracy of the calibration of GIS geospatial data and C-R-PM data. Calibrations of GIS geospatial data to changing C-R-PM data must be chronological and reflect accurate historical highway alignments because TSN must</p>

#	Data Quality Issues	Description
	complexity and delays in maintaining the GIS LRS.	be maintained on a chronologically accurate basis.
6.	<p>Data is not current</p> <ul style="list-style-type: none"> Out-of-date data can pose problems for analysis of traffic safety data. Caltrans staff is not fully informed of data obsolescence issues. 	<p>Caltrans staff not always aware that traffic safety data is not current:</p> <p>It is not widely communicated in Caltrans that collision data, AADT data, and C-R-PM data is not current at all locations along the highway system. This can create business issues with Division of Traffic Operations staff who are analyzing data for the purpose of selecting highway locations with potential safety issues.</p>
7.	<p>Lack of Data Quality and Completeness</p> <ul style="list-style-type: none"> Quality control is not systematic and occurs only on an ad hoc basis by district staff that discovers errors while performing other tasks. Staff entering C-R-PM updates into TSN does not follow the same processes when performing data updates. 	<p>TSN C-R-PM data quality control needs to be improved:</p> <p>Quality control reviews are not a consistent part of the alignment update process. Instead, TSI is dependent upon district feedback for discovery of problems with C-R-PM updates in TSN. This district feedback occurs on an ad hoc basis only when districts discover the C-R-PM error while performing other tasks. There is no consistent quality control process to ensure the accuracy of TSN C-R-PM data updates. Also, staff performing the C-R-PM updates does not necessarily follow a consistent process for updating C-R-PM data in TSN, which can also lead to errors and/or inconsistencies in the data that is updated in TSN.</p>
8.	<p>Unrealistic Data Standards</p> <ul style="list-style-type: none"> Data standards call for postmile location precision that is not realistic and may not be necessary. <p>Unclear Data Standards</p> <ul style="list-style-type: none"> Caltrans staff is not always aware of the meaning, interpretation, and proper use of C-R-PM data standards. Poor recording of C-R-PM data leads to errors in displaying items on the highway in the GIS LRS. 	<p>Data Standards are unclear and may be unrealistic:</p> <p>While there is a data standard for the type, length, and description of C-R-PM data, there is an issue with the accuracy of the data. The current data standard for C-R-PM data is to define the postmile location to the thousandths of a mile (three decimal places, i.e. 10.136); but the actual location data is not accurate to that degree of precision. This results in data that is beyond the current capabilities of Caltrans (to define locations to the thousandth of a mile). The data may not need to be that precise to be useful to Caltrans for providing locations on the state highway system.</p> <p>Also, the use and interpretation of the official C-R-PM data standard (particularly postmile prefixes and suffixes) are not known by all Communities of Interest in Caltrans. This can lead to inaccurate recording and communicating of postmiles in documents and spreadsheets that leads to further errors in the business processes impacted by the mistaken communications.</p> <p>When data standards for C-R-PM are not followed, it can impact the GIS LRS. Sometimes items on the highway system are entered into the GIS LRS without appropriate postmile suffixes or prefixes, or using rounded or estimated postmiles, causing the item to appear on the GIS map at a location that is different than its actual location.</p>

#	Data Quality Issues	Description
9.	<p>Limited Data Access</p> <ul style="list-style-type: none"> Access to TSN C-R-PM data is limited to TASAS coordinators. 	<p>TASAS data accessibility may be too restrictive:</p> <p>Because of Caltrans' policies regarding access to TASAS data, only TASAS coordinators are allowed to access the TSN C-R-PM data directly (although other users have access to postmile web-service tools associated with the GIS LRS). Users wanting TSN C-R-PM data must go through the TASAS coordinators to get the data which may be time consuming.</p>
10.	<p>Duplicate Data Entry</p> <ul style="list-style-type: none"> Data is entered by multiple staff. 	<p>Duplicate data entry relative to intersections:</p> <p>Because of TSN system issues, data related to intersections must be entered both by the project engineer and by the Traffic Census staff in Traffic Operations, resulting in duplicate data entry.</p>
11.	<p>Data does not meet federal reporting requirements</p> <ul style="list-style-type: none"> Need for future development of GIS LRS data for local roads to meet federal reporting requirements. 	<p>Future needs to incorporate local roads into LRS to meet federal reporting needs:</p> <p>Currently, the federal government requires information related to local roads to be reported through the Highway Performance Monitoring System (HPMS – a tool for federal reporting). In the future, federal reporting requirements will include GIS data for local roads that is currently not developed by Caltrans GIS LRS.</p>
12.	<p>Data Reporting Inconsistencies</p> <ul style="list-style-type: none"> TSN and HPMS route segmentation data is different, which results in reporting inconsistencies for certain transportation system data. 	<p>Differences in route segmentation business rules between TSN and HPMS:</p> <p>C-R-PM segments in TSN is based on a highway sequence listing in which the highway segments begin and end at various landmarks (i.e. intersections or interchanges) or changes in highway features (i.e. changes in number of lanes). Route segments in the HPMS are dependent on different factors that begin and end a segment (e.g. a change in the functional classification of a highway). Functional classification and other factors in HPMS are not contained in TSN. As a result, these two systems have different route segmentation. These differences create reporting difficulties for certain transportation system data (i.e. VMT – see below). The LRS systems currently in use by Caltrans can accommodate (process and display) any type of segmentation one would need to use, as long as this segmentation is defined by either beginning and ending postmile, or beginning and ending odometer measure.</p>

3.4.4 Annual Average Daily Traffic (AADT) Background & Overview

AADT is an important data set used throughout Caltrans for planning, design, and operational issues. The definition and description of AADT from the Caltrans 2009 Traffic Volumes on California State Highways is below:

“Annual average daily traffic is the total volume for the year divided by 365 days. The traffic count year is from October 1st through September 30th. Very few locations in California are actually counted continuously. Traffic Counting is generally performed by electronic counting instruments moved from location to location throughout the State in a program of continuous traffic count sampling. The resulting counts are adjusted to an estimate of annual average daily traffic by compensating for seasonal influence, weekly variation and other variables which may be present. Annual ADT is necessary for presenting a statewide picture of traffic flow, evaluating traffic trends, computing accident rates, planning, and designing highways, and other purposes.”

The Traffic Volumes data includes data elements for AADT, “Peak Month” volumes, and “Peak Hour” volumes. Peak Month volumes provide data for seasonal variations in traffic volume. Peak Hour volumes provide data for daily traffic congestion.

AADT is produced by the Caltrans Traffic Census Program in the Division of Traffic Operations. The Traffic Census Program performs traffic counts on state highways across the state. District staff performs the traffic counts, records the data on spreadsheets, and reports the data to headquarters staff, which then enters the district data into an Oracle-based system, designed for collecting traffic volume data. Traffic counts include numbers of vehicles, classification of vehicles, and vehicle weights (at locations with Weigh-in-Motion facilities).

Depending on the highway location, traffic counts can be performed continuously, quarterly, or every three, six, or ten years. The PeMS system records the traffic volume data that is used for the Traffic Census Program and to calculate AADT. At locations without PeMS, an Automatic Traffic Recorder (ATR) device is established to count traffic volumes. District staff is responsible for recording and reporting the traffic volume data captured by the ATR.

AADT data is used by Caltrans for determining VMT, traffic collision rates, and areas of congestion, and is used for operational analysis, design analysis, and future planning needs, among other purposes. In addition to Caltrans, AADT supports a variety of other customers, including metropolitan and local planning agencies, the Federal Highway Administration (FHWA), Federal Environmental Protection Agency (EPA), and the public. The federal government uses AADT extensively for apportionment, administrative, legislative, analytical and national highway database purposes, and the EPA uses it for air quality monitoring.

Caltrans assists some local agencies to collect traffic count data when local agencies don’t have the resources to gather and report the traffic counts on their own. Even though data collected by local agencies is of unequal quality, Caltrans collects local agency data and reports it to the FHWA through HPMS.

Caltrans is currently investigating new technologies for capturing traffic count data and estimating travel time and traffic speeds. Some of the technologies researched include roadside detectors that are equipped to detect “blue tooth” and other wireless phone technologies used by drivers on the highway system.

3.4.5 Annual Average Daily Traffic (AADT) Assessment

The results from the AADT interviews are presented in terms of the data assessment guidelines provided in “Table 7: Questions for a Data Quality Assessment” and summarized in Appendix J. The table below provides an overview of the issues and a summary statement of the issues to be addressed. As noted earlier, these high level assessment findings will likely lead to one or more projects to address the data quality issues.

Table 10: AADT Data Quality Issues

#	Data Quality Issues	Background Description
1.	Data Entry Duplication Duplicate data entry increases resource needs and the possibility for errors.	Traffic count data gathered by district staff must be entered twice; first by district staff into a spreadsheet that is then sent to Headquarters and re-entered again into the traffic counts database. This duplication of effort creates inefficiencies and increases the potential for data entry errors.

#	Data Quality Issues	Background Description
2.	<p>Data Anomalies</p> <p>Data anomalies in traffic count data are an indication of problems in data collection or data reporting. These anomalies lead to inaccurate traffic count and AADT data being reported.</p>	<p>Drastic, unexplained changes in traffic count data from year to year indicate potential data errors in one of the counts. This is resolved by smoothing the data over the years where the drastic differences occur. AADT data is calculated from the smoothed data. This is an indication of errors in data collection or data reporting and causes inaccuracies in the calculated AADT data.</p>
3.	<p>Differing Data Definitions and Accuracy</p> <p>Uneven data quality based on different traffic counting processes based on PeMS or ATRs.</p>	<p>Caltrans has two primary traffic counting processes. They are:</p> <ul style="list-style-type: none"> Automated, continuous data collection from PeMS traffic detectors located in the highway. These detectors are located in urban areas and provide a continuous flow of real-time data regarding traffic volumes at functioning PeMS locations. This continuously counted data is of generally high quality and can provide actual traffic volumes without the need for estimation based on sample data. Data collected by ATRs at prescribed frequencies and durations. From these data samples at each location, estimates of AADT, peak month, peak hour, and other calculations are performed for the location at which the count was taken. <p>The two different traffic counting processes lead to a difference in the data quality for the traffic counts at different locations on the state highway system. This can create varying levels of data reliability for AADT data dependent upon where the count was taken and the traffic counting process used. This situation may be resolved if PeMS is expanded to more of the highway system beyond its current concentration on urban areas.</p>
4.	<p>Local Data Incomplete, Inaccurate</p> <p>Caltrans does not have the means to assure quality data submittals from local agencies for reporting to the federal government.</p>	<p>Caltrans collects traffic count data for reporting to the federal government through HPMS. Caltrans has little quality control over the data reported by local agencies, and often the data is problematic and unreliable. This is often the result of inadequate resources at the local level to gather and report updated traffic volume data. This can cause underreporting of traffic levels in California, which may adversely impact federal calculations for federal assistance to the state and local agencies.</p>
5.	<p>Delayed Traffic Census Results in Incomplete Data</p> <p>Reporting of traffic counts from the districts is often delayed, creating delays in updating AADT data and other data associated with the traffic census. This also results in delays in creating statewide data and reporting data to FHWA.</p>	<p>Updated traffic counts and related AADT data cannot be produced until the traffic census is completed for the whole state highway system. For various reasons (primarily resource related) the completion of traffic counts is often delayed as district staff have their efforts transferred to other, higher priority tasks. The delays in performing the traffic counts result in delays in compiling the traffic census data and computing AADT data, and delays in submitting data to internal Caltrans customers and to the FHWA for federal reporting purposes.</p>
6.	<p>C-R-PM Data Backlog Results in Incomplete Data</p> <p>This creates a consequent backlog in the updating and reporting of traffic count and AADT data.</p>	<p>If locations at which traffic counts are performed are not entered into the C-R-PM data in TSN, the traffic count and AADT data cannot be entered and calculated. This causes a backlog of traffic count and AADT data that is not available for analysis and reporting.</p>

#	Data Quality Issues	Background Description
7.	<p><i>Increased FHWA Demands for Timely Data</i></p> <p>FHWA increases its requirements for data to be real-time and differentiated between lanes.</p>	<p>FHWA would like the states to be able to report additional traffic volume and AADT data that is increasingly in real-time, and differentiates between mixed flow lanes and HOV lanes. The cost of gathering and maintaining the new types of AADT data must be estimated and a determination made regarding Return On Investment and cost effectiveness.</p>

3.4.6 Vehicle Miles Traveled (VMT) Background & Overview

VMT measures the total of all miles traveled by all vehicles on a highway or highway segment. It is measured for a particular period of time (i.e. year, week, day, etc.) and is calculated based on AADT. VMT is the product of AADT and the highway segment length. It is used for a variety of purposes, including safety analysis, system planning, project design, and public information.

Within Caltrans, VMT is reported in TSN (previously in TASAS) and HPMS. Other state agencies that report VMT are the California Air Resources Board (CARB) and the California Energy Commission (Energy Commission).

VMT reported in TSN is collected by the Traffic Census program in the Division of Traffic Operations. This information is used for a variety of safety related purposes, including the Highway Safety Improvement Program and the calculation of accident rates on state highways. This data is available to other state and local agencies.

VMT reported in HPMS is reported to the FHWA for various uses, including determining apportionments and allocations of federal funds to states, analyzing highway performance, determining future highway system needs, and numerous other federal travel data reports. The information is also made available to the federal EPA for air quality reporting and forecasting. The frequency of VMT reporting for state highways is dependent upon the functional classification of each state highway segment.

Future VMT calculations are produced by travel modeling efforts within Caltrans. These models include: the California Statewide Travel Demand Model and the Motor Vehicle Stock Travel and Fuel Forecast (MVSTAFF). These models are used by various Caltrans divisions, other state agencies (i.e. CARB and the California Highway Patrol), Regional Transportation Planning Associations (RTPAs), Metropolitan Planning Organizations (MPOs), and federal agencies (i.e. FHWA and EPA). Model calculations of future VMT use current VMT as a basis for calculating future VMT.

Caltrans received complaints because of the different sources of VMT data used by and provided by Caltrans. As a result, a study was conducted by TSI and the Division of Traffic Operations to review the issues with these complaints and develop solutions. The study recognized that both sources of VMT (TSN and HPMS) have valid and necessary reasons for reporting VMT. In particular, TSN uses VMT in the calculation of accident rates using the detailed accident history data within TSN. HPMS does not have the accident history or other traffic safety data within TSN, and therefore cannot be used to report the accident rates or other traffic safety data. The study recommended that HPMS should be the “system of record” for VMT for all uses, except traffic safety within Caltrans, which would still be reported out of TSN. This study, called the “Vehicle Miles Traveled Data Sources Fact Sheet,” is included below in Appendix N.

3.4.7 VMT Assessment

The results from the VMT interviews are presented in terms of the data assessment guidelines provided in “Table 7: Questions for a Data Quality Assessment” and summarized in Appendix J.

The table below provides an overview of the issue and a summary statement of the issues to be addressed. As noted earlier, these high level assessment findings will likely lead to one or more projects to address the data quality issues.

Table 11: VMT Data Quality Issues

#	Data Quality Issues	Background Description
1.	<p><i>Inconsistent, Duplicate VMT Reporting by Caltrans</i></p> <ul style="list-style-type: none"> • Caltrans reports multiple versions of VMT data which are inconsistent. • No decision has been made regarding the choice of a single, official Caltrans VMT data source. 	<p>Caltrans maintains two sets of VMT data, one in HPMS and the other in TSN. HPMS data is developed for federal reporting purposes, and TSN data is used for calculation of safety-related issues such as accident rates on state highways. Because of the different business rules for each system (i.e. highway segmentation and reporting deadlines are different in each system) the reporting of VMT from each system is different, which can create confusion and errors for those using the data. This can be particularly problematic in cases where different responses to the same question are received by external customers.</p> <p>NOTE: This issue was the subject of the “VMT Reporting Committee” made up of members of the Division of Transportation System Information and the Division of Traffic Operations. Please see “Vehicle Miles Traveled Data Sources Fact Sheet, June 2010” in Appendix N.</p>
2.	<p><i>Inconsistent, Duplicate Data Reporting by State Agencies</i></p> <ul style="list-style-type: none"> • Potential inconsistency of VMT data reported from various state agencies. 	<p>The California Air Resources Board (CARB) and California Energy Commission also report VMT. To the degree that these agencies report VMT that is different than those reported by Caltrans, this could create confusion and errors for those using the data. It may also give rise to criticism of the state from the public, the media, control agencies, and the federal government who expect consistent information from state agencies.</p>
3.	<p><i>Inconsistent, Duplicate Data Reporting by Local Agencies</i></p> <ul style="list-style-type: none"> • Local agency data reported to HPMS is problematic due to issues of completeness, accuracy, and timeliness of data coming from local agencies. • Local agency VMT data and calculation methodologies may be inconsistent with state agencies. 	<p>VMT data is reported by local agencies for state and federal (HPMS) reporting purposes. Often, the local data is incomplete, inaccurate, and/or out-of-date. This is a reflection of inadequate resources at the local agencies for performing the data collection needed to calculate VMT. The FHWA mandates that Caltrans collect and report local agency VMT data. Therefore, Caltrans is accountable for this data and will be required to help resolve issues with local agency data quality and reporting.</p>
4.	<p><i>Inconsistent/Inaccurate VMT Calculations</i></p> <ul style="list-style-type: none"> • Inconsistency and/or inaccuracy in current VMT calculations can have negative impacts on future VMT calculations in Caltrans travel models. • VMT data accuracy is partially dependent on the resolution of other data quality issues involved with AADT, timely completion of the Traffic Census, and reducing the backlog of highway inventory updates in TSN. 	<p>Travel Models such as the California Statewide Travel Demand Model and MVSTAFF use current VMT calculations to help determine future VMT and to compare with the VMT calculations of other agencies to ensure accuracy and consistency. MVSTAFF customer expectations include “reasonable, consistent estimates and forecasts” of VMT and other data.</p> <p>Current VMT calculations are derived from AADT data and can only be as accurate and up to date as the AADT data used in VMT calculations. Calculations of updated AADT data are impacted by delays in completing traffic counts and delays in updates to the highway inventory. As such, VMT accuracy is dependent in part on addressing issues with updates to the highway inventory in TSN and the timely completion of the Traffic Census.</p>

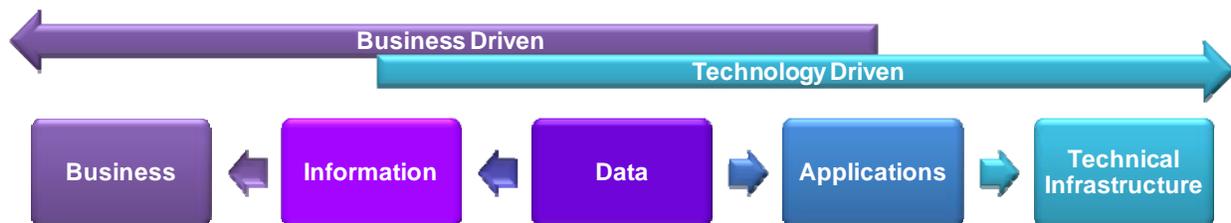
Chapter 4 Data Architecture

4.1 Data Architecture

4.1.1 Data Architecture Defined

An organization's needs are constantly changing as business strategies change and the business environment evolves. In order to provide useful information to users in response to all of these changes, well-defined data architecture must exist. Data Architecture is "a discipline, process, and program focusing on integrating sets of information."¹ As shown in the "Data Architecture Environment" diagram below, data is the foundation for both business and technology driven activities. Data governance at Caltrans primarily focuses on Communities of Interest; however, sound data architecture should consider Caltrans' technology environment as well.

Figure 13: Data Architecture Environment



As shown above, there are five components that comprise an organization's data architecture, including:

1. Business includes business rules, business requirements, organization structure, success factors, business process design and modeling, and/or the mission and vision of the organization.
2. Information includes data integration, data management, metadata management, data architecture, data delivery, data modeling, data quality, dashboards, business intelligence, reporting, performance management, and/or content management.
3. Data includes output from systems, data content, and may involve a collection of numbers, graphs, objects, letters, and/or images.
4. Applications include rudimentary workbooks with macros, database tools such as Access or FileMaker solutions, corporate custom-developed solutions, or COTS (Commercial Off-The-Shelf) solutions.
5. Technical Infrastructure includes desktops, middleware, servers, networks, operating systems, database infrastructure, security, storage and/or other hardware.

This chapter is largely focused on a particular type of information, metadata, or 'data about data'. Metadata related to data elements and data sets will support both the business and technical communities within Caltrans. The remainder of this section describes metadata, its importance, and relevance to data architecture and Caltrans, its relationship to data sets and data elements, and its use to Communities of Interest.

4.1.2 What is Metadata?

"Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data or information about information."² Metadata is fundamental to data architecture and is the focus of the recommendations provided in this chapter.

¹ Data Governance Institute, <http://www.datagovernance.com/glossary-governance.html>

Although metadata comes in a variety of different forms, there are three types of metadata:

1. Descriptive: This type of metadata is the resource used to search for and locate an object, such as the title, author, keywords, or publisher.
2. Structural: This type of metadata describes how the components of the objects are organized, such as tables, columns, or indexes.
3. Administrative: This type of metadata contains the technical information to manage a resource, such as file type, date it was created, or user accessibility.

As previously stated, metadata is “data about data,” but it also has an important purpose, which is to discover relevant information between different data sources. Additional methods in which metadata can assist in organizing and locating information are outlined below.

1. Resource Discovery: Metadata can help locate resources by:
 - a. Allowing resources to be found by relevant criteria;
 - b. Identifying resources;
 - c. Bringing similar resources together;
 - d. Distinguishing dissimilar resources; and
 - e. Giving location information.
2. Organize Electronic Resources: Electronic resources are largely organized by audience or topic to assist in finding the information faster. Metadata can be organized by creating:
 - a. Static WebPages: This is where the names and locations of the resources are “hardcoded” in the HTML.
 - b. Dynamic WebPages: This is where the web page is built directly from the metadata databases.
3. Interoperability: Metadata is understood by both humans and machines which can promote interoperability and allows multiple systems, different hardware/software, data structures and interfaces to communicate successfully.
4. Digital Identification: Most metadata is defined using unique data elements (e.g. standard numbers) that correspond to the data being identified. Sometimes metadata is defined using the location of the original data; however, the issue here is that sometimes the location or file name of the data changes, causing the metadata to be invalid. Metadata can also be combined to become a “data set” which also has a different set of data elements, making it unique.
5. Archiving and Preservation: Metadata is fundamental to reducing the risk of losing important digital information. Sometimes digital information can be compromised by users altering information, or it can become corrupted without knowledge to the user. Another issue is that software and hardware technologies sometimes change making the data inaccessible. Metadata helps to eliminate these issues by archiving and preserving the information to ensure it will continue to be accessible and reliable in the future. Archiving and preserving metadata requires unique elements to “track the lineage of a digital object, detail its physical characteristics, and to document its behavior” to ensure it will be compatible with future technologies.

Quality issues may arise if metadata is created by information gatherers that are not familiar with how to define metadata or if metadata is created automatically. The National Information Standards Organization (NISO) defines good metadata in “The Framework of Guidance for Building Good Digital Collections.” Six principles applying to good metadata are defined below:

1. Good metadata should be appropriate to the materials in the collection, users of the collection, and the intended, current and likely use of the digital object.
2. Good metadata supports interoperability.
3. Good metadata uses standard controlled vocabularies to reflect the what, where, when and who of the content.
4. Good metadata includes a clear statement on the conditions and terms of use for the digital object.
5. Good metadata records are objects themselves and therefore should have the qualities of archivability, persistence, and unique identification. Good metadata should be authoritative and verifiable.
6. Good metadata supports the long-term management of objects in collections.

The metadata for data elements and data sets described later in this document are designed to support quality metadata as outlined above.

4.1.3 Data Architecture at Caltrans

As noted earlier, metadata is fundamental to the recommendations provided in this chapter. Metadata standards are a key component of data architecture and this section describes how the proposed metadata fits within the data architecture at Caltrans.

In the past, Caltrans has defined data architecture using the Zachman Framework. The Zachman Framework employs a matrix based on six communication questions (What, Where, When, Why, Who and How) which are referenced in the below table. The framework is not a methodology, but rather a taxonomy for organizing data, linking it to "who" the data impacts, targets, and the issues being addressed. Caltrans has also developed corporate data models in the past using the Oracle Custom Development Methodology (CDM). These are both acceptable methods in defining a corporate data architecture methodology. *These methods are summarized in this document to demonstrate that the proposed Data Element Catalog and Data Set Catalog metadata is consistent with and supports these Caltrans methods.*

There many different layers of the organization that relate to the Zachman Framework. This framework includes six columns (or communication questions) and five rows (or layers). The layers include:³

1. Scope (or Contextual Row): The scope describes the context, constraints, architecture and range of the data framework that is important to the business. This section states what the framework is supposed to do as it is seen by the user. The scope doesn't describe the inner workings of the framework. **The Board operates primarily at this level by providing business guidance regarding issues, needs, constraints, and business solutions.**
2. Business Model (or Conceptual Row): The business model defines the goals, strategies, processes and structure of the framework that are used to support the mission and vision of the organization as it relates to the framework. **The Board provides guidance regarding business or conceptual models. The TAC ensures business and conceptual models are technically sound.**
3. System Model (or Logical Row): The system model states how the system should perform its functions. It has system activities, objects, functions and requirements that are used in implementing the business model defined above. The system model describes the inner workings of the framework. **The TAC vets activities, objects, functions and requirements that are used in implementing business and conceptual models and summarizes their findings in a format suitable for the Board.**

³ Journal of Enterprise Architecture, 2006

4. Technology Model (or Physical Row): The technology model lists the constraints that may be felt by tools, technology, materials, and even the constraints of people. **The Caltrans Information Technology (IT) stakeholders operate at this layer and provide technical details to the TAC.**
5. Detailed Representation (or Out-of-Context Row): The detailed representation contains data definitions that may sometimes be too detailed, depending on the audience. **The Caltrans IT stakeholders operate at this layer and provide technical details to the TAC. These technical details, in the form of data sets and data elements, are also of interest to Business Users.**

Table 12: Zachman Framework²

Name of Framework	What (data)	How (function)	Where (network)	Who (people)	When (time)	Why (motivation)
1. Scope (contextual)	<i>Of interest to the Board, Caltrans Management, and the TAC</i>					
2. Business Model (conceptual)	<i>Of interest to the Board and the TAC</i>					
3. System Model (logical)	<i>Of interest to the TAC</i>					
4. Technology Model (physical)	<i>Of interest to IT Stakeholders</i>					
5. Detailed Representation (out-of-context)	<i>Of interest to IT Stakeholders and Business Users</i>					

For purposes of this project, we are focusing on the first (What – Data) and fourth (Who – People) columns. These columns describe the:

- What (or data): The “what” is the data that describes the entities, attributes, and subject areas that are important to different perspectives within Caltrans. Examples of ‘what’ may include counties, routes, pavement statistics, and bridge health index, etc.
- Who (or people): The “who” represents the people responsible for assessing capabilities and performance within Caltrans. These people are also responsible for assessing the duties, responsibilities and authorities of others within the organization. This also relates to relationships between work performed by people, and the systems that may interface with a person’s duties. Examples may include: Data Stewards, Communities of Interest and specific categories of users like planners, owners, designers, builders, or subcontractors.

When defining the ‘what’ through formal data modeling efforts, data modelers must break data down to the atomic level and then build it back up. Generally speaking, the atomic level is a data element (or data field) and data elements are logically grouped or built back up into entities (and eventually tables). The Data Architect is the person responsible for doing this and (for the purposes of this project) uses two traditional architectural processes as described below:

- Conceptual: Represents all business entities and may include employees, customers, products, equipment, etc.
- Logical: Represents the logic of how entities are related and may include the business name, relationships between entities, unique key for an entity, and list of attributes.

- Physical: the realization of the data tables and their relationships to support a specific type of functionality.

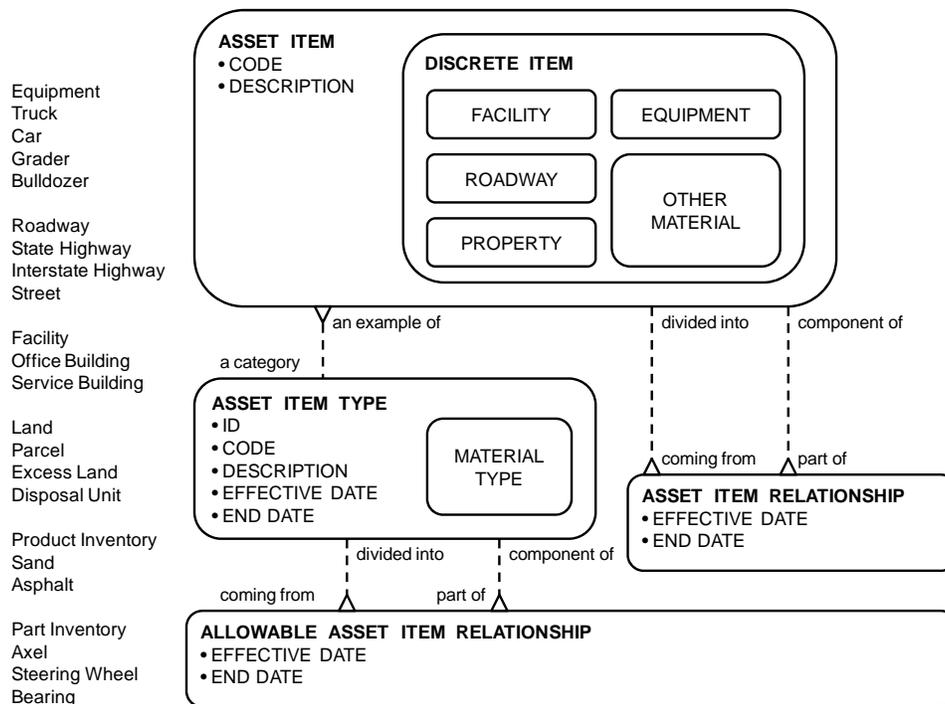
As noted earlier, Caltrans has experience developing conceptual, logical and physical data models using the Oracle Custom Development Methodology (CDM). There are many ways that Conceptual and Logical models can be valuable to assist in creating a data architecture framework. The Physical data models are omitted from these discussions and are generally outside the scope of the Board or TAC. These benefits are discussed below.⁴

Conceptual Models (high-level models), of interest to the Board and TAC:

1. Put the immediate focus on the data. All parties involved are able to discuss the target environments so everyone is on the same page.
2. Resolve organizational conflicts and political issues early in the process before becoming a major issue by mapping out Business Ownership and identifying where it may overlap.
3. Allow a manageable view of the data involved and allocates appropriate resources in advance.
4. Gives the opportunity to group certain subject areas and allocate the correct resources to each subject area.

The figure below represents a sample conceptual data model for “Assets” such as Facilities, Roadway, Property, Equipment, Part Inventory, etc. Conceptual data models contain entities (the boxes – ASSET ITEM, FACILITY, and COST) and attributes (the items in the boxes below – ID, CODE, DESCRIPTION, etc.). The conceptual and logical models look quite similar at first. However, the logical model requires substantially more effort to develop, contains more entities, and all attributes have been defined.

Figure 14: Sample Conceptual Data Model for “Assets”



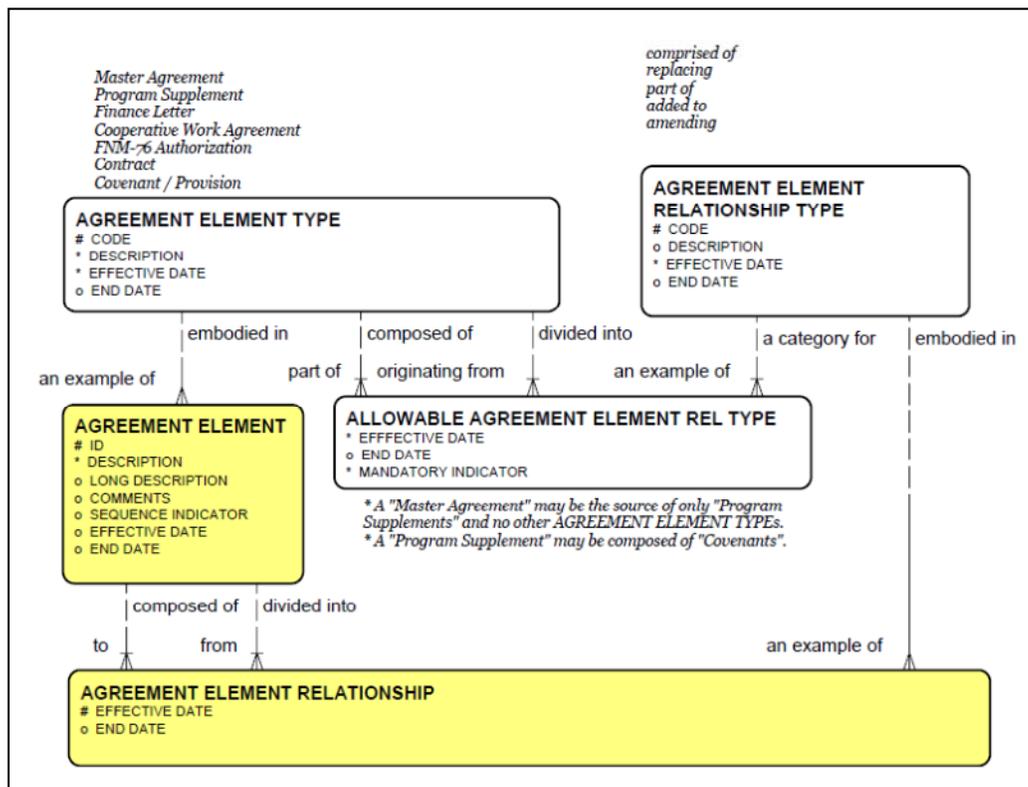
⁴ Data Migration Pro, “8 data migration benefits of logical and conceptual data models.” 2008

Logical Models (lower-level models), of interest primarily to the TAC:

1. Quickly identifies data gaps (when compared to the high-level model) that may require further research.
2. Allow for a more easily and relevant focused business process analysis that provides a deep understanding of the target environment.
3. Gain a maximum advantage by defining the design and build areas of the project that identifies which data elements are critical and which items can be delayed to reach the target environment.
4. Help align the target design and build areas without severe delays and risk that is jointly agreed upon by the parties involved.

The next figure below shows a sample logical data model for “Agreements” such as Master Agreements, Program Supplements, Finance Letters, Contracts, etc. Logical data models contain entities (the boxes – AGREEMENT ELEMENT) and attributes (the items in the boxes below – Description, Long Description, Comments, etc.). In a physical data model, the entities become database tables and the attributes become database columns or fields. The next sections outline a business approach to capture conceptual, entity and attribute information using a Data Element Catalog (Section 4.2) and Data Set Catalog (Section 4.3).

Figure 15: Sample Logical Data Model for “Agreements”



4.1.4 Data Products and Data Architecture

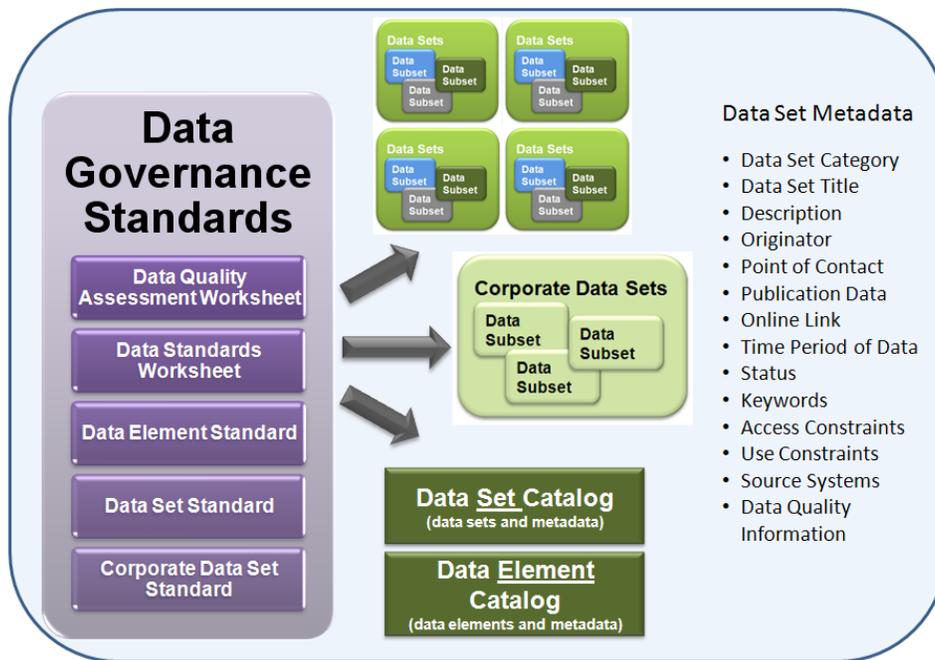
For Caltrans’ data governance purposes, data products are defined from a business perspective and not from an IT or technical perspective. Data products may exist as raw data in a database table, data in an Excel Workbook, data in a corporate computer system, or data contained in a report. In the prior section, we presented a data architecture overview and ended by describing the relationships between conceptual, logical and physical data models. The table below shows the additional relationships to two data products: Data Element Catalog and Data Set Catalog.

Table 13: Two Data Products – Data Set Catalog and Data Element Catalog

Data Products	Conceptual	Logical	Physical
Data Set Catalog – a collection of data from one or more tables or entities. Relationships between data are limited to key words and general categories. Provides data sets of interest to Caltrans as a whole or to specific Communities of Interest.	Entities – contains most, shows logical relationship between entities	Entities – contains all, shows logical relationship between entities	Tables
Data Element Catalog – a collection of data elements or attributes that may be from one or more tables or entities. Provides data elements of interest to Caltrans as a whole or to specific Communities of Interest.	Attributes – primary	Attributes – all	Columns

As displayed in the figure below, data governance and data governance standards lead to recognized data sets, higher priority corporate data sets, a Data Set Catalog and a Data Element Catalog that are defined using metadata.

Figure 16: Data Governance Standards, Data Catalogs and Metadata



The next two sections describe the Data Set Catalog and the Data Element Catalog and their associated metadata.

4.2 Data Element Catalog

Caltrans uses thousands of transportation system data elements. Each of these data elements has (or should have) established characteristics or standards (metadata) that are defined so that users understand the data element and can use, update, and/or report the data properly. Data governance will include the establishment of a Data Element Catalog which will include those Caltrans transportation

system data elements for which standards and metadata have been formally established and recognized. Contained in this Data Element Catalog will be a listing of metadata for each data element.

4.2.1 Data Element Metadata

The following are the metadata categories that will be contained in the Data Element Catalog and sample values for the data element 'County' (for illustration purposes only).

Table 14: Data Element Metadata

#	Data Element Metadata	Definition	Sample Values
1.	Long Name	The full title of the data element.	<i>California County</i>
2.	Short Name	The abbreviated title of the data element.	<i>County</i>
3.	Description	A brief description of the data element.	<i>One of the 58 California jurisdictions defined as a county per x, y, z.</i>
4.	Format	The format of the data element (e.g. text, integer, data, currency)	<i>Text</i>
5.	Length	The number of spaces required for the data element, including decimal places if applicable.	<i>30</i>
6.	Valid Values	The values which are valid for this data element (e.g. list of valid county codes, values greater than zero).	<i>Amador, Butte, and continued list of remaining 56 counties...</i>
7.	Business Rules or Specifications	Specific rules or specifications for creating valid values or using this data element.	<i>Not applicable</i>
8.	Data Sets	The data set(s) in which this data element is included.	<i>Not applicable</i>
9.	Keywords	Keywords which users of the data element can search to find information related to the data element.	<i>County Jurisdiction</i>
10.	Source Systems	Systems containing the data element.	<i>TSN</i>
11.	Publication Date	The last date the data element metadata was updated.	<i>June 30, 1991</i>
12.	Status	The current status of the data element, including problems or issues with the data.	<i>This list of counties as developed in May of 1991 and has remained unchanged since that time.</i>
13.	Originator	The Caltrans organization that owns the data element.	<i>Division of Transportation System Information</i>
14.	Point of Contact	The person to contact for information or questions regarding the data element.	<i>To be determined</i>
15.	Metadata Date	Date the metadata of the data element was last updated or changed.	<i>Unchanged since May 12, 2011 posting</i>
16.	Metadata Contact	The contact person for questions or information regarding data element	<i>Coco Briseno, Chief, Division of Transportation</i>

#	Data Element Metadata	Definition	Sample Values
		metadata. (May or may not be same person as Point of Contact).	System Information

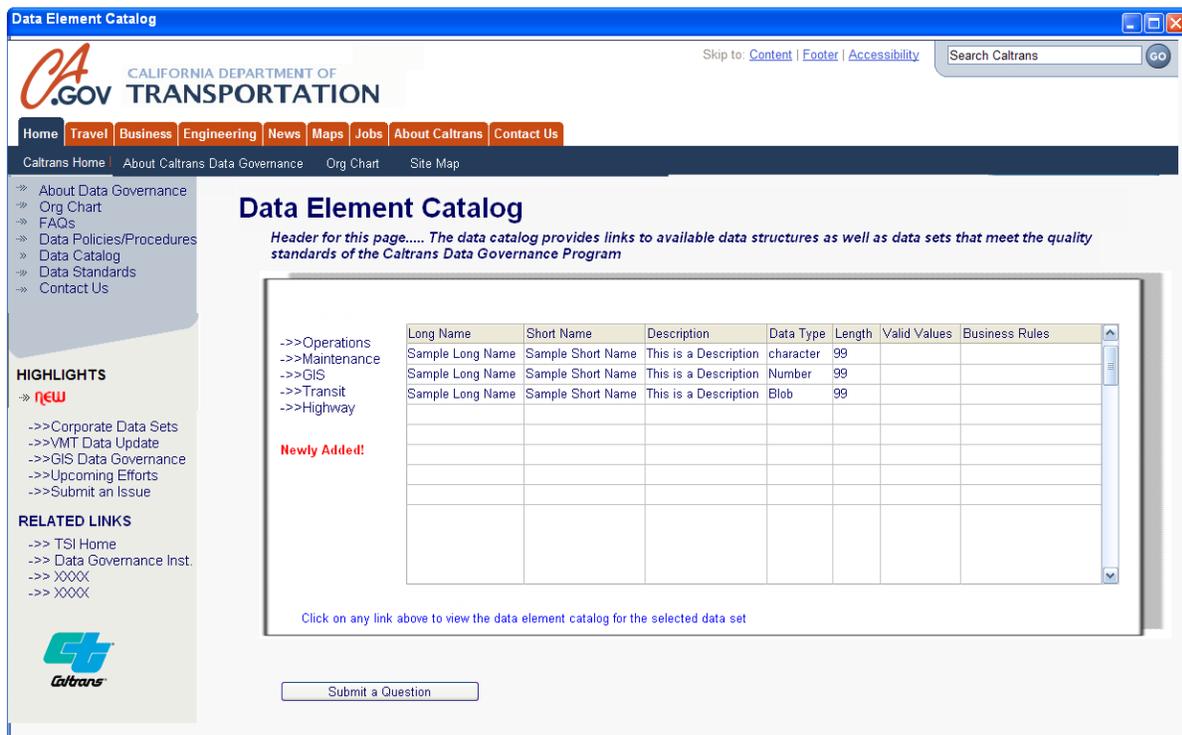
Appendix K, “Data Element Analysis Worksheet” contains some key questions to guide the analysis of a data element for needed data standard development. It is used to identify issues, develop recommendations, and state implementation factors, costs, and potential return on investment.

Appendix H, “Data Element Metadata Form” contains a worksheet for the use of those establishing and/or updating data element metadata.

4.2.2 Data Element Catalog

As data elements undergo data governance processes, the data element metadata will be established, and the data element will be included in the Data Element Catalog. The Caltrans data governance website will be updated whenever new data elements are added to the Data Element Catalog, along with their metadata description, or when metadata for an existing data element is modified. Below are examples of webpage mock-ups for the Data Element Catalog, showing the data elements included in the catalog, and a drill down to the metadata for one or more data elements.

Figure 17: Data Element Catalog Web Page



During the process of applying data governance to a data element, the data element metadata will be analyzed and established and/or updated, as needed (using the “Data Element Analysis Worksheet” in Appendix K if necessary). The Board will direct the TAC to use the Data Element Metadata Form” in Appendix H to define the data element metadata. Once the metadata is complete, this information will be entered or updated in the catalog on the website for reference by users and others who may need the metadata information.

4.3 Data Set Catalog

At Caltrans, a data set is a group of related data elements that are typically gathered in one or a few related data tables and are often used together in analysis and producing information. Data governance at Caltrans will focus on “corporate data sets,” which are data sets that span multiple business functions within the organization and therefore have an increased importance across the corporate structure of Caltrans.

Data governance will include the establishment of a Data Set Catalog which will include Caltrans corporate data sets (as well as other data sets, in order of priority). Contained in this Data Set Catalog will be a listing of the metadata for each data set.

4.3.1 Data Set Metadata

The following are the metadata categories that will be contained in the Data Set Catalog:

Table 15: Data Set Metadata

#	Data Set Metadata	Definition
1.	Data Set Category	The broad category the data set belongs to. <i>(Data Set Categories will be defined through a future project.)</i>
2.	Data Set Title	The full title of the data set
3.	Description	Description of the data set
4.	Originator	The Caltrans organization that owns the data set
5.	Point of Contact	The person to contact for information or questions regarding the data set
6.	Publication Data	The last date the data set metadata was updated
7.	Online Link	The hyperlink to the data set source
8.	Time Period of Data	The time period covered by the data set
9.	Status	The current status of the data set, including problems or issues with the data
10.	Keywords	Keywords which users of the data set can search to find information related to the data set
11.	Access Constraints	Constraints to user access to the data set
12.	Use Constraints	Constraints to the use of the data from the data set
13.	Source Systems	Systems containing the data set
14.	Data Quality Information	Specific data quality issues related to the data set (i.e. duplicate data, incomplete data, inaccurate data, etc.)
15.	Primary Data Elements	The most important and/or widely used data elements within the data set
16.	Metadata Date	Date the metadata of the data set was last updated or changed
17.	Metadata Contact	The contact person for questions or information regarding data set metadata. Note: May or may not be the same person as the “Point of Contact” defined above in number 5.

Please see Appendix I, “Corporate Data Set Metadata Form,” which is for the use of those establishing and/or updating data set metadata.

4.3.2 Data Set Metadata – Sample Content from Three Corporate Data Sets

As was explained in detail in a prior chapter (“Data Products and Data Assessment”), Caltrans will initiate data governance efforts for three corporate data sets: County/Route/Postmile - Linear Reference System (LRS), Annual Average Daily Traffic (AADT), and Vehicle Miles Traveled (VMT). Below is a table which includes the metadata for each of these corporate data sets:

Table 16: Metadata for Three Corporate Data Sets

#	Metadata	County/Route/PM – LRS	AADT	VMT
1.	Data Set Category	<ul style="list-style-type: none"> Location 	<ul style="list-style-type: none"> Traffic Volume Congestion Collisions 	<ul style="list-style-type: none"> Traffic Volume Congestion
2.	Title	County/Route/Postmile - LRS	Annual Average Daily Traffic	Vehicle Miles Traveled
3.	Description	County/Route/Postmile data represents the foundation of Caltrans’ Linear Referencing (LRS) System and provides essential location data along the state highway system. This data is used in many systems but the primary system of record for County/Route/Postmile is TSN. Caltrans also built a GIS-based LRS tool that links County/Route/Postmile location data to geospatial coordinate data.	AADT data is derived from formulas applied to traffic census data that measures the level of traffic on the state highway at individual locations for specific dates and times. Traffic census data is gathered either through PeMS detector systems built into the travel way, or through other traditional traffic census methods. AADT calculations are contained in TSN (formerly TASAS) and are used for various operations, planning, and design purposes throughout Caltrans.	VMT data is derived from AADT data and measures the total number of miles traveled by vehicles, over a specific length of highway, for a specific time period. VMT is currently reported in both TSN and the Highway Performance Monitoring System (HPMS), although the VMT amounts in each system differ because of the use of different highway segments in calculating VMT.
4.	Originator	TSI	Traffic Operations (Traffic Census)	Traffic Operations (Traffic Census) and TSI
5.	Point of Contact	Mark Samuelson, TSI	Nick Compin, Traffic Operations	Mark Samuelson, TSI
6.	Publication Date	(enter last date the data set metadata was updated)	(enter last date the data set metadata was updated)	(enter last date the data set metadata was updated)
7.	Online Link	<i>(To be developed)</i>	<i>(To be developed)</i>	<i>(To be developed)</i>
8.	Time Period of Data	Updated continuously.	Traffic census is collected annually from October 1 through September 30. AADT is calculated from the annual traffic census.	Vehicle Miles Traveled is updated monthly in the “Trend” Report.

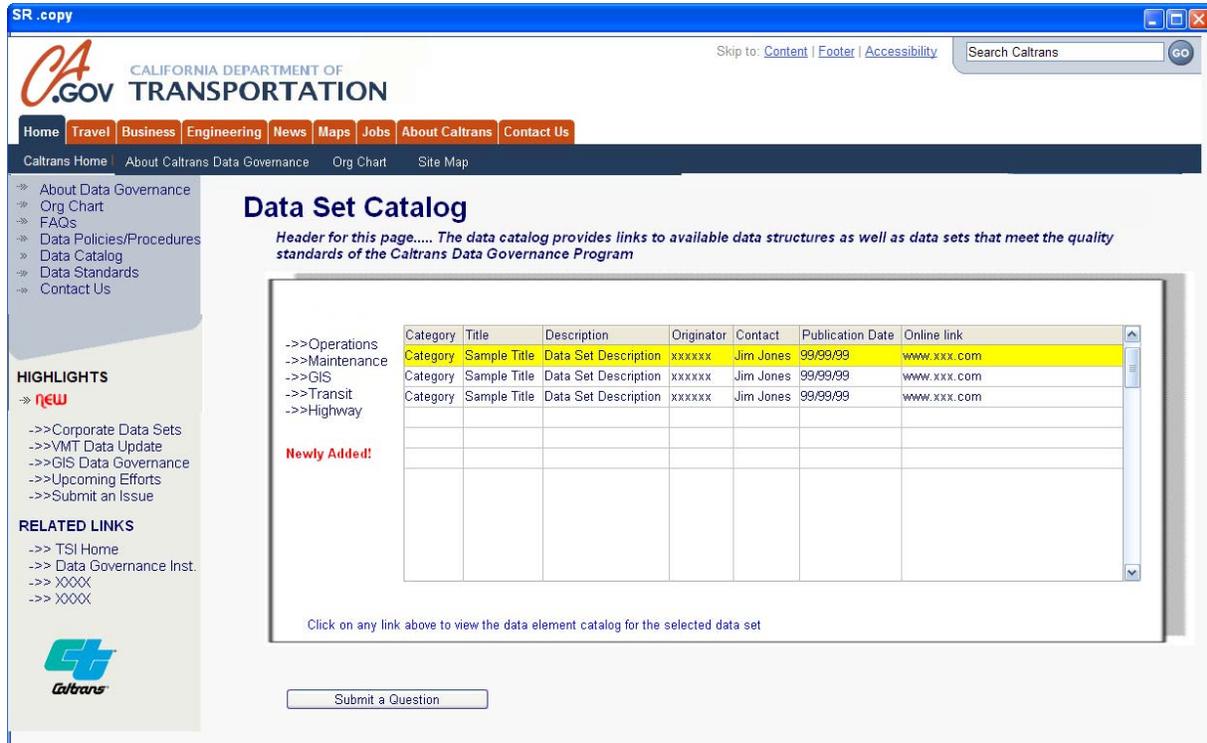
#	Metadata	County/Route/PM – LRS	AADT	VMT
9.	Status	Backlogged. Updates occur as information is available.	Backlogged. Delays in Traffic Census create delays in AADT. Updates of AADT in TSN also dependent on updates to C-R-PM data, which is backlogged.	VMT is duplicated in Caltrans in TSN and in HPMS because of differing highway segmentation in the two systems.
10.	Keywords	<ul style="list-style-type: none"> Location Linear reference Odometer 	<ul style="list-style-type: none"> Traffic Volume Congestion Peak 	<ul style="list-style-type: none"> Traffic Volume Congestion Mileage
11.	Access Constraints	This data is public domain and available to share with Caltrans partners.	This data is public domain and available to share with Caltrans partners.	This data is public domain and available to share with Caltrans partners.
12.	Use Constraints	This data is public domain and available to share with Caltrans partners.	This data is public domain and available to share with Caltrans partners.	This data is public domain and available to share with Caltrans partners.
13.	Source Systems	<ul style="list-style-type: none"> TSN HPMS State Roadway Network 	<ul style="list-style-type: none"> TSN PeMS 	<ul style="list-style-type: none"> TSN HPMS
14.	Data Quality Information	<ul style="list-style-type: none"> Duplicate data Incomplete data Untimely updates Poor data quality assurance 	<ul style="list-style-type: none"> Duplicate data Incomplete data 	<ul style="list-style-type: none"> Duplicate data Inaccurate data
15.	Primary Data Elements	<ul style="list-style-type: none"> District County code Route Number (incl. any suffix) Postmile (incl. any prefix or suffix) 	<ul style="list-style-type: none"> District County/Route/PM Peak Hour Peak Month Direction of travel 	<ul style="list-style-type: none"> District County/Route/PM AADT Route Segment Length
16.	Metadata Date	<i>(To be determined)</i>	<i>(To be determined)</i>	<i>(To be determined)</i>
17.	Metadata Contact	<i>(To be developed)</i>	<i>(To be developed)</i>	<i>(To be developed)</i>

4.3.3 Data Set Catalog

As these and other corporate data sets undergo data governance processes, the data set metadata will be established, and the data set will be included in the Data Set Catalog. The Caltrans data governance website will be updated whenever new data sets are added to the Data Set Catalog, along with their metadata description, or when metadata for an existing data set is modified. Below is an example of a

webpage mock-up for the Data Set Catalog, showing the data sets included in the catalog and the metadata categories associated with each data set.

Figure 18: Data Set Catalog Web Page



During the process of applying data governance to a data set, the data set metadata will be analyzed and established and/or updated, as needed. Using the Corporate Data Set Metadata Form in Appendix I, the data set metadata will be established and entered or updated in the catalog on the website for reference by users and others who may need the metadata information.

Chapter 5 Data Governance Program Management and Implementation Plan

5.1 Data Governance, a Program Management Perspective

Caltrans corporate data governance requires a program perspective in order to achieve goals and objectives that extend over a multiyear time horizon. A program is a group of related projects managed in a coordinated manner to obtain benefits that are not available from managing them individually. Programs may include elements of related work outside of the scope of the discrete projects in the program. Projects have finite timelines, whereas programs are conducted on an ongoing basis. At the highest level, program management involves the following core functions:

- **Program Management** is to ensure that the larger-scale initiatives are being managed properly, and that the relationships between projects are understood and managed; and
- **Project Management** is to ensure that the individual initiatives are being executed according to plan and methodology.

The “Data Governance Program Management and Implementation Plan” is based on best practices for both program management and project management. In order for the Board to be successful, it must oversee ongoing actions such as conducting regular board meetings; it must direct small scale projects such as analyzing a single data set to determine issues and solutions; and it must support larger projects such as developing a Data Governance Website complete with a Data Element Catalog and a Data Set Catalog.

Caltrans’ Data Governance Program will be a long-term, ongoing effort that is devoted to achieving goals and objectives as shown in *Table 2: Goals and Objectives* in Chapter 1 at the beginning of this document.

Table 17: Caltrans Data Governance Program Goals and Objectives

#	Goal Title	Goal Description	Objectives
1.	Leadership	Champion data solutions, ensure accountability, and increase the value of data assets	<ul style="list-style-type: none"> • Promote data governance within Caltrans • Communicate data-related changes to all Communities of Interest • Monitor progress and ensure accountability of data governance tasks and projects
2.	Quality	Oversee efforts to provide high quality data that is accurate, clear, and easy to access	<ul style="list-style-type: none"> • Establish a data quality assurance program • Increase the accuracy and clarity of data • Improve accessibility of data
3.	Prioritization	Prioritize efforts to address data gaps and needs	<ul style="list-style-type: none"> • Establish clear priorities to address data gaps and needs • Communicate priorities to Caltrans business units
4.	Cooperation	Facilitate cross-organizational collaboration, data sharing, and integration (break down barriers between business units, reduce data silos)	<ul style="list-style-type: none"> • Increase opportunities for data sharing • Eliminate data silos and other barriers • Ensure business units know the identity of Business Owners/Data Custodians of data elements • Ensure Business Owners know the identity of their Communities of Interest

#	Goal Title	Goal Description	Objectives
5.	Flexibility	Encourage creative and innovative solutions to data needs	<ul style="list-style-type: none"> Identify innovative data solutions throughout Caltrans Communicate innovative solutions to Business Owners and Communities of Interest

The Board will need to weigh trade-offs and consider issues in the context of broader Caltrans' objectives as it pursues the Data Governance Program goals. Caltrans priorities may influence which issues become priorities that the Board identifies for the Data Governance Program. This will be particularly important as the Board develops priorities and allocates the resources available to data governance.

Like many new initiatives, the difference between the success and failure of data governance at Caltrans will be the ability of the organization to understand, accept, and promote data governance and the changes needed to transform the new initiative into a mature program. The change management plan described in the next section (5.2 – *Data Governance Change Management Plan*) will provide Caltrans the opportunity to embrace the changes necessary to make data governance a success. This chapter identifies the actions and activities to promote and advance data governance at Caltrans through change management practices. The phrase "change management" has a variety of definitions. In this context, change management is an approach to effecting change, both from the perspective of an organization and on the individual level. Change management means defining and implementing tools, processes, training, and communication to accomplish changes in the business environment in order to achieve program goals.

This chapter also outlines an overarching, multi-year program management plan that includes the development of projects and action plans to implement the Data Governance Program and achieve the program's goals and objectives. Data governance projects will consist of larger-scale, longer-term efforts to 'stand up' the data governance organizational structure and infrastructure. A "Project Summary Form" is provided in Appendix D for the use of those developing data governance projects. In addition, a variety of "Action Plans" will also be developed, which are smaller scale action steps to support the development of the Data Governance Program.

This chapter also contains tasks and action steps for eight projects in Section 5.3 (*Data Governance Program of Projects*) below. Appendices O - S contain completed project summaries for the three data governance pilot projects (County-Route-Postmile, Annual Average Daily Traffic, and Vehicle Miles Traveled), the creation of the TAC, and the creation of the Data Governance Website.

5.2 Data Governance Change Management Plan

Change management activities are fundamental to sustaining a Data Governance Program over time. The purpose of the Data Governance Change/Communication Plan component of the Data Business Plan is to identify a series of change management activities necessary to keep Communities of Interest comfortable with the Data Governance Program and ongoing changes.

The table below provides high level issues the Data Governance Program Change/Communication Plan will address during the first year. Addressing these issues will launch the Data Governance Program and sustain the program through the resulting changes. Additional issues will be identified as the Data Governance Program rolls out and challenges regarding acceptance of data governance become evident.

Table 18: Key Change Issues to be Addressed

#	Change Issue
1.	Caltrans is unaware of the new the data governance effort.

2.	Communities of Interest have difficulty finding important data, data standards, and information about data quality and validity (i.e. Information about data is found instead through informal communication channels).
3.	Communities of Interest must understand how to use the tools, processes, and organizational structure supporting data governance, in order to effectively participate in the new program.
4.	Communities of Interest, managers, and executives are used to making data decisions to support business needs and processes specific to a single business area or need (i.e. there may be resistance to the Board’s decisions which best meet the strategic needs of Caltrans vs. the operational needs of a division or office).

The table on the following page presents *Sample Data Governance Change/Communication Plan Content* that will support and sustain the Data Governance Program (a more complete table can be found in Appendix T). These activities are organized in the following framework for understanding change.

5.2.1 Framework for Understanding Change

- Awareness of the need for change – activities in this group are designed to promote overall awareness of the Data Governance Program, both initially and an ongoing basis (for new staff, Business Owners, or Communities of Interest). These activities are primarily targeted at addressing issue #1 above.
- Desire and Motivation to support and participate in the change – activities in this group are designed to encourage participation in and support for the Data Governance Program through illustrating management support, and periodically communicating the value (results and benefits) the Data Governance Program provides. These activities are primarily targeted at addressing issue #3 above.
- Knowledge of how to change and the ability to implement required skills and behaviors – activities in this group are designed to educate Caltrans Communities of Interest on how to access and use the tools and results of the Data Governance Program. These activities are primarily targeted at addressing issues #2 and #3 above.
- Reinforcement to sustain the change – activities in this group are designed to reinforce the use of Data Governance Program decisions, tools, and standards to help avert the tendency to revert back to old processes and ways of doing business. These activities are primarily targeted at addressing issue #4 above.

Although each activity falls into a single framework category in the plan, it is important to recognize that these activities frequently support multiple categories. Implementation of these activities must be coordinated by the Board, and could be led by a TAC member on behalf of the Board. The table below represents a sample of the activities designed to increase ‘Awareness’ of the new data governance structure and processes. See Appendix T for the complete communication plan table that outlines how to increase ‘Desire and Motivation’ to improve data governance; how to provide ‘Knowledge and Ability’ to implement data governance; and ‘Reinforcement’ communications to help ensure Caltrans continues to support and improve data governance.

Table 19: Sample of Data Governance Change Management/Communication Plan Content

Communication or Activity	Audience	Medium	Owner
Introduce Data Governance Program	Executive Team	Presentation at Director's Meeting	Board Chair Person
Data Governance Accomplishments Highlights	Caltrans General (All Staff)	Director's Video Caltrans Magazine	Director/Exec Staff Public Info Staff
Data Governance Website	<ul style="list-style-type: none"> • Business Owners • Communities of Interest • Data Custodians • IT System Developers • Caltrans Staff 	Develop and Implement a data governance website with access via On-Ramp	TAC or TSI

5.3 Data Governance Website Concept Overview

This section presents a website concept that is designed to function as a critical tool to support both ongoing communications about data governance progress, as well as, serve as Caltrans' central portal for all data governance products. The website is a key component of the Data Governance Change Management/Communication Plan presented in the previous section of this document. The website concept is organized in seven web pages with a common sidebar on each page. The web pages are summarized below, along with the change management components they support and a mock-up of what each screen could look like. See Appendix S for the Data Governance Website Mockups and Project Summary.

Table 20: Summarized Website Concept

#	Web Page Title	Purpose	Supports Change Component			
			Awareness	Desire	Knowledge	Reinforcement
1	About Data Governance	Home page for Data Governance – provides information on the overall mission and goals for the Data Governance Program.	X			
2	Data Governance Organization Chart	Provides the Data Governance organization structure and identifies current members.	X			
3	Frequently Asked Questions	Provides answers to common data governance questions that Communities of Interest may ask about and a mechanism to submit new questions.			X	
4	Data Governance Policies and Procedures	Provides easy access to data governance policies developed through the Data Governance Program. Provides easy access to forms, tools, and/or	X		X	

#	Web Page Title	Purpose	Supports Change Component			
		procedures developed to support data governance.				
5	Data Standards	Provides a central location for all standards developed, accepted and/or approved by the Data Governance Program.			X	
6	Data Set Catalog	Provides access to data formats, and potentially, data sets that have been approved for enterprise access by the Data Governance Program.			X	
7	Corporate Data	Provides a working definition of “corporate data” within the context of the Data Governance Program. Provides links to project updates regarding active data governance efforts for those data sets.	X	X	X	
8	Data Element Catalog	Provides access to data element formats that have been approved for enterprise access by the Data Governance Program.			X	
9	Sidebar on all pages	Provides standard links to all key pages within the website, as well as, new highlights to keep Communities of Interest updated on progress, in which the Board wants to publish.		X		X

In summary, this section presents a website concept that supports the data governance effort through providing a location for anyone in Caltrans to find the latest information about data governance accomplishments, products announcements, and data. When implemented, this website becomes the central location for data governance information and addresses key Caltrans data issues noted previously.

The lack of immediately available resources is a risk to the overall data governance effort at Caltrans, although the risk is mitigated by reducing the scope and size of the Data Governance Program. Resources are not the only risk facing the Board. The Board will likely face a variety of risks as it begins operations, and will need a process to manage those risks so they do not undermine the data governance effort. The following section (5.4 – *Risk Management Strategies and Controls*) describes the implementation of risk management to avoid or mitigate risks that arise.

5.4 Risk Management Strategies and Controls

The risk management process includes the development and continuous maintenance of a Risk Management Plan in accordance with standard project management methodology. Risk management is a critical component of project management used to address potential issues on the project before they significantly impact the overall project direction. All projects and activities must take into consideration potential risks and how to mitigate and respond to risks as they arise. In order to prepare for likely scenarios, the Board should approve a Risk Management Plan that documents the process and procedures that will be used to manage risks: identify the people responsible for managing various areas of risk, how risks will be tracked throughout the life cycle, how contingency plans will be implemented, and how resources will be allocated to handle risks.

5.4.1 Risks Identified to Date

During the process of establishing the data governance project, some preliminary risks to the success of a data governance effort were noted that may become part of the initial risk management effort of the Board. These risks are shown in the table below:

Table 21: Potential Data Governance Risks

#	Potential Risk	Response
1.	Lack of resources	Reduce scope/size of data governance effort
2.	Resistance to change from organization	Develop and implement an effective Change Management Plan
3.	Lack of knowledge of data governance in Caltrans	Develop and implement an effective Communication Plan

Other risks will be identified as Caltrans implements and operates data governance. The risks will be recorded and tracked to facilitate the management and mitigation of each risk.

5.4.2 Responsible Parties

Risk management is a dynamic process that occurs throughout any activity life cycle. Therefore, several parties will be responsible for developing and implementing the Risk Management Plan. The Board Chairperson will have primary responsibility for managing the data governance risk management process. The specific roles of parties to the risk management process are described below:

- Caltrans Directorate — The Caltrans Directorate, as leader of the activities of the Board, will address critical risks that have been elevated from the Board.
- Board Chairperson — The Board’s Chairperson will support the Caltrans Board in managing risk. The Chairperson will oversee development of the Risk Management Plan and will work with the Board to ensure that the Board is informed of and acts on risks. The Chairperson will monitor ongoing risks, develop mitigation measures and contingency plans, and implement the contingency plans as necessary.
- Board Members — Members of the Board will be involved in identifying potential risks and will work with the Board Chairperson to carry out mitigation actions and/or contingency plans.
- Data Custodians and/or the TAC – Data custodians and/or the TAC will be responsible for identifying risks and logging them for the attention of the Board. These groups will also implement any risk mitigation decisions made by the Board.

5.4.3 Risk Management Process

Controlling risk is supported by a risk management process that is tailored to the specific needs of the Data Business Plan. This process will continue as part of ongoing planning, until dissolution of the Board. The procedure for consistent and ongoing evaluation of risk items and development of risk strategies includes the following process components:

- Identify the risk
- Evaluate the risk
- Define a mitigation and contingency strategy
- Document the risk status, priority, severity, mitigation strategy, and responsible party
- Communicate new high severity risks to the Board, Caltrans Directorate, and other appropriate parties, as necessary

In documenting ongoing risks, the Board uses standard tools such as a Risk Management Worksheet.

5.4.4 Risk Management Worksheet

To document and control risk, a Risk Management Worksheet may be utilized. The Risk Management Worksheet is a display of the identified risks and the key attributes or characteristics of each. See Appendix L for a blank Risk Management Worksheet. The information to be included in the Risk Management Worksheet is shown in the table below:

Table 22: Risk Management Worksheet

Risk	Probability	Impact	Affected Area	Warning Signs	Preventive Measures
<A concise statement of the risk and its consequence>	<Likelihood of a risk occurring>	<Impact of risk if not avoided or mitigated>	<The areas affected should the risk occur. May include: Budget, Schedule, Resources, etc.>	<Indications that a risk exists>	<The mitigations and/or contingency actions to be taken to reduce the probability of the risk occurring and/or the severity of the risk should it occur.>

5.4.5 Phases of Risk Management

Phase 1: Risk Assessment

Risk assessment involves identifying, analyzing, quantifying, and prioritizing risks. Data Custodians and/or the TAC will identify risks and log them for the attention of the Board. The Board and the other responsible parties involved in the risk management process will regularly review new risk assessments as well as ongoing risk efforts to:

- Evaluate and determine the risk exposure and severity
- Identify appropriate action to avoid or mitigate the risk
- When appropriate, elevate the risk assessment and response to the Caltrans Directorate

Phase 2: Risk Response

Risks are managed during the risk response phase, including actions such as avoidance, acceptance, mitigation, and sharing. The Board should consider a variety of factors during risk response, including resources available, and the time needed to respond. The following are standard responses to risks:

- Risk Avoidance: Risk avoidance is very effective if it can be employed. It involves reducing risk by developing an alternate approach or eliminating the risk completely.
- Risk Acceptance: Risk acceptance is the result of a determination that the risk is unavoidable and that the consequences will have to be accepted.
- Risk Mitigation: Risk mitigation is employed when a risk cannot be avoided but does not have to be accepted. It involves mitigating the impact of the risk occurrence through the development of a strategy to address aspects of the risk to reduce its impact.
- Risk Sharing: Risk sharing can be employed where others are able to share in the risk and where that sharing is in the best interests of all parties. It typically is associated with negotiated tradeoffs to gain the acceptance of all parties.

Phase 3: Risk Tracking and Control

Risk tracking and control maintains the status of identified risks and the actions that have been taken to address those risks. Risk tracking will provide accurate and timely information to the Board during risk management implementation and help prevent risks from adversely affecting the activities of the Board.

Risks should be logged in the Risk Management Worksheet (Appendix L) for the duration of the activities of the Board. The project team should also establish and maintain metrics for measuring performance and progress toward resolving risks. Performance metrics may include, for example:

- Number of risks which are open, completed, incomplete, or past due
- Severity of risks which are open, completed, incomplete, or past due
- Age of risks which are open or incomplete
- Number and severity of risks which are escalated to the Caltrans Directorate

Information about risks logged in the worksheet and their performance metrics may be reported through regular status reports. Status reports will identify tasks accomplished, outstanding issues/risks to be resolved, new issues to be tracked, and the next steps. Additionally, the top ten risks will be identified with special attention called to the unresolved items that could impact critical data governance efforts. Particular emphasis will be paid to providing accurate and timely information to the Board to enable risk management and help prevent risks from adversely affecting the activities.

5.5 Data Governance Program of Projects

Data governance is still a new concept for most of Caltrans. And for those who are familiar with data governance and have participated in the initial efforts to establish data governance at Caltrans, much still must be done to clarify data governance and establish a strong foundation for data governance in Caltrans. The list below is a specific program of projects and actions which must be developed and completed to ensure the establishment and on-going maintenance and operation of the Data Governance Program. These projects are described at a summary level and do not contain an identification of those parties that will be responsible for forming the project teams and implementing the projects. As shown in the Data Governance Processes above, the identification of potential project team members and the establishment of project teams is the joint responsibility of the Board and the TAC (Project Sponsors and Stakeholders for some of these projects can be found in the Project Summary Forms in Appendices O through S).

Program of Projects

1. Establish and Maintain Data Governance Organizational Structure and Operations
2. Introduce Data Governance to Caltrans and Support Ongoing Communications
3. Develop and Maintain Corporate Data Standards
4. Oversee Data Governance Projects
5. Develop Data Governance Website
6. Conduct Data Governance Pilot Projects
 - a. County/Route/Postmile – Linear Referencing System (LRS)
 - b. Annual Average Daily Traffic (AADT)
 - c. Vehicle Miles Traveled (VMT)
7. Develop Data Set Catalog
8. Develop Data Element Catalog

This section is devoted to projects 1 through 4 that initiate data governance oversight and communications. Section 5.6 (*Near-term Government Projects*) is devoted to projects 5 through 8, setting up a website to support Caltrans data governance, conducting Pilot Projects, and developing the Data Set Catalog and the Data Element Catalog.

In the following sections, each of these projects will be described in more detail to state what the project is and what the primary actions steps will be to accomplish the project and further the establishment of data governance at Caltrans. Schedules for these actions and projects and presented in Section 5.8 (*Implementation Schedule*).

Project 1: Establish and Maintain Data Governance Organizational Structure and Operations

The Caltrans Data Governance Program has been under development during the course of the project to develop the Data Business Plan. The concept of establishing the groundwork for the program in parallel with developing the Data Business Plan has provided the opportunity to proceed with an Implementation Plan that is farther advanced than if the Data Business Plan had been developed in isolation.

To date, an Interim Data Governance Board has been established (from the Steering Committee overseeing the development of the Data Business Plan) that has made important initial decisions regarding the initial structure of data governance, including its mission and goals, a charter for the Board, and the selection of initial pilot projects related to important corporate data sets. This chapter builds upon what the Interim Board has already established and prepared for a long-term, ongoing organizational structure and operational elements for Caltrans’ data governance. The table below outlines the elements and action steps for this project:

Table 23: Tasks and Action Steps for Project 1

#	Task	Primary Action Steps
1.	Establish and Maintain Post-Interim Data Governance Board	<ol style="list-style-type: none"> 1. Select Board Members, Ensure Positions Filled (Chair, Headquarters’ Membership, and District Membership) 2. Conduct Kick Off and Quarterly Meetings 3. Execute Deputy Directive recognizing Board role and responsibilities
2.	Support Data Governance Board	<ol style="list-style-type: none"> 1. Plan, Schedule, Facilitate, and Follow Up Communications for Quarterly Meetings 2. Maintain Charter and revise as needed per Board direction 3. Determine new membership as needed
3.	Establish the TAC	<ol style="list-style-type: none"> 1. Create initial TAC <ol style="list-style-type: none"> a. Develop Charter (goals, objectives, membership) b. Select participants, identify Chairperson c. Conduct Kick Off and Regular Meetings
4.	Support the TAC	<ol style="list-style-type: none"> 1. Plan, Schedule, Facilitate and Follow Up Communications for Regular Meetings 2. Maintain Charter and revise as per TAC or Board direction 3. Determine new TAC membership as needed
5.	Determine Staffing / Resource Needs to Support Data Governance	<ol style="list-style-type: none"> 1. Prepare estimate of staffing/resource needs for Board and TAC 2. Present to Board and adjust according to Board direction 3. Monitor staff/resource usage over one year 4. Review at year end – prepare upcoming year’s estimate

#	Task	Primary Action Steps
6.	Review Data Governance Structure, Determine if Aligned with Business Needs	<ol style="list-style-type: none"> 1. Review data governance structure (following completion of first year's activities) 2. Determine needed changes to existing structure (if any) 3. Determine need for expansion (if any) 4. Determine resources available for change/expansion (if any) 5. Board decision to implement structural changes 6. Update Board or TAC charters to reflect changes

Project 2: Introduce Data Governance to Caltrans and Support Ongoing Communications

An integral part of the Data Business Plan is the Change Management/Communication Plan that outlines the strategy for introducing the Data Governance Program to Caltrans and for maintaining ongoing communication and updates.

This project, described in the table below, is intended to introduce and establish important elements of change management and communication for data governance at Caltrans. It contains important elements such as identifying leadership for the communication effort, implementing efforts to introduce data governance to Caltrans, building a website to help manage ongoing communications and updates of key data governance components, and a process for providing ongoing data governance status updates to Caltrans.

Table 24: Tasks and Action Steps for Project 2

#	Task	Primary Action Steps
1.	Identify Communication Lead	<ol style="list-style-type: none"> 1. Board defines Communication Lead roles and responsibilities 2. Board nominates and confirms Communication Lead for data governance
2.	Introduce Data Governance Program	<ol style="list-style-type: none"> 1. Analyze and enhance Communication Plan based on available resources and near-term data governance needs 2. Identify staff to provide data governance training to Caltrans executives and staff 3. Develop training program and materials to be delivered to audiences (executives and staff) 4. Educate executive team, and gain buy-in, convey business objectives and expected results first year 5. Educate staff, inform, convey business objectives and expected results first year 6. Stakeholder Training - Provide an overview of the goals, objectives, roles, responsibilities and initiatives of the Data Governance Program (for those with new data related responsibilities)
3.	Communicate Data Governance Accomplishments and Updates	<ol style="list-style-type: none"> 1. Record and provide updates of recent accomplishments and new initiatives of the Data Governance Program 2. Provide directions to access general data governance information for staff unfamiliar with Data Governance Program
4.	Provide Data Set, Data Element Catalog, and Website Training	<ol style="list-style-type: none"> 1. Develop training for data governance processes, procedures, and tools available through the Data Governance Website 2. Provide training for documented processes, procedures and tools on the Data Governance Website

#	Task	Primary Action Steps
5.	Post Frequently Asked Questions (FAQs) to Website	<ol style="list-style-type: none"> 1. Identify and develop an extensive list of “Frequently Asked Questions” (FAQs) about general data governance questions, the Caltrans Data Governance Program, and the Data Governance Website 2. Publish FAQs on Data Governance Web Site

Project 3: Develop and Maintain Corporate Data Standards

A variety of documents, forms, and tools have been created during the course of the Data Business Plan project. Over the next months and years, these documents, forms and tools will likely be refined to better serve Caltrans’ changing data governance needs. Caltrans must ensure that all changes are executed in controlled manner and that only the latest versions are available to support effective data governance. This project will ensure that these Data Business Plan documents, forms, and tools are established and maintained to provide the information and functionality required to help guide the successful implementation of data governance at Caltrans.

Table 25: Tasks and Action Steps for Project 3

#	Task	Primary Action Steps
1.	Maintain Data Governance Standards	<ol style="list-style-type: none"> 1. Finalize and maintain data governance standards, including: <ol style="list-style-type: none"> a. Board Charter b. Communication Plan c. Process diagrams and narratives d. Data governance forms (Data Issue Form, Project Prioritization Form, Project Summary Form, Quarterly Project Progress Form) e. Data Governance Website content (catalogs, organization, policies/procedures, corporate data, FAQ, etc.) f. Data element and data set metadata g. Data Quality Assessment tool
2.	Oversee maintenance of Data Set Catalog and associated metadata	<ol style="list-style-type: none"> 1. Finalize metadata for data set under review 2. Post new metadata to website 3. Review of new metadata by Business Owners or Communities of Interest 4. Address comments from Business Owners or Communities of Interest 5. Provide support and updates as needed
3.	Oversee maintenance of Data Element Catalog and associated metadata	<ol style="list-style-type: none"> 1. Finalize metadata for data element under review 2. Post new metadata to website 3. Review of new metadata by Business Owners or Communities of Interest 4. Address comments from Business Owners or Communities of Interest 5. Provide support and updates as needed

Project 4: Oversee Data Governance Projects

In Chapter 2, “Process to Identify Data Business Needs,” five primary data governance processes have been developed to help guide the core functions of data governance at Caltrans. The diagram below outlines these five processes:

Figure 19: Five Primary Data Governance Processes



In the table below, each of the five processes is related to the essential questions that each process addresses:

Table 26: Data Governance Processes and Related Questions

#	Process	Data Governance Questions
1.	Submitting data governance issues for consideration	<ul style="list-style-type: none"> How does Caltrans staff formally raise a data issue to the Board? What information is needed? Who is the issue submitted to?
2.	Filtering issues prior to Board consideration	<ul style="list-style-type: none"> Who prepares the issues for Board review? Does the Board hear all issues, or issues on any subject?
3.	Prioritizing data governance issues for action	<ul style="list-style-type: none"> How does the Board decide what issues will be addressed? Who else participates in the prioritization process? What happens to issues that do not reach the Board?
4.	Monitoring progress on existing tasks and projects	<ul style="list-style-type: none"> How will the Board be kept apprised of efforts to solve data governance issues? What progress has been made on Board recommended initiatives? Who assists the Board in monitoring and evaluating projects? How are project risks and issues being managed?
5.	Establishing and implementing new data governance policies	<ul style="list-style-type: none"> How does the Board create new policies? How do policies get communicated to appropriate parties? What is the Board's role in implementation? Who develops policy details and implements policies?

5.6 Near-term Data Governance Projects

This section is devoted to projects 5 through 8, setting up a website to support Caltrans data governance, conducting Pilot Projects, and developing the Data Set Catalog and the Data Element Catalog

1. Develop Data Governance Website
2. Conduct Data Governance Pilot Projects
 - a. C-R-PM LRS
 - b. AADT
 - c. VMT
3. Develop Data Set Catalog
4. Develop Data Element Catalog

These projects produce some of the key data governance data products that will be needed by the Data Governance Program and by Communities of Interest as the program matures over the next few years.

Project 5: Develop Data Governance Website

As shown above in Project 2, regarding the initial implementation of the Communications Plan, a Data Governance Website is a valuable tool for providing information and training to Caltrans regarding the Data Governance Program. For more information on this project, see the Project Summary Form in Appendix D, for a complete project description, risks and mitigation factors, project benefits, project sponsor and other key project positions, project milestones, and, where possible, an estimated cost to implement.

Table 27: Tasks and Action Steps for Project 5

#	Tasks	Primary Action Steps
1.	Identify staff to develop and maintain the Data Governance Website	<ol style="list-style-type: none"> 1. Identify services and staff needs for website development 2. Request services from appropriate source (i.e. Division of Information Technology)
2.	Develop and implement a Data Governance Website with access via the On-Ramp	<ol style="list-style-type: none"> 1. Design web page 2. Develop content 3. Test (system and user) 4. Go live and make available to Caltrans
3.	Provide general website maintenance	<ol style="list-style-type: none"> 1. Identify staff to provide ongoing maintenance 2. Receive and respond to requests from Business Owners and Communities of Interest

Project 6: Conduct Data Governance Pilot Projects

The Board has selected three corporate data sets to be the subject of initial data governance pilot projects. The pilot Project Summary Forms can be found in Appendices O, P, and Q. A description of the three pilot corporate data sets is shown below:

- **County/Route/Postmile – Linear Referencing System (LRS)** (Appendix P)
County/Route/Postmile data represents the foundation of Caltrans’ LRS and provides essential location data along the state highway system. This data is used in many systems but the primary system of record for County/Route/Postmile is TSN (formerly TASAS). Caltrans also built a GIS-based LRS tool that links County/Route/Postmile location data to geospatial coordinate data.
- **Annual Average Daily Traffic (AADT)** (Appendix Q)
AADT data is derived from formulas applied to traffic census data that measures the level of traffic on the state highway at individual locations for specific dates and times. Traffic census data is gathered either through detector systems built into the travel way, or through other traditional traffic census methods. AADT calculations are contained in TSN (formerly TASAS) and are used for various operations, planning, and design purposes throughout Caltrans.
- **Vehicle Miles Traveled (VMT)** (Appendix O)
VMT data is derived from AADT data and measures the total number of miles traveled by vehicles, over a specific length of highway, for a specific time period. VMT is currently reported in both TSN and the Highway Performance Monitoring System (HPMS), although the VMT amounts in each system differ because of the use of different highway segments in calculating VMT.

A preliminary assessment of each of these data sets has been completed and is included in Chapter 3, “Data Products and Data Assessment.” These assessments were based on a limited number of interviews with Subject Matter Experts (SMEs) for each of the data sets and resulted in the identification of a variety of data issues within each data set. Each pilot project should address some of the data issues identified by the preliminary assessments. The number of issues addressed and the overall project scope will be determined by the resources available to the project team in conducting the pilot project.

The table below outlines the common elements and action steps for each of the three pilot data governance projects.

Table 28: Tasks and Action Steps for Project 6

#	Tasks	Primary Action Steps
1.	Create Project Team	<ol style="list-style-type: none"> 1. Identify business areas, Data Custodians and Communities of Interest related to the project 2. Develop a list of potential project team members 3. Establish a project team and select a team leader 4. Identify resources (i.e. staff time) available for the project
2.	Develop project summary	<ol style="list-style-type: none"> 1. Leverage Initial Data Set Assessment from the Data Products and Data Assessment chapter <ol style="list-style-type: none"> a. Choose data issues to address in the project (scope must be achievable within resources identified above) 2. Determine if a Data Quality Assessment needs to be completed 3. Determine whether data standards will be developed for the data elements within the data set 4. Develop a project summary using the Project Summary Form (See Appendix D)
3.	Present Project Proposal to TAC/Board	<ol style="list-style-type: none"> 1. Present the project at the next TAC and Board meetings 2. Revise the project per changes required by the TAC and/or Board
4.	Execute Project	<ol style="list-style-type: none"> 1. Identify SMEs to interview related to data issues 2. Conduct SME interviews 3. Conduct additional research, as needed 4. Complete Data Quality Assessment Worksheet (Appendix J), if necessary 5. Develop Quarterly Project Progress Form (Appendix F) and submit to the TAC and the Board 6. Present project status at TAC and Board meetings, as necessary
5.	Develop Metadata	<ol style="list-style-type: none"> 1. Define Data Set(s) and associated metadata 2. Define supporting data elements and associated metadata, if necessary 3. Update Data Element and Data Set Catalogs with new metadata
6.	Provide Recommendations	<ol style="list-style-type: none"> 1. Develop draft recommendations related to data issues 2. Present to TAC and Board for approval 3. Incorporate TAC and Board changes into final recommendation 4. Develop new data governance policies, if applicable
7.	Communicate Project Outcome	<ol style="list-style-type: none"> 1. Announce decisions regarding the data set to related Data Custodians, Business Owners, and Communities of Interest 2. Distribute Data Governance Policy Form (See Appendix G), if new policies have been developed 3. Post project results and new policies on the Data Governance Website
8.	Close Out Project	<ol style="list-style-type: none"> 1. Develop a final project report 2. Determine a final project cost (staff hours, and any operating expenditures)

See Appendices O, P, and Q for complete project descriptions, risks and mitigation factors, project benefits, project sponsors and other key project positions, project milestones, and, where possible, an estimated implementation cost.

Project 7: Develop Data Set Catalog

The Data Set Catalog is a listing of all data sets (beginning with corporate data sets) for which metadata has been analyzed and established. There are many data sets at Caltrans (including dozens that may be considered corporate data sets) in various IT systems, databases, spreadsheets, and other sources. A preliminary listing of corporate data sets has already been developed and is located in Appendix M. This listing will need to be analyzed and refined to include any additional data sets that qualify as corporate data sets. The projects summarized above (Project 6) will also provide information for at least three corporate data sets. Defining data sets and establishing related metadata will ensure that related Communities of Interest know what each data set is, where to find it, and how it should be used. The Data Set Catalog will be a central repository available to all Caltrans staff, increasing the availability and appropriate use of data.

Table 29: Tasks and Action Steps for Project 7

#	Tasks	Primary Action Steps
1.	Finalize draft list of corporate data sets	<ol style="list-style-type: none"> 1. Review the existing list of corporate data sets 2. Add data sets to the list, if applicable 3. Remove data sets from the list, if applicable
2.	Define metadata information for corporate data sets and load into Data Set Catalog on website	<ol style="list-style-type: none"> 1. Define metadata for each corporate data set 2. Enter metadata information into the Data Set Catalog 3. Ask Communities of Interest to review information on the website for accuracy and usability
3.	Load additional Data Set metadata into Data Set Catalog on website	<ol style="list-style-type: none"> 1. Enter the data set metadata from the data governance projects into the Data Set Catalog 2. Enter the data set metadata received from other sources into the Data Set Catalog
4.	Maintain Data Set Catalog	<ol style="list-style-type: none"> 1. Ensure that the Data Set Catalog is available to Communities of Interest 2. Implement data changes to correct any inaccuracies identified by Communities of Interest

Project 8: Develop Data Element Catalog

The Data Element Catalog is a listing of all data elements for which metadata has been analyzed and established. There are thousands of transportation system data elements at Caltrans in various IT systems, databases, spreadsheets, and other sources. Defining data elements and establishing related metadata will ensure that Communities of Interest know what each data element is, where to find it, and how it should be used. The Data Element Catalog will be a central repository available to all Caltrans staff, increasing the availability and appropriate use of data. The data catalog serves as a good starting point for Caltrans when identifying who owns transportation system data at Caltrans, where it is located, and how it is defined. This data element catalog is intended to be a living document that is updated and expanded as Caltrans’ transportation system data needs continue to evolve.

Caltrans currently has a preliminary data catalog, the TSI Data Catalog 20100324, containing over 1,700 data elements along with the data element definition, owner, source systems, and other useful information. The development of the Data Element Catalog will ideally leverage the existing information in the TSI Data Catalog 20100324 to populate many of the metadata fields for data governance. The Data Element Catalog will be available on the Data Governance Website.

The figure below shows a screen-shot of the existing TSI Data Catalog and the variety of information already stored in this catalog.

Figure 20: TSI Data Catalog

The screenshot displays the 'DataElement - Transportation System Data Inventory and Catalog' application. The interface includes a ribbon with tabs for 'Home', 'Create', 'External Data', 'Database Tools', and 'Add-Ins'. The main workspace is divided into several sections:

- Navigation Pane:** Located on the left, it shows the current record ID '231'.
- Form Fields:**
 - Name:** Emission Process
 - Description:** Emission Process: (B) Blowby, (G) Partial day resting losses, (K) Hot soak, (L) Partial day diurnal breathing, (R) Running losses, (S) Starting emissions, (M) Multiple day diurnal breathing, (N) Multiple day resting losses.
 - Mode:** Environmental
 - Category 1:** Air Quality
 - Category 2:** Emissions Rates
 - Category 3:** N/A
 - Category 4:** N/A
 - Data Steward:** EAD
 - Data Steward (long):** Caltrans Division of Environmental Analysis
 - Contact Name:** Gregg Erickson
 - Telephone:** (916) 654-6296
 - Email:** gregg_erickson@dot.ca.gov
 - Reports that use Data Element:** None
 - Systems that use Data Element:** EMFAC
 - Users of Data Element:** DTIM
- Footer:** Shows 'Record: 223 of 1776' and a search bar.

The TSI Data Catalog database was developed using MS Access 2003, allowing users of Access to be able to navigate the database easily. To keep the information centrally managed and consistent, data in the database is centrally administered and published by a Data Catalog Administrator, who oversees additions, changes, and deletions to data elements or their characteristics. The catalog contains 1,777 data elements, which span the transportation modes, as shown in the table below:

Table 30: Number of Data Elements, by Transportation Mode

Mode	Number of Data Elements	Percent of Data Elements
Highway Mode	1157	65.1%
Transit Mode	271	15.3%
Aeronautics Mode	213	12.0%
Rail Mode	78	4.4%
Environmental	42	2.4%
Bike/Walk Mode	13	0.7%
Intermodal/Interregional	3	0.2%
Grand Total	1777	100.0%

The data elements related to each mode include the following data characteristics:

- Data element name
- Description of the data element
- Data Steward/Custodian of the data element (a summary of the list of Data Stewards/Custodians, their short names, long names, and contact information, is included in the database)
- Users of the data element
- Systems used to store the data
- Reports on which the data element is used

These and other efforts demonstrate that there are many sources of existing information that will be beneficial to data governance at Caltrans. These sources should be identified and exploited by the Board to assist in addressing data issues without repeating work that may have already been completed in another forum. The ability of Board to capitalize on prior efforts and advance data governance at Caltrans will be restrained by the availability of resources.

Table 31: Tasks and Action Steps for Project 8

#	Tasks	Primary Action Steps
1.	Analyze Existing TSI Data Catalog	<ol style="list-style-type: none"> 1. Review the data elements in the TSI Data Catalog <ol style="list-style-type: none"> a. Confirm the accuracy of data element information 2. Identify data elements from the TSI Data Catalog to be loaded into the Data Element Catalog on the Data Governance Website
2.	Load Information from TSI Data Catalog into Data Element Catalog on website	<ol style="list-style-type: none"> 1. Transfer data element information from the TSI Data Catalog to the Data Element Catalog on the Data Governance Website 2. Ask Communities of Interest to review information on the website for accuracy and usability
3.	Load additional Data Elements and associated metadata into Data Element Catalog on website	<ol style="list-style-type: none"> 1. Enter the data element metadata from the data governance projects into the Data Element Catalog 2. Enter the data element metadata received from other sources into the Data Element Catalog
4.	Maintain Data Element Catalog	<ol style="list-style-type: none"> 1. Ensure that the Data Element Catalog is available to Communities of Interest 2. Implement data changes to correct any inaccuracies identified by Communities of Interest

The Data Element Catalog and Data Set Catalog projects outlined above describe the requirements for projects that result in two catalogs' metadata and webpage concepts. Both catalogs should start relatively small and grow over time to include additional data elements and data sets of sufficient importance to Caltrans. The Board and the TAC will guide the growth of these two catalogs.

The next section provides a detailed perspective of the implementation schedule for the projects discussed above. The schedule is presented in two forms: a Gantt chart and a PERT chart.

5.7 Implementation Schedule

This section includes two types of diagrams to support documenting and tracking schedules for data governance actions and projects outlined in the prior two sections. The first diagram, the Gantt chart, displays tasks or projects and related timelines to allow the user to see durations and relative timing. The second diagram, the PERT chart, displays tasks or project dependencies and provides a more tactical perspective of 'what has to happen by when.' Both of these methods are commonly used by project managers when managing a project or group of projects.

5.7.1 Gantt Chart

The Gantt chart is a graphical representation of a project schedule and is helpful when monitoring a single project's progress or planning and scheduling multiple projects. This chart displays the start dates and end dates of each task throughout the life of each project. The timeline may differ depending on the length of the project, how many tasks are involved, and the breadth of detail the chart must represent. A Gantt chart is primarily useful for initial planning and then illustrating if the project is behind or ahead of schedule.

The Gantt chart represented on the next page is a display of the actions and projects associated with the Caltrans' Data Governance Program. It contains a total of eight projects, with a span of three years, broken down by four quarters in each year.

Figure 21: Data Governance Project Gantt Chart

Projects	Year 1				Year 2				Year 3			
	FY 2011/2012				FY 2012/2013				FY 2013/2014			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
1) Establish and Maintain Data Governance Organizational Structure and Operations												
a. Establish Data Governance Board												
b. Support Data Governance Board												
c. Establish Technical Advisory Committee												
d. Support Technical Advisory Committee												
e. Determine Staffing / Resource Needs to Support Data Governance												
f. Review Data Governance Structure, Determine if Aligned with Business Needs												
2) Introduce Data Governance to Caltrans and Support Ongoing Communications												
a. Identify Communication Lead												
b. Introduce Data Governance Program												
c. Communicate Data Governance Accomplishments and Updates												
d. Provide Website Data Set, Data Element Catalog, and Website Training												
e. Create and Update Website Frequently Asked Questions												
3) Develop and Maintain Enterprise Data Standards												
a. Maintain Data Governance Standards												
b. Oversee maintenance of Data Set Catalog & associated metadata												
c. Oversee maintenance of Data Element Catalog & associated metadata												
4) Oversee Data Governance Projects												
5) Develop Data Governance Website												
a. Identify staff to develop and maintain the data governance website												
b. Develop and implement a data governance website												
c. Provide general website maintenance												
6.1) Data Governance Pilot Project - County/Route/Postmile												
a. Create Project Team												
b. Develop project summary												
c. Present project proposal to TAC/Board												
d. Execute project												
e. Develop metadata												
f. Provide recommendations												
g. Communicate project outcome												
h. Close out project												
6.2) Data Governance Pilot Project - Annual Average Daily Traffic												
a. Create Project Team												
b. Develop project summary												
c. Present project proposal to TAC/Board												
d. Execute project												
e. Develop metadata												
f. Provide recommendations												
g. Communicate project outcome												
h. Close out project												
6.3) Data Governance Pilot Project - Vehicle Miles Traveled												
a. Create Project Team												
b. Develop project summary												
c. Present project proposal to TAC/Board												
d. Execute project												
e. Develop metadata												
f. Provide recommendations												
g. Communicate project outcome												
h. Close out project												
7) Develop Data Set Catalog												
a. Finalize draft list of Corporate Data Sets												
b. Define metadata for Corporate Data Sets & load into catalog												
c. Load additional data set metadata into Data Set Catalog												
d. Maintain Data Set Catalog												
8) Develop Data Element Catalog												
a. Analyze and Update Preliminary Data Catalog												
b. Load info from Preliminary Data Catalog into Data Element Catalog												
c. Load additional data elements & metadata into Data Element Catalog												
d. Maintain Data Element Catalog												

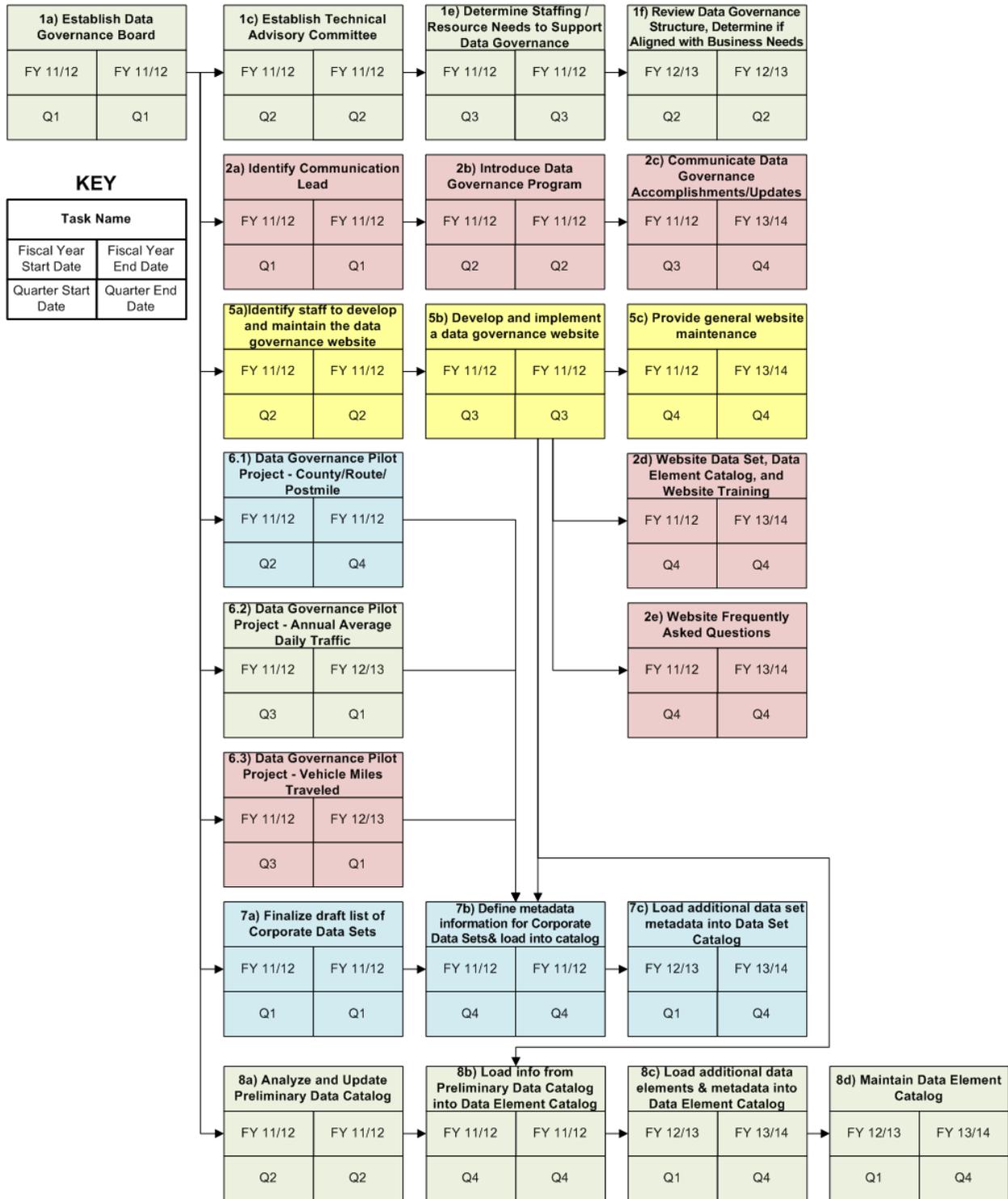
5.7.2 PERT Chart

A Program Evaluation Review Technique (PERT) chart is presented in the form of a network diagram that contains boxes that represent tasks (or milestones) in the project. These boxes are linked by lines with arrows to the next task in the sequence. The primary assumption in developing a PERT Chart is that no activity can begin until its immediately preceding task is completed. The PERT chart is also used to identify the critical path for a project or program of projects.

The PERT chart is beneficial because it provides a timeframe for the program of projects and the “critical path” of projects that impact the program completion. This critical path displays dependencies and relationships between projects and tasks and also identifies opportunities for moving resources from one task to another. The PERT chart below is designed to assist the Board in managing the initial data governance projects. Projects #3 and #4 are not on the critical path and therefore are excluded from the PERT chart.

This PERT chart highlights critical dependencies, shows the critical path for initial data governance efforts, and provides an overall timeframe. It can be updated as circumstances dictate over time.

Figure 22: Data Governance Program PERT chart



5.8 Project Summary Form

This section contains an overview of a template that will be used for the development of data governance projects. The Project Summary Form (Appendix D) will assist in the development of the near-term data governance projects described in the prior section.

Project planning and project management is essential to the successful delivery of complex projects. Studies have shown that poor planning and management (among other reasons) have been significant causes of project failure and project cost overruns. Proper program and project management will be essential for data governance to succeed at Caltrans. The Project Summary Form (in Appendix D) is one tool that can be used to properly plan a data governance project and increase the project and program management capabilities of the Data Governance Program.

As Caltrans completes the Data Business Plan and enters the implementation phase, it will need a tool such as the Project Summary Form to help prioritize workload and budgeting decisions, provide a consistent format for project presentations to benefit the Board, and improve the accuracy and consistency of project planning.

The Project Summary Form was developed with the assumption that it will be easy to use and will require no special software or other tools to use. This will make it useful to a wide variety of potential Communities of Interest at no extra cost in terms of staff time or operating expenses. The form:

- Provides a well-thought out, clear project description to decision makers
- Defines the problem, opportunity and objectives the project is addressing
- Identifies project risks, mitigation strategies and expected benefits
- Identifies who is responsible for and who will be impacted by the project

See Appendix D for the Project Summary Form with a description about each item.

Appendix A Glossary of Terms

1. AAADT – Annual Average Daily Traffic
2. ATR – Automatic Traffic Recorder
3. Board – The Transportation System Data Governance Board (Board) is the decision-making and policy-making authority for matters related to transportation system data for all divisions and districts at Caltrans. (Note: Some decisions may have to be referred to the Directorate – see below for a definition)
4. Board Chairperson — The Board’s Chairperson is selected by the Board and serves as the Board’s leader for a limited-term during which time the Chairperson facilitates Board meetings, communicates with the Board Sponsor, ensures the Board carries out its mission and achieves its goals, and facilitates implementation of Board decisions.
5. Board Members — Board members make key decisions regarding the activities of data governance, ensure that Caltrans is progressing toward effective data governance practices, raise issues and challenges that hinder effective data governance practices, select data governance projects, and hold project teams accountable for progress toward objectives.
6. Business Owners – Business Owners are the managers of the business units that own data at Caltrans. Business Owners are managers responsible for ensuring that a data product is being well managed and is providing value commensurate with the level of required investment for ongoing upkeep. They also appoint Data Custodians to serve as subject matter experts for transportation system data in their business unit and to address projects identified by the Board.
7. CDM – Custom Development Methodology
8. Communities of Interest (COI) – Communities of Interest are groups of internal and external stakeholders of data that have an interest in transportation system data and the data governance actions that impact data in which they have an interest.
9. Corporate Data Set – A corporate data set is an important subset of transportation system data. Corporate data sets span multiple business functions within Caltrans and therefore have an increased importance across the corporate structure of Caltrans. Corporate data sets are typically of great value to both internal Caltrans users and external users (e.g. the Legislature, the Governor’s Office, local agencies, etc.). Corporate data sets are priorities for data governance because of the broad and visible impact improvements in corporate data sets have across Caltrans.
10. CRS – California Road System
11. C-R-PM – County-Route-Postmile
12. Communities of Interest – An individual or group that could be affected by the data that is in-scope for the Data Governance Program. They come from across the organization or from outside the organization. They include groups who create data, those who use data, and those who set rules and requirements for data.
13. Data Custodians – The subject matter expert and point of contact for their assigned data products. Data Custodians represents the Business Owner and maintain communication with the Community of Interest. Data Custodians are existing Caltrans staff and are the recognized data experts for their functional areas (i.e. Division, Branch, Office, etc.).
14. Data Governance – The exercise of decision-making and authority for data-related matters
15. Data Element Catalog – A collection of data elements or attributes that may be from one or more tables or entities. The Data Element Catalog provides data elements of interest to Caltrans as a whole or to specific Communities of Interest.
16. Data Product – A collection of data that Communities of Interest require in order to manage the transportation system.

17. Data Set Catalog – A collection of data sets from one or more tables or entities. The relationships between data sets are limited to key words and general categories. The Data Set Catalog provides data sets of interest to Caltrans as a whole or to specific Communities of Interest.
18. Directorate – The Directorate is responsible for addressing critical risks that have been elevated from the Board.
19. DLG – Digital Line Graph
20. Enterprise Architecture Council (EAC) – The Enterprise Architecture Council helps guide Caltrans' direction related to enterprise-wide information technology and related systems.
21. Financial Policy Board – The Financial Policy Board is an executive-level entity that guides Caltrans' direction related to high-level financial policies and decisions. Some data governance issues with significant financial implications may have to be referred to the Financial Policy Board.
22. Gantt Chart – A Gantt Chart is a graphical representation of a project schedule. It includes start dates and end dates of each task throughout the life of each project.
23. Geospatial Data Management Committee (GDMC) – The GDMC was formed to address GIS data issues. The committee is advisory in nature and does not have an executive sponsor. The GDMC was important in the development of the GIS Corporate Structure Value Analysis Study (see below).
24. GIS – Geographic Information System
25. GIS Corporate Structure Value Analysis Study – The GIS Corporate Structure Value Analysis Study was undertaken by a variety of divisions in Caltrans to develop recommendations to improve the corporate organization of GIS in Caltrans and to facilitate a coordinated approach to GIS.
26. GIS Management Committee – This committee was previously responsible for developing GIS governance. It included executive sponsorship and a detailed organization. Although it was disbanded in 2005, the committee's projects and ideas may provide useful information for Caltrans' data governance effort.
27. HPMS – Highway Performance Monitoring System
28. IT Governance Board – The IT Governance Board is an executive-level entity that guides Caltrans' direction related to high-level information technology policies and decisions. Some data governance decisions impacting information technology systems or other infrastructure may have to be referred to the IT Governance Board.
29. LIGER – A Caltrans modification of the United States Census Bureau's "TIGER" GIS line data that references Caltrans LRS
30. LOS – Level of Service
31. LRS – Linear Referencing System
32. Metadata – Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data or information about information. (Information from: "Understanding Metadata," 2004 by the National Information Standards Organization)
33. MVSTAFF – Motor Vehicle Stock Travel and Fuel Forecast
34. PeMS – Performance Measurement System
35. PERT Chart – A Program Evaluation Review Technique (PERT) Chart is presented in the form of a network diagram that contains boxes that represent tasks (or milestones) in the project. These boxes are linked by lines with arrows to the next task in the sequence. The primary assumption in developing a PERT Chart is that no activity can begin until its immediately preceding task is completed.
36. PMS – Pavement Management System

37. ROI – Return on Investment
38. SHOPP Executive Committee – SHOPP Executive Committee is Caltrans’ executive-level entity that makes decisions regarding the State Highway Operations and Protection Plan (SHOPP). The purpose of the SHOPP is to maintain the state’s investment in the State Highway System and its infrastructure. Some data governance decisions impacting the SHOPP and/or the State Highway System may have to be referred to the SHOPP Executive Committee.
39. SME – Subject Matter Expert
40. TASAS – Traffic Accident Surveillance and Analysis System
41. Technical Advisory Committee (TAC) – An advisory body of technical data experts who review detailed information and provide advice and recommendations to the Data Governance Board. It consists of existing Caltrans staff with particular subject matter expertise in existing data and/or data requirements at Caltrans.
42. Transportation System Data – Transportation system data is data related to the definition, condition, or performance of the physical transportation system, including all transportation modes. It is a subset of all Caltrans data and does NOT include administrative data such as financial system data, human resources data, procurement data, and other data related to the operations of administrative functions.
43. TSN – Transportation System Network
44. TIGER – Topologically Integrated Geographic Encoding and Referencing System, the United States Census Bureau’s GIS line data.
45. VMT – Vehicle Miles Traveled
46. WIM – Weigh-in-Motion
47. Zachman Framework – The Zachman Framework is the taxonomy for organizing data in use at Caltrans. It links data to "who" the data impacts, targets, and the issues being addressed. It is organized in a matrix with five rows (or layers) [Scope (contextual), Business Model (conceptual), System Model (logical), Technology Model (physical), and Detailed Representation (out-of-context)] and six columns (or questions) [What (data), How (function), Where (network), Who (people), When (time), and Why (motivation)].

Appendix B Data Governance Board Charter

OVERVIEW

Transportation System Data Governance Board Purpose

The Transportation System Data Governance Board (Board) is the decision-making and policy-making authority for matters related to transportation system data for all divisions and districts at Caltrans (see Glossary of Terms for definitions).

The Board addresses transportation system data issues raised by Caltrans’ business units and recommends solutions, strategies and alternatives regarding transportation system data at Caltrans.

The Board will also oversee the implementation of transportation system data standards and quality assurance to ensure that divisions and districts are developing, maintaining, and providing transportation system data for the use of others.

*“Data Governance is the exercise of decision-making and authority for data-related matters.”
(From the Data Governance Institute)*

Mission

The Caltrans Transportation System Data Governance Board ensures that Caltrans creates and maintains reliable transportation system data that is accessible to Caltrans and its partners.

Goals

The following are strategic goals for the high-level direction of the Board:

Data Governance Strategic Goals

#	Goal Title	Goal Description
1.	Leadership	Champion data solutions, ensure accountability, and increase the value of data assets
2.	Quality	Oversee efforts to provide high quality data that is accurate, clear, and easy to access
3.	Prioritization	Prioritize efforts to address data gaps and needs
4.	Cooperation	Facilitate cross-organizational collaboration, data sharing, and integration (break down barriers between business units, reduce data silos)
5.	Flexibility	Encourage creative and innovative solutions to data needs

Guiding Principles

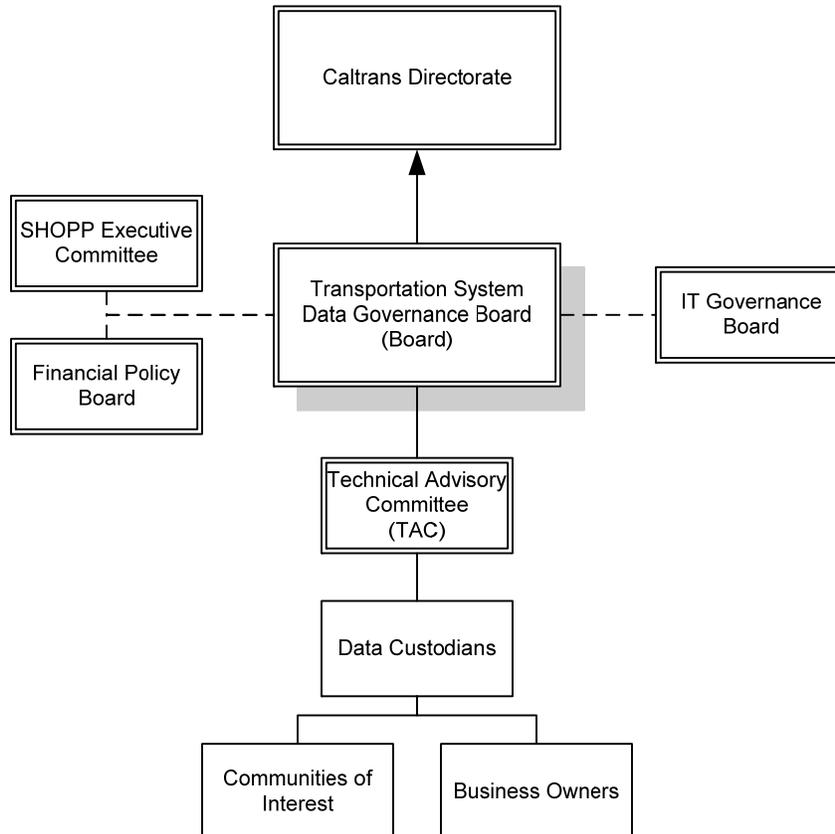
The following principles guide the Board:

- Promote data as a valued asset within Caltrans
- Ensure that all decisions are transparent, documented, and communicated to all stakeholders
- Support creative solutions to data governance.
- Facilitate cooperation and partnering across divisions and districts for corporate data
- Promote best practices and standardization for corporate data management and data quality.

Relationship to Caltrans Management

The Board will report to the Caltrans Directorate. The Board may also refer decisions to the SHOPP Executive Committee, the Financial Policy Board, and the IT Governance Board. The figure below displays these reporting and collaborative relationships.

The Board and Other Caltrans Management Bodies



Performance Objectives for First Year

During the first year of operation (January 2011 through December 2011), the Board will accomplish the specific objectives listed below:

- Establishment and regular meetings of the Board (following dissolution of the Interim Board)
- Select, initiate, and oversee one data governance pilot project
 - Establish criteria and evaluate corporate data sets
 - Evaluate and resolve data governance issues
- Establish data governance policies, rules, and guidelines, and develop recommendations for monitoring and compliance

- Establish and regularly convene the Technical Advisory Committee (TAC)

The Board will establish new objectives for calendar year 2012 during the final Board meeting of calendar year 2011.

MEMBERSHIP

The membership of the Interim Board is (as of June 20, 2011):

- Chairperson Coco Briseno, Chief, Division of Transportation System Information
- Robert Copp, Chief, Division of Traffic Operations
- Terry Abbott, Chief, Division of Design
- Jay Norvell, Chief, Division of Environmental Analysis
- Tony Tavares, Chief, Division of Maintenance
- Brent Green, Chief, Division of Right of Way
- Bill Bronte, Chief, Division of Rail
- Jane Perez, Chief, Division of Mass Transportation
- Sharon Scherzinger, Chief, Division of Transportation Planning
- Gary Cathey, Chief, Division of Aeronautics
- Doug Kempster, Division of Information Technology
- Sponsor Marty Tuttle, Deputy Director, Planning and Modal Programs

Board members shall be selected from divisions and districts, with one (1) Chairperson and ten (10) representatives from various divisions and three (3) district representatives. Membership shall not exceed fourteen (14) members. Members will serve under this charter for two years. Board members should be knowledgeable and passionate about the issues facing transportation system data and bring energy, creativity and innovation to the resolution of data issues facing Caltrans. Board member should have a sufficient breadth and depth across the entire span of Caltrans organization so as to understand the background and context of issues that the Board will consider.

The Board Chairperson will be selected by the current membership to provide consistency and mitigate turnover.

Member Roles and Responsibilities

The following are the roles and responsibilities of members of the Board:

Data Governance Roles and Responsibilities

Roles	Responsibilities
TSDGB Sponsor (Deputy Director of Planning and Modal Programs)	<ul style="list-style-type: none"> • Support the activities of the Board • Provide direction • Champion proposals to the Caltrans Directorate • Assess Board performance

Roles	Responsibilities
TSDGB Chairperson	<ul style="list-style-type: none"> Facilitate Board meetings Communicate with Board Sponsor Ensure the Board carries out the mission, goals, and guiding principles Facilitate implementation of Board decisions
TSDGB Members	<ul style="list-style-type: none"> Attend and actively participate in meetings Ensure that Caltrans is progressing towards effective data governance practices Raise issues and challenges that hinder effective data governance practices Select data governance projects and hold project teams accountable for progress towards objectives Assess Board performance

MEETINGS

Meeting Frequency

The Board meets quarterly in person at Caltrans Headquarters (1120 N Street, Sacramento). The Board's annual meeting schedule will be posted at the TSI website at <http://www.dot.ca.gov/hq/tsip/>. Regular meetings are mandatory for all members, or their alternates, and are used to discuss, develop, and recommend transportation system data solutions (see Use of Alternates).

The Board Chairperson can call additional meetings as needed to facilitate timely resolution of proposals. Typically, these meetings are used to address time-sensitive items. The additional meetings should focus on addressing single items of importance and bringing said item(s) to a completed decision point.

Meeting Notice

Meeting notices for regular meetings are distributed at the beginning of each calendar year to all members active at the time of distribution for the entire calendar year. Every attempt will be made to provide a minimum of ten (10) working days advance notice for additional meetings. All meeting notices are developed and distributed using Caltrans standard appointment scheduling tool (Lotus Notes) to ensure that the notices appear on the members' schedules.

Communications

Matters which are to be considered at a meeting shall be distributed to members for review at least five (5) working days in advance of the scheduled meeting. Notification also will include the time and place of the meeting and other essential information.

Notes are recorded at each meeting and will be made available to all participants within ten (10) working days following the meeting. Notes will record key discussion points of all motions and decisions, as well as any required action items.

Meeting Quorum

Before a meeting can conduct business it requires a quorum--the minimum number of members who must be present at the meeting before business can be transacted. For the Board, a quorum exists when a super majority (over 75%) of the voting members is present at the meeting.

Decision-making

Decision-making will be by consensus of the Board. A quorum of Board members is required to make decisions. The agenda shall clearly identify agenda items which require a decision.

Use of Alternates

Board members must designate an alternate to represent that area in the event the member cannot attend the Board meeting. Alternates are to attend meetings in the absence of the appointed member. Alternates must be able to speak for the member and have decision-making authority when substituting for a member.

Amendments to the Charter

Amendments to the Board Charter must be approved by a consensus of the Board.

Glossary of Terms

Board – The Transportation System Data Governance Board (Board) is the decision-making and policy-making authority for matters related to transportation system data for all divisions and districts at Caltrans.

Business Owners – The manager responsible for ensuring that the data product is being well managed and is providing value commensurate with the level of required investment for ongoing upkeep.

Communities of Interest – An individual or group that could be affected by the data that is in-scope for the Data Governance Program. They come from across the organization. They include groups who create data, those who use data, and those who set rules and requirements for data.

Data Custodians – The subject matter expert and point of contact for their assigned data products. Represents the Business Owner and maintains communication with the Community of Interest.

Data Governance – Data Governance is the exercise of decision-making and authority for data-related matters.

Directorate – The Directorate is responsible for addressing critical risks that have been elevated from the Board.

Transportation System Data – Transportation system data is data related to the definition, condition, or performance of the physical transportation system, including all transportation modes. It does NOT include administrative data such as financial system data, human resources data, procurement data, and data related to the operations of administrative functions.

Technical Advisory Committee (TAC) – An advisory body of technical data experts who review detailed information and provide advice and recommendations to the decision-making body.

Appendix C Data Issue Form

Purpose: The purpose of this document is to record and submit data governance issues and potential projects to the Transportation System Data Governance Board and Technical Advisory Committee.

Data Issue Overview

Data Issue Title	<Succinct data issue title>
Data Issue Description	<Briefly describe the data issue>
Benefit to Solving Issue	<Briefly describe the benefit to Caltrans if this data issue is successfully addressed>
Division or Office	<Division/Office that 'owns' the issue>
Division or Office Chief	<Division/Office Chief that 'owns' the issue>
Data Custodian	<Name> <Division/Office>

Impacted Data and Systems

Impacted Data	<ul style="list-style-type: none"> <List key data groups that may be impacted by this issue>
Impacted Caltrans Systems	<ul style="list-style-type: none"> <List Caltrans systems that may be impacted by this issue>

Project Description

<p><Provide a description of the proposed project to address the data issue defined above and the benefits that Caltrans and other stakeholders will realize. Please use as much space as necessary to describe your project adequately.</p> <ul style="list-style-type: none"> If estimated resource needs and schedule are known, please include them in the project description If potential project team members are known, please include their names and divisions>

Business Impact Description

No.	Risk
1.	<List business impacts if project is not undertaken>
2.	
3.	
4.	
5.	

Appendix D Project Summary Form

Item	Description
Project Title	<Insert Project Title here>
Project Description	<Enter a detailed description of the project that includes: 1. Objectives, 2. Business opportunities or problems to be addressed, and 3. Expected results>
Known Implementation Risks/Issues	<Describe any known implementation risks and/or issues that must be considered before and during project implementation>
Known Mitigation Factors	<Enter any viable options for addressing known risks and/or issues>
Expected Benefit	<Describe who will benefit, and the expected benefits>
Dependencies With Other Data Governance Projects or Activities	<Describe any data governance activities or projects that impact this project>
Project Sponsor or Business Owner	<Enter the name of project champion or business owner (the person or group who controls the resources for the project)>
Key Project Stakeholders or Communities of Interest	<Describe those people who will be most affected by the project>
Project Manager	<Enter the name of the person who will execute the project>
Data Custodian(s)	<Enter the name of the Data Custodian(s) responsible for the data impacted by this project>
Additional Roles / Responsibilities	<Additional people needed on the project team or needed to support the project team>
Key Milestone Dates	<Enter the dates for key project milestones (non-negotiable dates)>
Project Duration	<Enter the date the project will begin, will end, and total duration>
Cost to Implement	<Enter both the number of PYs (or hours) and the project dollar cost>

Appendix E Proposed Project Prioritization Form

Purpose: The purpose of this document is to present the Technical Advisory Committee’s preliminary prioritization of proposed data governance projects to the Transportation System Data Governance Board for review and final prioritization.

Prioritization Date: _____

This table provides a brief listing of proposed data governance projects. The table on the following pages presents a brief summary of each proposed project.

Priority No.	Proposed Project Title	Proposing Division or Office
1.	< Enter Project Title >	<Enter Division/Office>
2.		
3.		
4.		
5.		
6.		
7.		
8.		
9.		
10.		
11.		
12.		
13.		
14.		
15.		
16.		
17.		
18.		
19.		
20.		
21.		
22.		
23.		
24.		
25.		

Proposed Project's Overview

Priority No.	Proposing Division or Office	<Enter Division/Office>	Resource Estimate	Hours: <XXX> Dollars: \$ <X.XX>
1	Project Title	<Succinct project title>	Estimated Timeframe	MM/YY to MM/YY
	Project Benefit	<Briefly describe the benefit to Caltrans if this project is successfully implemented>		
Business Impact Description:		<Briefly describe business impact if project is not undertaken>		

Priority No.	Proposing Division or Office	<Enter Division/Office>	Resource Estimate	Hours: <XXX> Dollars: \$ <X.XX>
2	Project Title	<Succinct project title>	Estimated Timeframe	MM/YY to MM/YY
	Project Benefit	<Briefly describe the benefit to Caltrans if this project is successfully implemented>		
Business Impact Description:		<Briefly describe business impact if project is not undertaken>		

Priority No.	Proposing Division or Office	<Enter Division/Office>	Resource Estimate	Hours: <XXX> Dollars: \$ <X.XX>
3	Project Title	<Succinct project title>	Estimated Timeframe	MM/YY to MM/YY
	Project Benefit	<Briefly describe the benefit to Caltrans if this project is successfully implemented>		
Business Impact Description:		<Briefly describe business impact if project is not undertaken>		

Priority No.	Proposing Division or Office	<Enter Division/Office>	Resource Estimate	Hours: <XXX> Dollars: \$ <X.XX>
4	Project Title	<Succinct project title>	Estimated Timeframe	MM/YY to MM/YY
	Project Benefit	<Briefly describe the benefit to Caltrans if this project is successfully implemented>		
Business Impact Description:		<Briefly describe business impact if project is not undertaken>		

Priority No.	Proposing Division or Office	<Enter Division/Office>	Resource Estimate	Hours: <XXX> Dollars: \$ <X.XX>
5	Project Title	<Succinct project title>	Estimated Timeframe	MM/YY to MM/YY
	Project Benefit	<Briefly describe the benefit to Caltrans if this project is successfully implemented>		
Business Impact Description:		<Briefly describe business impact if project is not undertaken>		

Appendix F Quarterly Project Progress Form

Purpose: The purpose of this document is to provide quarterly project status and progress towards meeting a data governance project goals and objectives. This report is completed periodically by the project team for review by the Technical Advisory Committee and the Transportation System Data Governance Board.

Project Overview

Project Title	<Succinct project title>		
Brief Project Description	<Briefly describe the project>		
Data Issue or Problem	<Briefly describe the data related issue or problem that this project is intended to address>		
Project Benefit	<Briefly describe the benefit to Caltrans if this project is successfully implemented>		
Resource Estimate	Hours: <XXX> Dollars: \$ <X.XX>	Estimated Timeframe	<MM/YY> to <MM/YY>

Project Progress

Report Date	<MM/DD/YY>	Progress Period	<Month YYYY>	Resources Expended	Hours: <XXX> Dollars: \$ <X.XX>
On Time?	<Yes/No>	On Budget?	<Yes/No>	On Scope?	<Yes/No>
Status Details	<If you answered 'No' to any of the above, please explain>				

This Period

Detailed Accomplishments
<Provide details of this period's accomplishments>

Next Period

Milestones, Tasks, and Deliverables
<Provide details of next period's milestones, tasks, and deliverables>

Risk/Issue Description and Mitigation Strategy

No.	Risk/Issue	Mitigation Plan
1.	<List current risk/issue>	<Describe what is being done to eliminate or mitigate the risk/issue>
2.		

Appendix G Data Governance Policy Form

Data Governance Policy

Number: <DG-##>
Relevant Policies: <A – name>
<B – name>
<C – name>
Effective Date: <Month DD, YYYY>

TITLE <Title of the Policy>

POLICY DESCRIPTION
< Policy Description>

BACKGROUND
< Background Description>

RESPONSIBILITY

< Office/Chief Name>

- <Responsibility Statement>
- <Responsibility Statement>

< Office/Chief Name>

- <Responsibility Statement>
- <Responsibility Statement>

APPLICABILITY
<Enter the applicable entities (i.e. All Caltrans' systems)>

<Policy Authorizers Name>
<Policy Authorizers Position>

<Date Signed>

Appendix H Data Element Metadata Form

Purpose: The purpose of this worksheet is to facilitate the development and recommendation of metadata for key Caltrans data elements.

Background: Often, data elements are duplicative or have definitions that overlap and lead to confusion, causing errors and/or extra effort to use the data to meet Caltrans business needs. One of the goals of data governance at Caltrans is to develop data standards that provide a consistent and clear definition of each data element along with additional metadata.

Directions: Please complete the table below by listing the key metadata values for a single data element.

[Data Element name] Metadata

#	Metadata Category	Definition	Metadata Value
1.	Long Name	The full title of the data element.	<Enter Value>
2.	Short Name	The abbreviated title of the data element.	<Enter Value>
3.	Description	A brief description of the data element.	<Enter Value>
4.	Format	The format of the data element (e.g., text, integer, data, currency)	<Enter Value>
5.	Length	The number of spaces required for the data element, including decimal places if applicable.	<Enter Value>
6.	Valid Values	The values which are valid for this data element (e.g., list of valid county codes, values greater than zero).	<Enter Value>
7.	Business Rules or Specifications	Specific rules or specifications for creating valid values or using this data element.	<Enter Value>
8.	Data Sets	The data set(s) in which this data element is included.	<Enter Value>
9.	Keywords	Keywords which users of the data element can search to find information related to the data element.	<Enter Value>
10.	Source Systems	Systems containing the data element.	<Enter Value>
11.	Publication Date	The last date the data element metadata was updated.	<Enter Value>
12.	Status	The current status of the data element, including problems or issues with the data.	<Enter Value>
13.	Originator	The Caltrans organization that owns the data element.	<Enter Value>
14.	Point of Contact	The person to contact for information or questions regarding the data element.	<Enter Value>
15.	Metadata Date	Date the metadata of the data element was last updated or changed.	<Enter Value>
16.	Metadata Contact	The contact person for questions or information regarding data element metadata. <i>(May or may not be same person as Point of Contact).</i>	<Enter Value>

Appendix I Corporate Data Set Metadata Form

Purpose: The purpose of this worksheet is to facilitate the development and recommendation of metadata for Caltrans corporate data sets.

Background: Data sets can include data tables or data elements that are duplicative of data found in other data sets and can lead to confusion, errors, and unnecessary effort on the part of those using and maintaining the data set. Establishing metadata will lead to a better, clearer understanding of each data set and how it is related to other data sets at Caltrans.

Directions: Please complete the table below by listing the key metadata values for the corporate data set.

[Corporate Data Set name] Metadata

#	Metadata Category	Definition	Metadata Value
1.	Data Set Category	Transportation system data category to which the data set is related	<Enter Value>
2.	Data Set Title	The title of the data set.	<Enter Value>
3.	Description	A brief description of the data set.	<Enter Value>
4.	Originator	The Caltrans organization that owns the data set.	<Enter Value>
5.	Point of Contact	The person to contact for information or questions regarding the data set.	<Enter Value>
6.	Publication Date	The last date the data set metadata was updated.	<Enter Value>
7.	Online Link	The Caltrans link for the data set.	<Enter Value>
8.	Time Period of Data	The time period covered by the data in the data set.	<Enter Value>
9.	Status	The current status of the data set, including problems or issues with the data.	<Enter Value>
10.	Keywords	Keywords which users of the data set can search to find information related to the data set.	<Enter Value>
11.	Access Constraints	Constraints to user access to the data set.	<Enter Value>
12.	Use Constraints	Constraints to the use of the data from the data set.	<Enter Value>
13.	Source Systems	Systems containing the data set.	<Enter Value>
14.	Data Quality Information	Specific data quality issues related to the data set (i.e. duplicate data, incomplete data, inaccurate data, etc.).	<Enter Value>
15.	Primary Data Elements	The most important data elements included in the data set.	<Enter Value>
16.	Metadata Date	Date the metadata of the data set was last updated or changed.	<Enter Value>
17.	Metadata Contact	The contact person for questions or information regarding data set metadata. <i>(May or may not be same person as Point of Contact).</i>	<Enter Value>

Appendix J Data Quality Assessment Worksheet

#	Assessment Question	Answer
1.	What is the data you are assessing?	<Answer here>
2.	If data currently exists, are there data duplications?	<List the data duplications and any additional details>
3.	If data currently exists, are there data overlaps (not exact duplications)?	<List the data overlaps and any additional details>
4.	If data does not currently exist, what resources are required to gather it?	<Answer here>
5.	If data does not currently exist, what is the value of capturing it now?	<Answer here>
6.	If data does not currently exist, what is the return on investment (ROI) for capturing it?	<Answer here>
7.	Does the data have a single, clear definition?	<p><Answer here></p> <p><i>Research and list all current definitions</i></p> <p><i>Clarify any related/competing definitions</i></p> <p><i>Choose which definition is the correct one and recommend establishment across Caltrans/the division/...</i></p>
8.	Is the data accurate (valid)?	<p><Answer here></p> <p><i>Research and measure the accuracy by comparing it to other data sources and interview data experts.</i></p> <p><i>Does the data fall into an acceptable accuracy range (validity)?</i></p>
9.	Is the data complete?	<p><Answer here></p> <p><i>Does the data appear in all of the fields that it is anticipated that it should appear?</i></p>
10.	Is the data maintained?	<p><Answer here></p> <p><i>How often is the data updated? Does the update frequency match expectations? Is that often enough?</i></p>
11.	Is the data accessible?	<p><Answer here></p> <p><i>Who can currently access the data?</i></p> <p><i>Who needs to access the data?</i></p> <p><i>Can the data be made more accessible, if necessary?</i></p>
12.	Is the data timely?	<p><Answer here></p> <p><i>Do the necessary users have access when needed?</i></p>
13.	Is the data sensitive/confidential?	<Answer here>

Appendix K Data Element Analysis Worksheet

The 'Data Element Analysis Worksheet' below guides analysis of data issues or problems at the data element level. The worksheet requires the following information:

- Description of the issues or problems
- Recommended solution to the issues or problems
- Implementation analysis, particularly regarding whether or not existing systems should be updated with new data standards
- An estimate of the cost and return on investment of implementing the new data standards in existing systems

Also, please list the recommended data standards that will be developed through this analysis and implementation. Make sure that the data element name is unique, that its key metadata is defined (format and length), and that the data element definition is clear and specific to that data element.

Below is a table with questions to use in applying a data element analysis. Included here is a sample assessment of the "County" data element that is displayed for sample demonstration purposes only and is NOT the result of any formal Caltrans analysis or recommendation.

Data Element Analysis (with sample analysis of the "County" data element)

#	Assessment Question	Answer
1.	What are the data element issues?	<p><List the issues></p> <p>Sample for "County" data element:</p> <ul style="list-style-type: none"> • There are duplications of the data element "County" in different systems that have different format and length. • There are related data elements to identify a county or a county associated with a highway route.
2.	What is the data standard recommendation to resolve the issue(s)?	<p><State and explain the recommendation></p> <p>Sample for "County" data element:</p> <ul style="list-style-type: none"> • Eliminate the duplication of the element "County" by creating a single standard for "County" • Establish a second standard if necessary for the duplicated element with a different element name • Establish data standards for all data elements in analysis • Update definitions of all data elements as necessary to ensure that their meaning and use are clear.
3.	Implementation Analysis: Should Caltrans employ the data standard going forward only? Or should existing systems containing these data elements be updated to the new standard?	<p>< State and explain the implementation analysis></p> <p>Sample for "County" data element:</p> <ul style="list-style-type: none"> • The new data standards should be adopted into all new systems. • Existing systems undergoing modernization should adopt the new standards if practical.
4.	What is the estimated cost and return on investment for completing the data standardization?	<p>< Provide an estimate of cost and return on investment></p> <p>Sample for "County" data element:</p> <ul style="list-style-type: none"> • It will be too costly to update county code data in all of the impacted systems throughout Caltrans.

Appendix L Risk Management Worksheet

#	Risk	Probability	Impact	Affected Area	Warning Signs	Preventive Measures
1.	<A concise statement of the risk and its consequence>	<Likelihood of a risk occurring>	<Impact of risk if not avoided or mitigated>	<The areas affected should the risk occur. May include: Budget, Schedule, Resources, etc.>	<Indications that a risk exists>	<The mitigations and/or contingency actions to be taken to reduce the probability of the risk occurring and/or the severity of the risk should it occur.>
2.						
3.						
4.						
5.						
6.						
7.						
8.						
9.						
10.						
11.						
12.						
13.						
14.						
15.						
16.						
17.						
18.						
19.						
20.						
21.						
22.						
23.						
24.						
25.						

Appendix M Corporate Data Sets – Complete List

#	Corporate Data Set	Definition	Business Functions	Systems
1.	Collision data	Collision data is recorded from the CHP's traffic collision reports that include all reported incidents on the state highway system. The data from these traffic collision reports is entered into TSN (formerly TASAS) from hard copies of the traffic collision reports. An estimated County/Route/Postmile location is also included and derived from the location description provided by the CHP officer.	Operations	TSN
2.	Traffic Investigations	Traffic Investigations contain data related to potential traffic issues for specific sections of the state highway system. Traffic investigations are triggered by data indicating a potential problem on a section of a highway. Following an analysis of the highway section and related data, the traffic investigation information is entered into TIRTS (Traffic Investigation Reports Tracking System).	Operations	TIRTS
3.	Worker safety data	Worker safety data is maintained by the Caltrans Safety Office and records highway accidents that result in Caltrans and/or contractor highway worker injuries or fatalities.	Safety Office	Unknown
4.	AADT (Annual Average Daily Traffic)	AADT data is derived from formulas applied to traffic census data that measures the level of traffic on the state highway at individual locations for specific dates and times. Traffic census data is gathered either through PeMS detector systems built into the travel way, or through other traditional traffic census methods. AADT calculations are contained in TSN (formerly TASAS) and are used for various operations, planning, and design purposes throughout Caltrans.	Operations Planning Design	TSN PeMS
5.	VMT (Vehicle Miles Traveled)	VMT data is derived from AADT data and measures the total number of miles traveled by vehicles, over a specific length of highway, at a specific time period. VMT is currently reported in both TSN and the Highway Performance Monitoring System (HPMS), although the VMT amounts in each system differ because of the use of different highway segments in calculating VMT.	Operations Planning Design	TSN HPMS

#	Corporate Data Set	Definition	Business Functions	Systems
6.	Truck Volumes	Truck volume data is similar to AADT except that it captures only the level of traffic traveled specifically by trucks. Trucks are divided into various classifications based on the number of axles on the truck. Truck data is gathered by Weigh-in-Motion (WIM) systems on the highways, or by other traditional traffic census methods.	Operations Planning Design	TSN HPMS
7.	Performance Measurement System (PeMS)	PeMS is real-time traffic volume data that is captured by automated counting devices (i.e. loop detectors) built into the travel way of the highway system. PeMS is currently only located in urban areas of the state. Data from the detectors is transmitted to the PeMS system, which is used as the basis for communicating real-time traffic information and developing traffic census data for specified locations.	Operations Planning	PeMS TSN
8.	Arterial Performance Measurement System (APeMS)	APeMS is similar to PeMS (see above) but records the real-time traffic volume data for local arterial roads. APeMS was developed by Caltrans and SANDAG (San Diego Area of Governments). Data from the detectors is transmitted to the PeMS system and is used as the basis for communicating real-time traffic information, as well as for developing traffic census data for specified locations.	Operations Planning	APeMS
9.	Transit Performance Measurement System (TPeMS)	TPeMS was created by Caltrans and SANDAG to record transit ridership information. This information allows transportation system managers to make intermodal decisions regarding the regional transportation system.	Operations Planning Modal Programs	TPeMS
10.	Ridership	Ridership data captures the use of transit and rail systems in California. Some of this data is reported to the National Transit Database.	Planning Modal Programs	Natl. Transit Database (NTD)
11.	Park and Ride	Park and Ride data consists of information related to the location, capacity, and usage of Park and Ride lots controlled or managed by Caltrans.	Operations Planning Modal Programs	Unknown

#	Corporate Data Set	Definition	Business Functions	Systems
12.	Capital Project data (Project Delivery Outputs)	Capital project data includes planning, project development, right-of-way, design, and construction data regarding capital projects for the entire life-cycle of the project. This data captures project delivery outputs and includes CADD/GIS data used in the development and design of capital projects. This data also contains lifecycle cost and asset management data related to capital projects.	Planning Environmental Right-of-Way Design Construction Modal Programs Local Asst Maintenance	PRSM XPM CMS STEVE IMMS PMS BMS GIS/CADD AIS (R/W)
13.	Rail/Transit Capital Project data	Rail/Transit Capital project data includes planning, project development, right-of-way, design, and construction data regarding rail/transit capital projects for the entire life-cycle of the project. This data captures project delivery outputs and includes CADD/GIS data used in the development and design of rail/transit capital projects. This data also contains lifecycle cost and asset management data related to rail/transit capital projects.	Planning Environmental Right-of-Way Modal Programs	Natl. Transit Database (NTD) Rail Data
14.	Right-of-Way Capital Project data	Right-of-way Capital project data includes planning, project development, right-of-way, design, and construction data regarding right-of-way capital projects for the entire life-cycle of the project. This data captures project delivery outputs and includes CADD/GIS data used in the development and design of right-of-way capital projects. This data also contains lifecycle cost and asset management data related to right-of-way capital projects.	Planning Environmental Right-of-way Design Construction	AIS (R/W)
15.	Inventory	Inventory data includes highway inventory (including all highway inventory elements within the right-of-way), non-highway inventory (e.g. maintenance stations, rest areas, office buildings, etc.), other land parcels/real estate (within the right-of-way owned by Caltrans), and Modal Program inventory related to rail, transit, and aeronautics facilities.	Planning Environmental Right-of-way Design Construction Modal Programs Local Assistance Maintenance	IMMS PMS BMS AIS (R/W)

#	Corporate Data Set	Definition	Business Functions	Systems
16.	Pavement Condition	Pavement condition information is derived from the "Pavement Condition Survey" which is an in-person assessment of the pavement's condition on the entire state highway system. The results from the in-person assessments are entered into the Pavement Management System (PMS) for use in determining future needs, planning projects, measuring the effectiveness of past pavement treatments, and other purposes.	Planning Design Construction Maintenance Operations	PMS HPMS
17.	Bridge Condition	Bridge condition information is derived from bridge inspections, which are in-person assessments of the condition of bridges on the state highway system and local agency bridges. Bridges are inspected every two years. The results from the in-person inspections are entered into the Bridge Management System (BMS) for use in determining and recording bridge condition, assessing future needs, planning projects, and measuring the effectiveness of past bridge repairs, among other purposes.	Planning Design Construction Maintenance Operations	BMS
18.	Environmental data	Environmental data includes a variety of data related to the impact of the highway system and other transportation modes on the environment. Some important environmental data includes: <ul style="list-style-type: none"> • Air quality (attainment) • Critical Habitats • Significant Wetlands • Storm water • Physical/Natural/Community 	Environmental Planning Design Construction Maintenance	STEVE
19.	Level of Service (LOS)	Two types of LOS are used in Caltrans. One is used to measure the performance of highway maintenance and a second is used to measure traffic performance. <u>Maintenance LOS</u> is based on field reviews conducted each year, consisting of 10% of the highway system's center-line miles. Reviewers determine whether or not the inventory (i.e. guardrail, striping, fencing, etc.) of a particular postmile passes or fails based on established maintenance thresholds. This data is entered into the Maintenance LOS system and is used to calculate LOS scores for	Maintenance Operations Environmental Planning	Maintenance LOS System Traffic LOS system unknown

#	Corporate Data Set	Definition	Business Functions	Systems
		<p>districts and regions based on accumulated review results.</p> <p><u>Traffic LOS</u> is a calculated measure of effectiveness for various types of highway facilities (freeways, ramps, highways, intersections, etc.) based on related performance measures (density, delay, speed, etc.). This LOS measure is used in Traffic Impact Studies, Route Concept Reports, etc.</p>		
20.	County/Route/Post mile	County/Route/Postmile data provides essential location data along the state highway system. This data is used in many systems but the primary system of record for County/Route/Postmile is TSN. Caltrans also built a GIS-based LRS tool that links County/Route/Postmile location data to geospatial coordinate data.	Multiple	TSN HPMS State Roadway Network (GIS)
21.	Technical Characteristics	Technical Characteristics contain the details of the highway system, including the number of lanes, width of lanes, shoulder type, key landmarks, and other recorded details on the Highway Log.	Multiple	TSN HPMS State Roadway Network (GIS)
22.	Functional Classification	Functional Classification provides a method for classifying highways and local roads according to the type of service they provide. It is used in the determination of federal aid eligibility, transportation planning processes, and project design, among other processes. Functional classification data is provided in the HPMS system for federal reporting purposes and is also recorded in California Road System (CRS) maps.	Multiple	HPMS CRS maps
23.	System Designations	System Designations data refers to the designation of state highways and highway segments according to Federal Aid System designations (i.e. the Interstate System, the National Highway System, and other related subsystems).	Multiple	TSN HPMS
24.	City/County/District	City/County/District data refers to city and county governmental designations and Caltrans District designations, which are used in multiple systems to provide location and jurisdiction information.	Multiple	TSN HPMS State Roadway Network (GIS)

#	Corporate Data Set	Definition	Business Functions	Systems
25.	Geographical Information System (GIS) Data	GIS data and geospatial data are developed in a variety of systems in Caltrans. The primary system is the State Roadway Network maintained in the Division of Transportation System Information. Other systems are maintained in the Division of Maintenance, and in some districts. In addition, the Division of Design has resources dedicated to GIS to manage the capital project GIS (and CADD) data used in project development.	Multiple	State Roadway Network (GIS) Other GIS systems
26.	Terrain	Terrain data describes the terrain features (e.g. elevation, climate region, urban/rural, agricultural, mountainous, coastal, forested, etc.) along the entire state highway system.	Multiple	TSN State Roadway Network (GIS)

Appendix N Vehicle Miles Traveled Data Sources Fact Sheet

Vehicle Miles Traveled Data Sources Fact Sheet June 2010

Purpose: To explain that different estimates of vehicle miles of travel (VMT) are being reported from various sources within the Department of Transportation. A VMT Reporting Committee was established to explore the reasons behind the conflicting VMT numbers and to develop recommendations for the Divisions of Transportation System Information (TSI), Traffic Operations and Caltrans management to identify a single and consistent source for VMT reporting.

Definition: Vehicle Miles of Travel (VMT) is an aggregate measure of travel occurring on all or part of a highway system. It is the sum of miles traveled by all vehicles during a fixed period of time on a fixed expanse of highways (Source: California Public road and Related Data).

Annual VMT = Segment length X AADT (annual average daily traffic) X 365
The unit is in millions of miles

Daily VMT = Segment length X AADT (annual average daily traffic)
The unit is in 1000's of miles

Background: The VMT Reporting Committee included a group of ten people from TSI and Traffic Operations staff and managers. The committee members included: Ayalew Adamu, Brad Boehm, Brian Domsic, Frank Law, Ioana Pana, Jahangir Kashkooli, Joe Avis, Sarah Chesebro (Chair), Soheila Khoii, and Tremain Downey.

The committee held several monthly meetings since April of 2008 to discuss the discrepancy of VMT data reported from the Department prepared by different offices and also compared it with what was reported by Federal Highway Administration (FHWA).

State Highway VMT data sources in Caltrans are:

1. Highway Performance Monitoring System (HPMS) by Division of TSI.
2. Traffic Accident Surveillance and Analysis System (TASAS) Data Base by Division of TSI (now), previously Division of Traffic Operations.
3. Traffic Operations Monthly Report (annually reconciled to agree with TASAS's year-end VMT)

During the various meetings the committee discussed and identified the sources of VMT data and method of estimation of segment lengths, lane miles, and VMT by each source.

Conclusion: The main differences between HPMS and TASAS are:

1. Segment lengths; HPMS has Discrete segments which are larger than TASAS
2. Updates of relinquishments and adoptions are not the same in those two sources
3. Reporting deadline is different between these two sources. HPMS and TASAS's vehicle miles of travel are calculated at different times of the year.
 - HPMS data is due June 15th
 - TASAS data is due whenever the accident rates for the calendar year have been entered
 - System changes could have been made in TASAS and not captured in HPMS

The Traffic Operations Monthly Report: is calculated using up to 20 Traffic Monitoring Sites. The monthly Average Daily Traffic at each site is compared to the monthly average daily traffic for the same month a year ago. A total volume change (of all sites available) is then calculated. The volume change is applied

to the monthly VMT of a year ago. The monthly VMT base was set in 1974 and has been a running report since. After the TASAS calculates the annual statewide VMT, the previous year's monthly VMT will be revised.

The following sections explain the usage of TASAS and HPMS VMT data by Caltrans and others.

Usage of TASAS VMT Data

VMT data is reported in the "Collision Data on California State Highways" annual publication. VMT is labeled as TRAVEL (MVM) and values are calculated statewide and county wide. The statewide TRAVEL is used to calculate the FATALITIES RATE or DEATH RATE which is reported in the "Collision Data on California State Highways" annual publication and reported to Caltrans' Public Affairs Office as a Safety Performance Measure.

TASAS VMT data contributes directly to the many Caltrans safety related activities and provides important data ultimately leading to improved efficiency of state operations. Data produced by this activity is used extensively throughout Caltrans and is the primary source of data for the Highway Safety Improvement Program. TASAS information and reports are available to all Caltrans' employees, SHSP partners and other California agencies. The list of Caltrans' partners includes the FHWA, TRB, Highway Safety Information System research program and other transportation safety researchers. Caltrans also publishes an Annual Statewide Collision Data Report that is available to the general public upon request.

TASAS VMT data is collected by the TRAFFIC CENSUS PROGRAM and published on its web site. Annual publications that result from this program are available to MPO's, RTPA's and Local Agencies in hard copy and web access.

Web site: <http://www.dot.ca.gov/hq/traffops/saferesi/trafdata/index.htm>.

Usage of HPMS VMT Data

The reports and procedures outlined in the "HPMS Field Manual for the Continuing Analytical and Statistical Data Base" are authorized under 23 U.S.C. 315, which places the responsibility on the Secretary of Transportation for management decisions which affect transportation. In addition, 23 CFR 1.5 and 49 CFR 1.48 provide the Federal Highway Administrator with authority to request such information deemed necessary to administer the Federal-aid highway program. Estimates of future highway needs of the Nation are mandated by Congress on a biennial basis (23 U.S.C 502(g)). Data are used for assessing highway system performance under FHWA's strategic planning and performance reporting process in accordance with requirements of the Government Performance and Results Act (GPRA, Sections 3 and 4) and for apportioning Federal-aid highway funds. Finally, 23 CFR 420.105(b) requires States to provide data that support FHWA's responsibilities to the Congress and the public.

The HPMS is not only a data collection and reporting system, but an analytical system that consists of a series of interrelated simulation models designed to serve the needs of the policy decision-making process of the agency. The HPMS data base, in concert with the inventory of the Nation's structures, consists of the continuing source of data used to prepare the legislatively mandated biennial report to Congress, Status of the Nation's Surface Transportation System: Condition and Performance. In addition, the HPMS serves as the single continuing data source used to carry out a host of special studies and operational functions including:

- Various State management systems.
- Vehicle size and weight studies.
- Federal-aid program apportionments and allocations.
- Environmental Protection Agency's (EPA's) Section 187, VMT Forecasting and Tracking

Guidance (Attachment I) and Transportation Conformity Rule (40 CFR Parts 51 and 93) for urbanized areas that are National Ambient Air Quality Standards (NAAQS) non-attainment areas.

- FHWA publications including Highway Statistics and Our Nation's Highways.
- California Highway Patrol (CHP) uses HPMS vehicle miles of travel data to forecast collision rate

Length, lane-mile, and travel data are also used for the apportionment and allocation of Federal-aid funds [23 U.S.C. 104(b)]. Numerous special travel data reports are prepared in response to requests from within the FHWA and National Highway Traffic Safety Administration (NHTSA).

Offices from within the Department of Transportation, EPA, Department of Defense, Congress, State governors and legislators, and numerous organizations and individuals in the private sector, as well as the general public, also request these data. The data collected in accordance with the HPMS Field Manual are unique in that they directly tie together roadway physical, operational, use (travel), pavement, condition, and performance data that can be analyzed and summarized at the sub-State, statewide, regional, and national levels. The HPMS is also unique in that it makes effective use of statistically based sampling. These data are not obtainable from any other known source.

The HPMS data is used by FHWA to assess the performance of the Nation's highway transportation system as well as identify future highway system needs. Among other things, the HPMS data is used as a basis for developing improvements to the overall highway system. The resulting improvements facilitate the mobility of the highway users while enhancing the economic growth and trade opportunities on the part of providers of goods and services and consumers.

VMT data is provided to the general public via the California Public Road Data which is published annually by the Division of Transportation System Information. This publication can be accessed on this web site.

<http://www.dot.ca.gov/hq/tsip/hpms/datalibrary.php>

Recommendation: The VMT reporting committee recommends that:

The VMT reported to FHWA with the annual HPMS submittal serves as the Department's official VMT. However, TSN-TASAS vehicle miles of travel data are the source for calculation of collision rates on the State Highway system and in the traffic safety analyses and reports.

Also, we recommend that the "historical monthly vehicle miles of travel" and "historical county vehicle miles of travel" reports on Traffic Operation's web site shall be consistent with HPMS vehicle miles of travel data.

Appendix O Project Summary – Vehicle Miles Traveled (VMT) Pilot

Item	Description
Project Title	Vehicle Miles Traveled (VMT)
Project Description	<p>Objectives:</p> <ul style="list-style-type: none"> • Develop consistent reporting of VMT data • Resolve data issues to reduce/remove conflicting VMT data from different sources <p>Business Opportunities/Problems:</p> <ul style="list-style-type: none"> • Understand differences in VMT data reported within Caltrans, by other state agencies, and by local agencies • Resolve differences in reporting, where possible • Develop processes to report proper and accurate VMT data <p>Expected Results:</p> <ul style="list-style-type: none"> • Understand all sources of VMT and remove differences, where possible • Develop clearly defined policy regarding reporting of VMT
Known Implementation Risks/Issues	<ul style="list-style-type: none"> • VMT used for collision data has specific needs and may not be able to conform to other VMT calculations • Other state and local agencies may not be willing to cooperate in developing standardized VMT calculations, depending on specific needs
Known Mitigation Factors	<ul style="list-style-type: none"> • Research special needs for VMT by various state and local agencies • Include other agencies in discussions of potential solutions
Expected Benefit	<ul style="list-style-type: none"> • Development of clearly defined process and policy for use of VMT within Caltrans • Coordination with other agencies in use of VMT
Dependencies With Other Data Governance Projects or Activities	<ul style="list-style-type: none"> • Development of TAC needed to precede development of data governance project teams • Changes in AADT processes may impact VMT, which uses AADT in calculations
Project Sponsor or Business Owner	Mark Samuelson, TSI
Key Project Stakeholders or Communities of Interest	<ul style="list-style-type: none"> • Division of Traffic Operations • Division of TSI • Division of Planning
Project Manager	<i>To be determined</i>
Data Custodian(s)	<i>To be determined</i>

Item	Description
Additional Roles / Responsibilities	<ul style="list-style-type: none"> • Data Governance Board – Review and approve project recommendations • TAC – Review and revise project recommendations
Key Milestone Dates	<p>October 2011 – Project kickoff January 2012 – Research complete March 2012 – Present initial recommendations April 2012 – TAC review – revise as necessary May 2012 – Board review – revise as necessary June 2012 – Recommendations approved</p>
Project Duration	October 2011 to June 2012
Cost to Implement	<i>To be determined</i>

Appendix P Project Summary – County/Route/Postmile- LRS Pilot

Item	Description
Project Title	County/Route/Postmile – Linear Referencing System (LRS)
Project Description	<p>Objectives:</p> <ul style="list-style-type: none"> • Select key data issues related to the County/Route/Postmile - LRS corporate data set to resolve. • Resolve data issues to improve accuracy, timeliness, and availability of County/Route/Postmile data and GIS LRS data <p>Business Opportunities/Problems:</p> <ul style="list-style-type: none"> • Increase the accuracy of County/Route/Postmile data • Reduce duplicate LRS systems in use by Caltrans • Improve the process for updating County/Route/Postmile data • Improve the relationship between County/Route/Postmile data and GIS LRS data • Improve the quality assurance processes related to updating County/Route/Postmile data <p>Expected Results:</p> <ul style="list-style-type: none"> • Increased accuracy and timeliness of updates to County/Route/Postmile data • Reduced duplication of GIS LRS systems • Improved quality assurance processes
Known Implementation Risks/Issues	<ul style="list-style-type: none"> • Staff availability may be limited because of competing priorities and workload • Legal and policy requirements related to collision data may limit alternatives for short-term improvements to updating County/Route/Postmile data • Current users of alternate GIS LRS systems must have their needs resolved before limiting the use of alternate systems
Known Mitigation Factors	<ul style="list-style-type: none"> • Manage project scope so as not to overly burden project team member schedules • Understand legal and policy limitations up front so project activities are not negatively impacted • Confer at length with owners and users of alternate LRS systems to ensure that their needs are met before implementing changes to LRS system availability and use.
Expected Benefit	<ul style="list-style-type: none"> • Improved accuracy and timeliness of County/Route/Postmile data • Improved accuracy and consistency of Caltrans GIS LRS data
Dependencies With Other Data Governance Projects or Activities	<ul style="list-style-type: none"> • Development of TAC needed to precede development of data governance project teams
Project Sponsor or Business	Mark Samuelson, TSI

Item	Description
Owner	
Key Project Stakeholders or Communities of Interest	<ul style="list-style-type: none"> • TSI • Division of Maintenance • Division of Transportation Planning • Division of Information Technology • All users of location and LRS information
Project Manager	<i>To be determined</i>
Data Custodian(s)	<i>To be determined</i>
Additional Roles / Responsibilities	<ul style="list-style-type: none"> • Data Governance Board – Review and approve project recommendations • TAC – Review and revise project recommendations
Key Milestone Dates	<p>October 2011 – Project kickoff January 2012 – Research complete March 2012 – Present initial recommendations April 2012 – TAC review – revise as necessary May 2012 – Board review – revise as necessary June 2012 – Recommendations approved</p>
Project Duration	October 2011 to June 2012
Cost to Implement	<i>To be determined</i>

Appendix Q Project Summary – Annual Average Daily Traffic (AADT) Pilot

Item	Description
Project Title	Annual Average Daily Traffic (AADT)
Project Description	<p>Objectives:</p> <ul style="list-style-type: none"> • Select key data issues related to the AADT corporate data set to resolve. • Resolve data issues to improve data gathering for state and local AADT and reduce duplicate data entry <p>Business Opportunities/Problems:</p> <ul style="list-style-type: none"> • Reduce duplicate data entry • Increase opportunities for automated data gathering • Improve processes for gathering local agency data • Improve ability to respond to FHWA requirements <p>Expected Results:</p> <ul style="list-style-type: none"> • Improved data gathering and data entry • Work with local agencies to improve local AADT data • Better response to FHWA
Known Implementation Risks/Issues	<ul style="list-style-type: none"> • Automated traffic counting through PeMS is constrained by the extent of PeMS coverage • Local agencies don't have resources to provide needed data • Traffic census limited by availability of staff and resources to complete on time
Known Mitigation Factors	<ul style="list-style-type: none"> • Work with local agencies to explore opportunities for improved data gathering • Submit BCP for increased traffic census resources • Improve PeMS maintenance to improve coverage of operational PeMS locations
Expected Benefit	<ul style="list-style-type: none"> • Increased availability of automated data for use in AADT calculations • Improved processes for gathering local agency data • Improved data entry processes
Dependencies With Other Data Governance Projects or Activities	<ul style="list-style-type: none"> • Development of TAC needed to precede development of data governance project teams • Changes in County/Route/Postmile update procedures may impact AADT updates, which are interrelated
Project Sponsor or Business Owner	Nick Compin, Traffic Operations

Item	Description
Key Project Stakeholders or Communities of Interest	<ul style="list-style-type: none"> • Division of Traffic Operations • Division of TSI • Division of Transportation Planning • Division of Design • Division of Maintenance
Project Manager	<i>To be determined</i>
Data Custodian(s)	<i>To be determined</i>
Additional Roles / Responsibilities	<ul style="list-style-type: none"> • Data Governance Board – Review and approve project recommendations • TAC – Review and revise project recommendations
Key Milestone Dates	<p>October 2011 – Project kickoff January 2012 – Research complete March 2012 – Present initial recommendations April 2012 – TAC review – revise as necessary May 2012 – Board review – revise as necessary June 2012 – Recommendations approved</p>
Project Duration	October 2011 to June 2012
Cost to Implement	<i>To be determined</i>

Appendix R Project Summary – Create the Technical Advisory Committee (TAC)

Item	Description
Project Title	Create the Technical Advisory Committee (TAC)
Project Description	<p>Objectives:</p> <ul style="list-style-type: none"> • Create a charter for the TAC • Find/nominate candidates for the TAC (by Board) • Determine TAC candidate availability • Select TAC members <p>Business Opportunities/Problems:</p> <ul style="list-style-type: none"> • Involve SMEs in the technical aspects of data governance • Identify Caltrans staff resources available to assist in data governance • Staff is overburdened and may not be available to serve on TAC <p>Expected Results:</p> <ul style="list-style-type: none"> • Establish a cross-functional TAC representing various Caltrans divisions and districts • Begin periodic meetings by November 2011
Known Implementation Risks/Issues	<ul style="list-style-type: none"> • Staff availability may be limited because of competing priorities and workload • Lack of familiarity of data governance may limit willingness to participate on TAC
Known Mitigation Factors	<ul style="list-style-type: none"> • Keep time commitment of TAC members limited to reduce impact on workload • Implement education/training elements of Change Management for potential TAC members
Expected Benefit	<ul style="list-style-type: none"> • Creating TAC will provide staff support for Data Governance Board • Provide technical expertise in proposing and assessing data governance projects
Dependencies With Other Data Governance Projects or Activities	<ul style="list-style-type: none"> • Other data governance projects are dependent on the existence of the TAC to provide project overview and guidance, and to support the ongoing activities of the Data Governance Board
Project Sponsor or Business Owner	Coco Briseno, Chief, Division of Transportation System Information
Key Project Stakeholders or Communities of Interest	<ul style="list-style-type: none"> • Data Governance Board members • Deputy Director Marty Tuttle
Project Manager	<i>To be determined</i>
Data Custodian(s)	N/A

Item	Description
Additional Roles / Responsibilities	Data Governance Board – to nominate and select TAC members
Key Milestone Dates	<ul style="list-style-type: none">• Develop TAC charter – September 2011• Nominate TAC members – October 2011• Select TAC members – November 2011• First TAC meeting – November 2011
Project Duration	September 1, 2011 to November 30, 2011
Cost to Implement	0.2 PY (approx. 350 hours); \$0

Appendix S Project Summary – Data Governance Website Mockups

Item	Description
Project Title	Create Data Governance Website
Project Description	<p>Objectives:</p> <ul style="list-style-type: none"> • Create a Data Governance Website to support data governance communication and implementation throughout Caltrans • Provide important data governance information and training via the website <p>Business Opportunities/Problems:</p> <ul style="list-style-type: none"> • Provide an easy one-stop-shop for new and experienced users of data governance to find data governance information at Caltrans • Provide storage of key data information for reference by Caltrans employees <p>Expected Results:</p> <ul style="list-style-type: none"> • A data governance website created by the end of June 2012
Known Implementation Risks/Issues	<ul style="list-style-type: none"> • Lack of IT staff availability may delay/inhibit the conduct of this project • Sufficient hardware and software resources may not be available • Information/content intended for website may not be available at the time the website is live
Known Mitigation Factors	<ul style="list-style-type: none"> • Provide resources (PY's and dollars) from Data Governance Program participants to provide any personnel, hardware, or software needs • Provide early notification of the information/content needs of the website to encourage owners of the information to provide it in a timely fashion
Expected Benefit	Easy access to data governance information by new and experienced data governance users at one central repository of data governance information
Dependencies With Other Data Governance Projects or Activities	Development of TAC needed to precede development of data governance project teams
Project Sponsor or Business Owner	Division of TSI and Division of Information Technology
Key Project Stakeholders or Communities of Interest	All Caltrans divisions and districts who will be consumers of data governance information
Project Manager	<i>To be determined</i>
Data Custodian(s)	<i>To be determined</i>
Additional Roles / Responsibilities	<ul style="list-style-type: none"> • Data Governance Board – Review and approve project recommendations

Item	Description
	<ul style="list-style-type: none"> • TAC – Review and revise project recommendations • IT Governance Board – review and approve any technical IT system or infrastructure decisions, including funding decisions
Key Milestone Dates	<p>October 2011 – Project kickoff January 2012 – Design website March 2012 – Complete website April 2012 – Test website (system and user) May 2012 – Revise and update website June 2012 – Go live</p>
Project Duration	October 2011 to June 2012
Cost to Implement	<i>To be determined</i>

Mockup 1 – About Caltrans Data Governance

This mockup represents a home page for the Caltrans Data Governance Program and provides a broad overview of the purpose of data governance. It includes:

- A definition of data governance
- A briefing of who governs Caltrans data
- A statement of the mission and goals of data governance
- Links to the most current Transportation System Data Governance Board Charter

The data governance website should be an internal website that is available via the Caltrans On-Ramp and could be linked to Division websites as well.

Caltrans Data Governance Mockup Screenshot

Home | [Travel](#) | [Business](#) | [Engineering](#) | [News](#) | [Maps](#) | [Jobs](#) | [About Caltrans](#) | [Contact Us](#)

Search Caltrans

Home | [About Caltrans Data Governance](#) | [Org Chart](#) | [Site Map](#)

About Caltrans Data Governance

In 2011, Caltrans initiated a Data Governance program to address Transportation System Data issues.

What is Data Governance?

"Data Governance is the exercise of decision-making and authority for data-related matters." (from the Data Governance Institute)

Who Governs Caltrans Transportation System Data?

The Transportation System Data Governance Board (TSDGB) is the decision-making and policy-making authority for matters related to transportation system data for all divisions and districts at Caltrans.

Mission: The Caltrans Transportation System Data Governance Board ensures that Caltrans creates and maintains reliable transportation system data that is accessible to Caltrans and its partners.

Transportation system data is data related to the definition, condition, or usage of the physical transportation system, including all transportation modes. It does NOT include administrative data such as financial system data, human resources data, procurement data, and data related to the operations of administrative functions

Data Governance Goals

#	Goal Title	Goal Description
1.	Leadership	Champion data solutions, ensure accountability, and increase the value of data assets
2.	Quality	Oversee efforts to provide high quality data that is accurate, clear, and easy to access
3.	Prioritization	Prioritize efforts to address data gaps and needs
4.	Cooperation	Facilitate cross-organizational collaboration, data sharing, and integration (break down barriers between business units, reduce data silos)
5.	Flexibility	Encourage creative and innovative solutions to data needs

[Click Here to view the Board Charter](#)
[Click Here to view the Board Organization Chart](#)

HIGHLIGHTS
-> **new**
->> Corporate Data Sets
->> VMT Data Update
->> GIS Data Governance
->> Upcoming Efforts
->> Submit an Issue

RELATED LINKS
->> TSI Home
->> Data Governance Inst.
->> XXXX
->> XXXX

Mockup 2 – Data Governance Organizational Chart

This mockup presents the organizational structure for data governance. This will be important, especially in the first year of the effort because Caltrans staff will be unfamiliar with the organization and the responsibilities of new individuals and groups (e.g. Data Custodians and Communities of Interest).

- The data governance organizational chart
- Listing of the Board Members
- Listing of the TAC members with links to launch emails to that individual
- Links to the most current Board Charter
- Links to a definition page of the roles and responsibilities of each group in the organizational chart

Data Governance Organizational Chart Mockup Screenshot

The screenshot shows a web browser window displaying the Caltrans website's organizational chart for data governance. The page title is "Data Governance Organization Chart". The main content area features a hierarchical diagram and two lists of board members.

Caltrans Directorate

- SHOIPP Executive Committee
- Financial Policy Board
- Transportation System Data Governance Board
- IT Governance Board

Transportation System Data Governance Board

- Technical Advisory Board
- Data Custodians
- Communities of Interest
- Business Owners

Transportation System Data Governance Board Members

- Board Chairperson - Coco Briseno - Division of Transportation System Information
- First Name Last Name - Division of XXXX
- First Name Last Name - Division of XXXX
- First Name Last Name - Division of XXXX
- First Name Last Name - Division of XXXX
- First Name Last Name - Division of XXXX

Technical Advisor Committee (Click name to contact/email)

- First Name Last Name - Division of XXXX
- First Name Last Name - Division of XXXX
- First Name Last Name - Division of XXXX
- First Name Last Name - Division of XXXX
- First Name Last Name - Division of XXXX

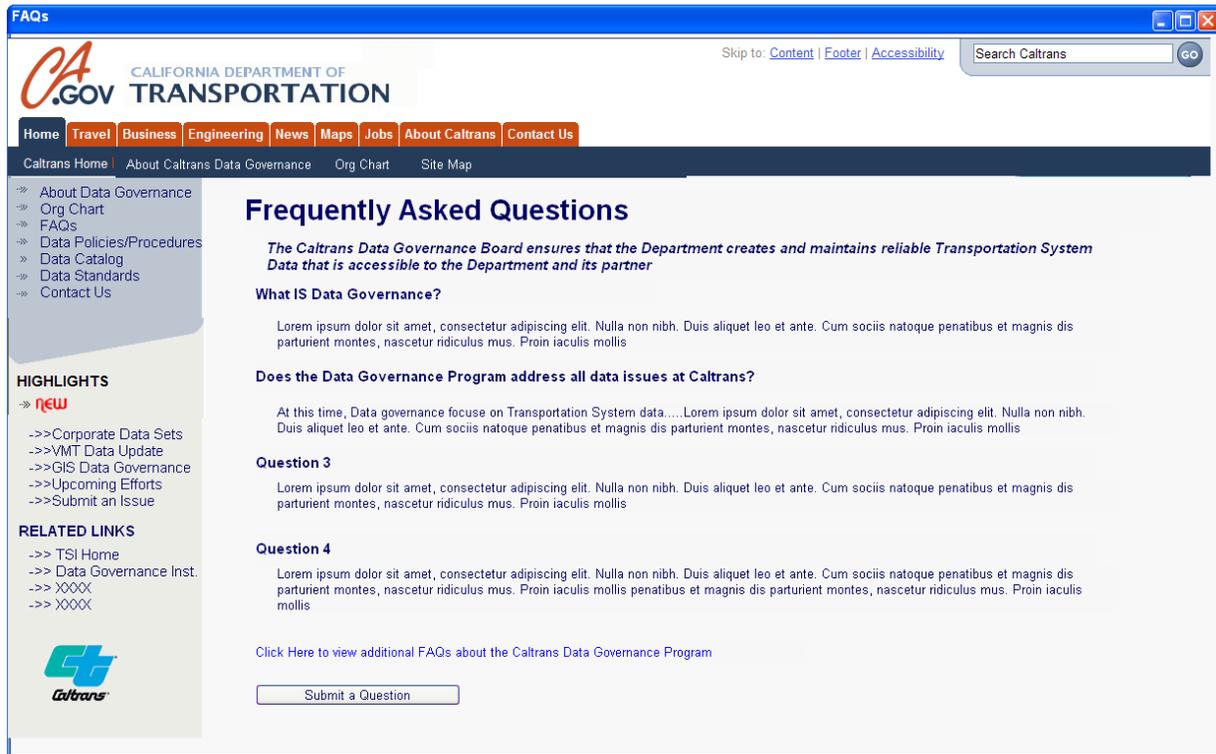
Click Here to view the Board Charter
Click Here to view additional Data Governance roles nad responsibilities

Mockup 3 – Frequently Asked Questions

This mockup provides the most frequently asked questions (FAQs) about the Data Governance Program. The content for FAQs can be determined by actual questions submitted to the Board, and/or those questions which the Board and the TAC feel that would likely come up given the new nature of data governance. This page includes:

- The most commonly asked questions
- Ability to either scroll through a long list or a link to a multi-page list of FAQs
- Ability to submit a question

Frequently Asked Questions Mockup Screenshot



Mockup 4 – Data Governance Policies and Procedures

As the Board determines policies, procedures, and tools to guide the use of data within Caltrans, this web page provides a single location for Communities of Interest to access the latest information. This mockup provides two information components:

- A scrolling list of Data Governance policies – users would see policy names and perhaps policy dates, and use this feature to select a policy to view. Once selected, a Word document or PDF would display the policy. Policies would be formatted in the standard format provided in *Appendix G – Data Governance Policy Form*.
- A scrolling list of Data Governance procedures and tools – users would be presented with a list of procedures and tools that have been developed to support data governance. For example, the standard template for performing data quality assessments could be located here. Once selected, the user would be presented with a file in the appropriate format for use.

Data Governance Policies and Procedures Mockup Screenshot

Data Policies

CA.GOV CALIFORNIA DEPARTMENT OF TRANSPORTATION

Skip to: [Content](#) | [Footer](#) | [Accessibility](#) Search Caltrans

Home Travel Business Engineering News Maps Jobs About Caltrans Contact Us

Caltrans Home | About Caltrans Data Governance Org Chart Site Map

→ About Data Governance
→ Org Chart
→ FAQs
→ Data Policies/Procedures
→ Data Catalog
→ Data Standards
→ Contact Us

Data Governance Policies and Procedures

Header for this page..... Search the lists below to find and review established Data Governance Policies or to view and use Data Governance process, forms and tools

Data Governance Policies

- Data Governance Business Plan
- Policy 2
- Policy 3
- Policy 4
- Policy 5
- Policy 6 **New! added 5/20/2012**

Click on any link above to review selected Policy

Procedures/Tools

Type	Title
Process	Submitting Issues for Board Consideration
Form	Issue Submission Form
Form	Data Governance Project Status Report
Procedure	Procedure 4
Tool	Data Quality Assessment Checklist
Process	Process 6

Click on any link above to review process or procedure

HIGHLIGHTS
→ **NEW**

->>Corporate Data Sets
->>VMT Data Update
->>GIS Data Governance
->>Upcoming Efforts
->>Submit an Issue

RELATED LINKS
->> TSI Home
->> Data Governance Inst.
->> XXXX
->> XXXX

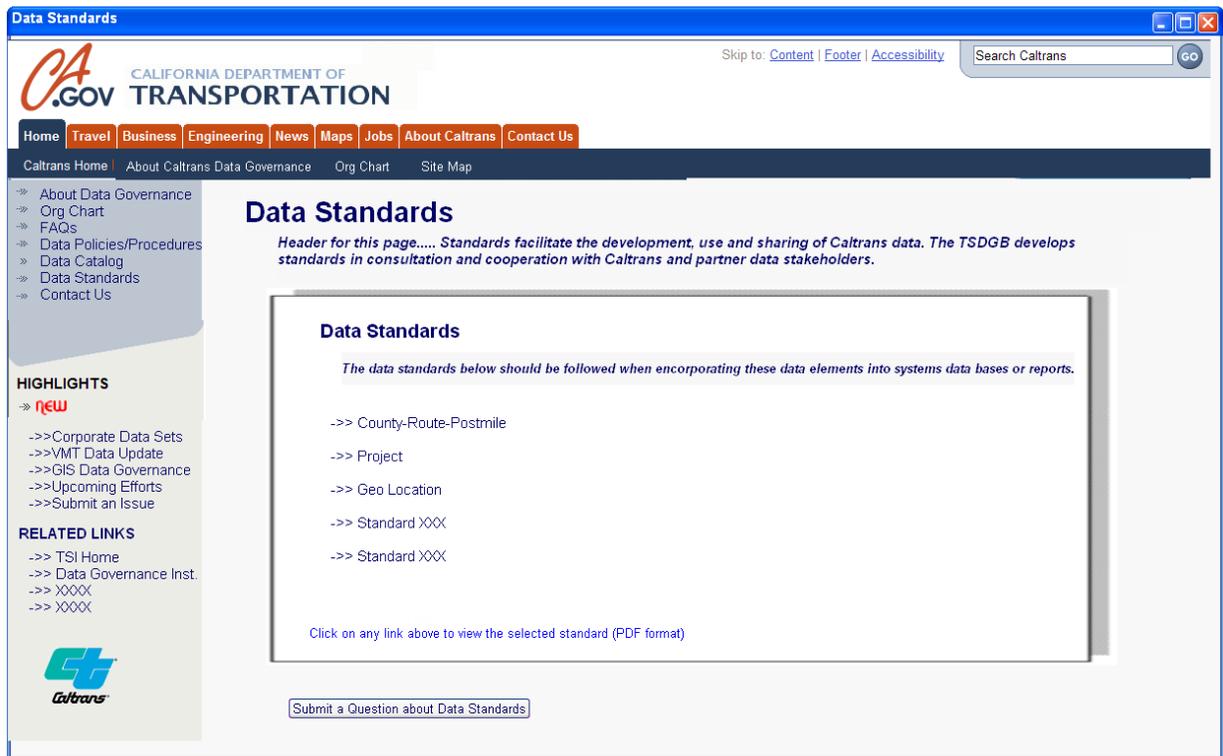
Mockup 5 – Data Standards

This mockup provides a single location for all data standards. This concept supports all Communities of Interest who want to know data set or data element standards. Information located here would include:

- Data standards that already exist at Caltrans
- Data standards developed by the Data Governance Program
- Data standards developed by other efforts within Caltrans and have received approval by the Board as appropriate for Caltrans' corporate use

Users may scroll through a list of standards and select the appropriate standards for viewing.

Data Standards Mockup Screenshot



Mockup 6 – Data Set Catalog

This mockup provides Data Set Catalog information to all Caltrans users. It can be used as the central location for all Communities of Interest to locate and access:

- Approved metadata for key data sets, along with key information such as source system, owner, and maintenance frequency.
- Actual data that has been approved for use by Caltrans users, for example:
 - An approved set of transit data
 - A valid set of VMT information by fiscal year

Data Set Catalog Mockup Screenshot

Data Set Catalog

Header for this page..... The data catalog provides links to available data structures as well as data sets that meet the quality standards of the Caltrans Data Governance Program

Category	Title	Description	Originator	Contact	Publication Date	Online link
Category	Sample Title	Data Set Description	xxxxxx	Jim Jones	99/99/99	www.xxx.com
Category	Sample Title	Data Set Description	xxxxxx	Jim Jones	99/99/99	www.xxx.com
Category	Sample Title	Data Set Description	xxxxxx	Jim Jones	99/99/99	www.xxx.com
Newly Added!						

Click on any link above to view the data element catalog for the selected data set

Submit a Question

Mockup 7 – Data Element Catalog

This mockup provides Data Element Catalog information to all Caltrans users. It can be used as the central location for all Communities of Interest.

Data Element Catalog Mockup Screenshot

The screenshot shows a web browser window titled "Data Element Catalog". The page header includes the California Department of Transportation logo and navigation links for Content, Footer, and Accessibility. A search bar is located in the top right corner.

The main navigation menu includes links for Home, Travel, Business, Engineering, News, Maps, Jobs, About Caltrans, and Contact Us. Below this, a secondary menu lists "Caltrans Home", "About Caltrans Data Governance", "Org Chart", and "Site Map".

The left sidebar contains a list of links: About Data Governance, Org Chart, FAQs, Data Policies/Procedures, Data Catalog, Data Standards, and Contact Us. Below this is a "HIGHLIGHTS" section with a "new" icon and links for Corporate Data Sets, VMT Data Update, GIS Data Governance, Upcoming Efforts, and Submit an Issue. A "RELATED LINKS" section follows with links to TSI Home, Data Governance Inst., and two placeholder links (XXXX).

The main content area is titled "Data Element Catalog" and features a header: "Header for this page..... The data catalog provides links to available data structures as well as data sets that meet the quality standards of the Caltrans Data Governance Program".

A table displays data elements with the following columns: Long Name, Short Name, Description, Data Type, Length, Valid Values, and Business Rules. The table contains three rows of sample data:

Long Name	Short Name	Description	Data Type	Length	Valid Values	Business Rules
Sample Long Name	Sample Short Name	This is a Description	character	99		
Sample Long Name	Sample Short Name	This is a Description	Number	99		
Sample Long Name	Sample Short Name	This is a Description	Blob	99		

Below the table, there is a "Newly Added!" section and a "Submit a Question" button. A note at the bottom of the table area states: "Click on any link above to view the data element catalog for the selected data set".

Mockup 8 – Corporate Data Sets

This mockup clarifies the definition of corporate data sets. This will be important, especially in the first year of the data governance effort, as Caltrans staff needs to understand what a corporate data set is within the context of the Data Governance Program. This web page may provide:

- The definition for corporate data
- Corporate data sets currently selected by the Board for data governance efforts
- A link to a Caltrans Earth display of corporate data

Corporate Data Sets Mockup Screenshot

The screenshot shows a web browser window titled "Corporate Data Sets". The page header includes the California Department of Transportation logo (CA.GOV) and navigation links: Home, Travel, Business, Engineering, News, Maps, Jobs, About Caltrans, and Contact Us. A search bar is located in the top right corner.

The main content area is titled "Corporate Data Sets" and features a header: "Header for this page.... The initial focus of the Data Governance Program is to address the highest priority corporate data sets for Transportation System Data."

What is Corporate Data?

Characteristics of high priority corporate data:

- Impacts multiple divisions or programs
- Answers frequently asked and important questions for the Department
- Required or requested by control agencies and/or partners
- Data discrepancies cause difficult issues across the Department
- High visibility Department data

Highest Priority Data Sets as Selected by the Data Governance Board

- >> AADT - Average/Annual Daily Traffic
- >> VMT - Vehicle Miles Traveled
- >> Pavement Conditions

Click on corporate data links to see latest Corporate Data Set project updates

The left sidebar contains a "HIGHLIGHTS" section with a "new" icon and a list of links: Corporate Data Sets, VMT Data Update, GIS Data Governance, Upcoming Efforts, and Submit an Issue. Below this is a "RELATED LINKS" section with links to TSI Home, Data Governance Inst., and two placeholder links (XXXX).

The Caltrans logo is visible in the bottom left corner of the page.

Appendix T Data Governance Change Management/Communication Plan Content

	Communication or Activity	Audience	Purpose	Priority	Frequency	Medium	Owner	Contributions
Awareness	Introduce Data Governance Program	Executive Team	Educate executive team, and gain buy-in, convey business objectives and expected results first year.	High	1x	Presentation at Director's Meeting	Board Chair Person	Data Governance Board
	Introduce Data Governance Program	Caltrans General (All Staff)	Educate staff, inform, convey business objectives and expected results first year. <i>Answers: What's in it for Me?</i>	High	1x	Director's Video or newsletter and/or Caltrans Quarterly Newsletter	Director/ Executive Staff	Chair Person/ Data Governance Board
	Training - Intro to Data Governance	New Communities of Interest (including new Data Custodians, or those with new data related responsibilities)	Provides an overview of the goals, objectives, roles, responsibilities and initiatives of the Data Governance Program.	Med	As Needed 1-2 x annually	Recorded webinar/ PowerPoint presentation	TAC	Data Governance Board

	Communication or Activity	Audience	Purpose	Priority	Frequency	Medium	Owner	Contributions
Desire and Motivation	Introduce Data Governance for Adopters	Project stakeholders that will begin to use Data Governance	Provide executive perspective on the benefits for Caltrans, divisions, and projects/ Individuals of adopting data governance standards. Describe activities and timing of the effort.	High	For each new group of adopters/ participants in Data Gov. Projects	Data governance kickoff meeting within a division	Division Executive Board Member	Data Governance Board/ TAC
	Data Governance Accomplishments Highlights	Caltrans General (All Staff)	Provide updates on recent accomplishments and new initiatives of the Data Governance Program. Provide directions to access general data governance information for staff unfamiliar with Data Governance Program.	Med	As Needed, Quarterly	Director's Video Caltrans Magazine	Director/ Exec Staff Public Info Staff	Chair Person/ Data Governance Board

	Communication or Activity	Audience	Purpose	Priority	Frequency	Medium	Owner	Contributions
	Data Governance Accomplishments Highlights – How it Helps	<ul style="list-style-type: none"> • Business Owners • Communities of Interest • Data Custodians 	Gain audience participation by illustrating Data Governance Program success.	Med	2x annually	Newsletter All Staff Meetings	TAC	<ul style="list-style-type: none"> • Data Governance Board • Data Custodians • Data Architect • Division representatives from successful data governance projects
Knowledge and Ability	Data Governance Website	<ul style="list-style-type: none"> • Business Owners • Communities of Interest 	Provide a central location for data governance knowledge repository.	High	On Demand	Develop and Implement a data governance website with access via On-Ramp	TAC or TSI	IT Support
	Training - Standard Processes	<ul style="list-style-type: none"> • Data Custodians • IT System Developers • Caltrans Staff 	This audience must to understand how to interact with the Data Governance Program.	Med	On Demand	Provide documented processes, procedures and tools on the Data Governance website	TAC or TSI	<ul style="list-style-type: none"> • Data Governance Board • TAC • Custodians • TSI • Data Governance Plan

	Communication or Activity	Audience	Purpose	Priority	Frequency	Medium	Owner	Contributions
	Training - Data Standards, Data Set Catalog, Data Element Catalog		Provide valuable information regarding available data, definitions, data standards, policies and projects to data and system efforts.	High	On Demand	Develop and publish data standards, policies and project results on the Data Governance website Maintain the Data Set and Element Catalogs on the website	TAC or TSI	<ul style="list-style-type: none"> • Data Governance Board • TAC • Custodians • TSI • Data Governance Plan
	Frequently Asked Questions		Answers commonly asked questions about the Data Governance Program and how to interact with it.	Med	On Demand	Publish on Data Governance Web Site	TAC	<ul style="list-style-type: none"> • Data Governance Board • Data Governance Plan
Reinforcement	Interactive Training to Provide Support	New adopters – anyone who has just started using the plan or is contemplating adoption	Support new adopters to help them understand standards, processes, and benefit from lessons learned.	High	As Needed	One on one sessions or interactive webinars	TAC	Data Governance Board

Communication or Activity	Audience	Purpose	Priority	Frequency	Medium	Owner	Contributions
Coaching - Understand and Adopt Data Governance Board Policies and Standards	<ul style="list-style-type: none"> • Business Owners • Communities of Interest • Data Custodians • IT System Developers 	Update business areas and project efforts regarding impact of a Data Governance policy or standard.	High	As Needed	<p>Facilitated sessions/ one-on-one meetings with this audience to ensure understanding of impacts and plan for changes</p> <p>Distribute Data Governance Policy Forms (similar to Deputy Directive distribution)</p>	TAC	Data Custodians
Adopt Data Governance Decisions	Data Governance Sponsor and/or Directorate	Brief on accomplishments, and decisions that meet resistance.	High	1x annually	Meeting/ briefing with subsequent resolutions/ actions identified	Data Gov. Board Chairman	<ul style="list-style-type: none"> • Data Governance Board • TAC • Custodians

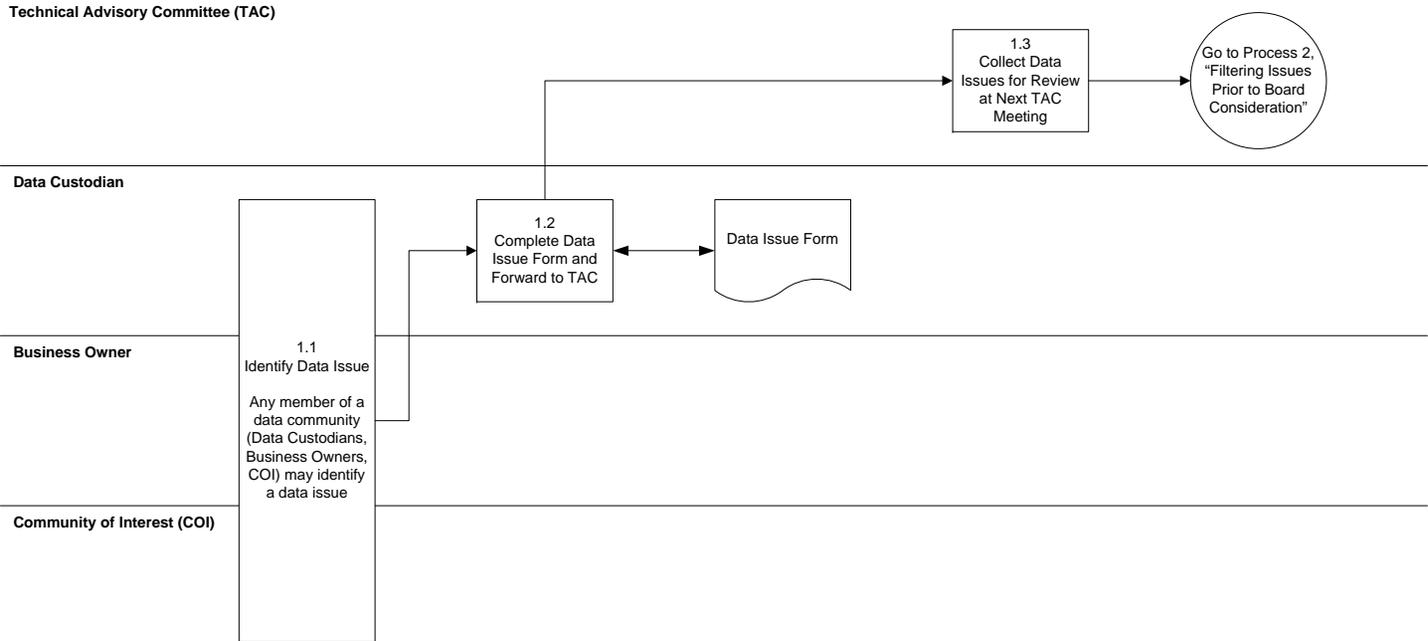
Appendix U Data Governance Process Maps & Narratives

Process 1: Submitting Data Governance Issues for Consideration – Process Map

Transportation System Data Business Plan - Data Governance Processes

Page 1

Process 1: Submitting Data Governance Issues for Consideration



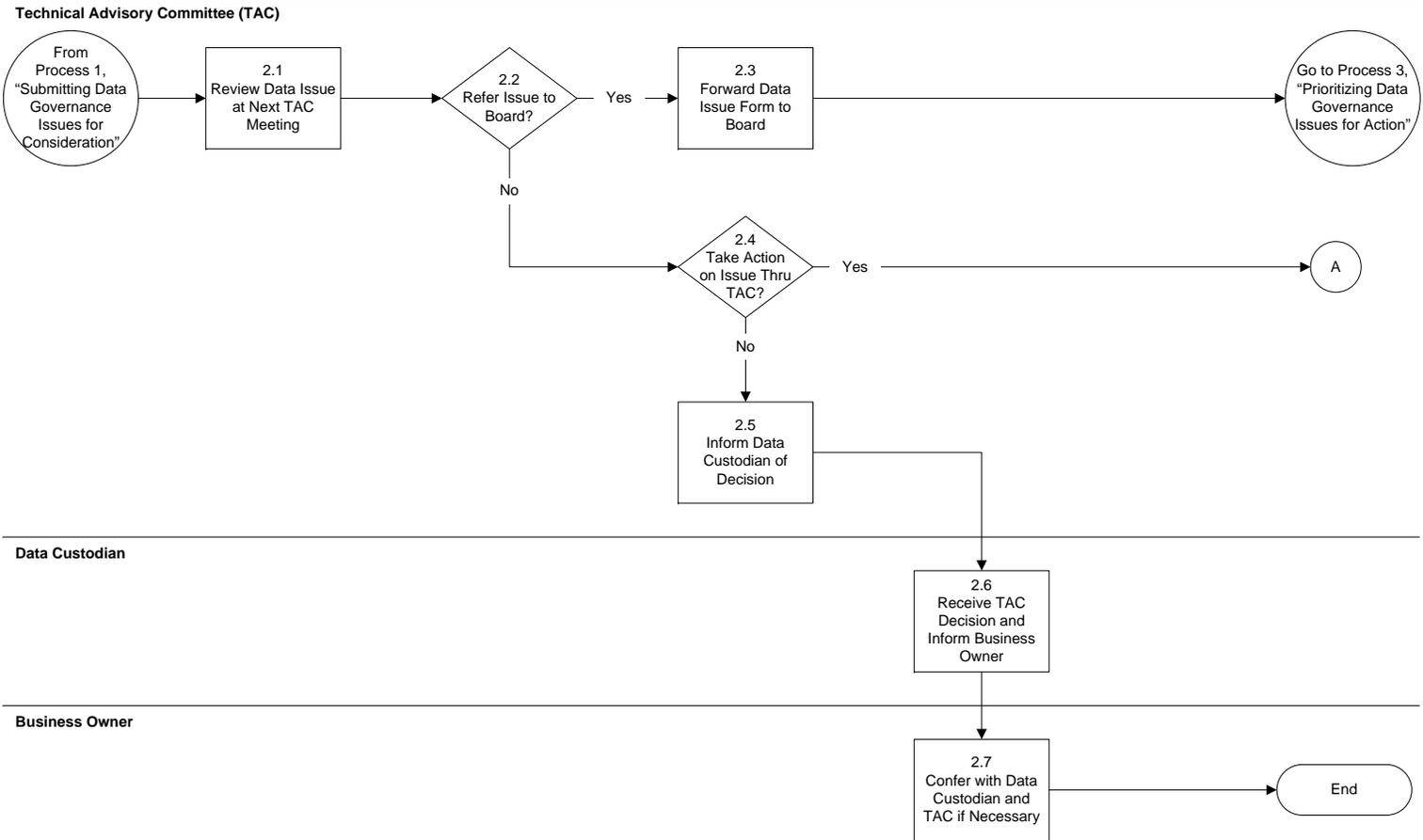
Process 1: Submitting Data Governance Issues for Consideration – Process Narrative

Description		Data issues are identified by a member of a data community (Data Custodian, Business Owner, or Community of Interest) and gathered by the TAC to determine the proper course of action.			
Assumptions		New data issues will be identified that will need to be reviewed by the data governance structure.			
Process Trigger		A new data issue is identified by a member of a data community.			
Completion Indicator		The data issue is sent to the TAC for further consideration.			
#	Activity	Description	Role	Input	Output
1.1	Identify Data Issue	A Data Custodian, Business Owner, or Community of Interest (COI) identifies a data issue.	Data Custodian, Business Owner, Communities of Interest	Identify data issue in their Business Area	None
1.2	Complete Data Issue Form and Forward to TAC	The Data Custodian completes the Data Issue Form and forwards it to the TAC for review.	Data Custodian	Data issue	Completed Data Issue Form
1.3	Collect Data Issues for Review at Next TAC Meeting (Go to Process 2, "Filtering Issues Prior to Board Consideration")	The TAC collects data issues submitted by Data Custodians between TAC meetings so that they can review the data issues at the next TAC meeting and develop recommendations for Board consideration or take other actions.	TAC	Completed Data Issue Forms	None

Process 2: Filtering Issues Prior to Board Consideration – Process Map

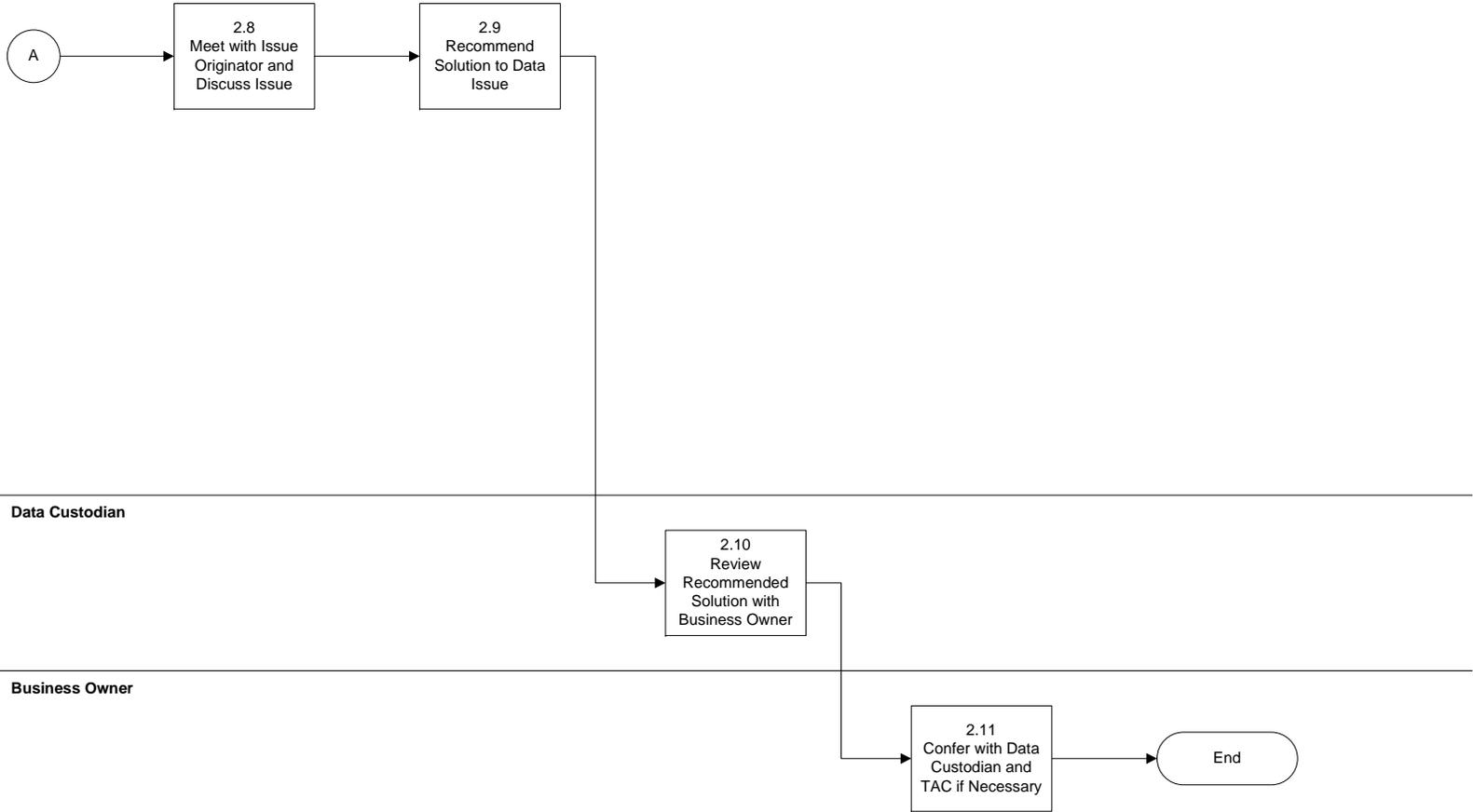
Transportation System Data Business Plan - Data Governance Processes

Process 2: Filtering Issues Prior to Board Consideration



Transportation System Data Business Plan - Data Governance Processes
Process 2: Filtering Issues Prior to Board Consideration

Technical Advisory Committee (TAC)



Process 2: Filtering Issues Prior to Board Consideration – Process Narrative

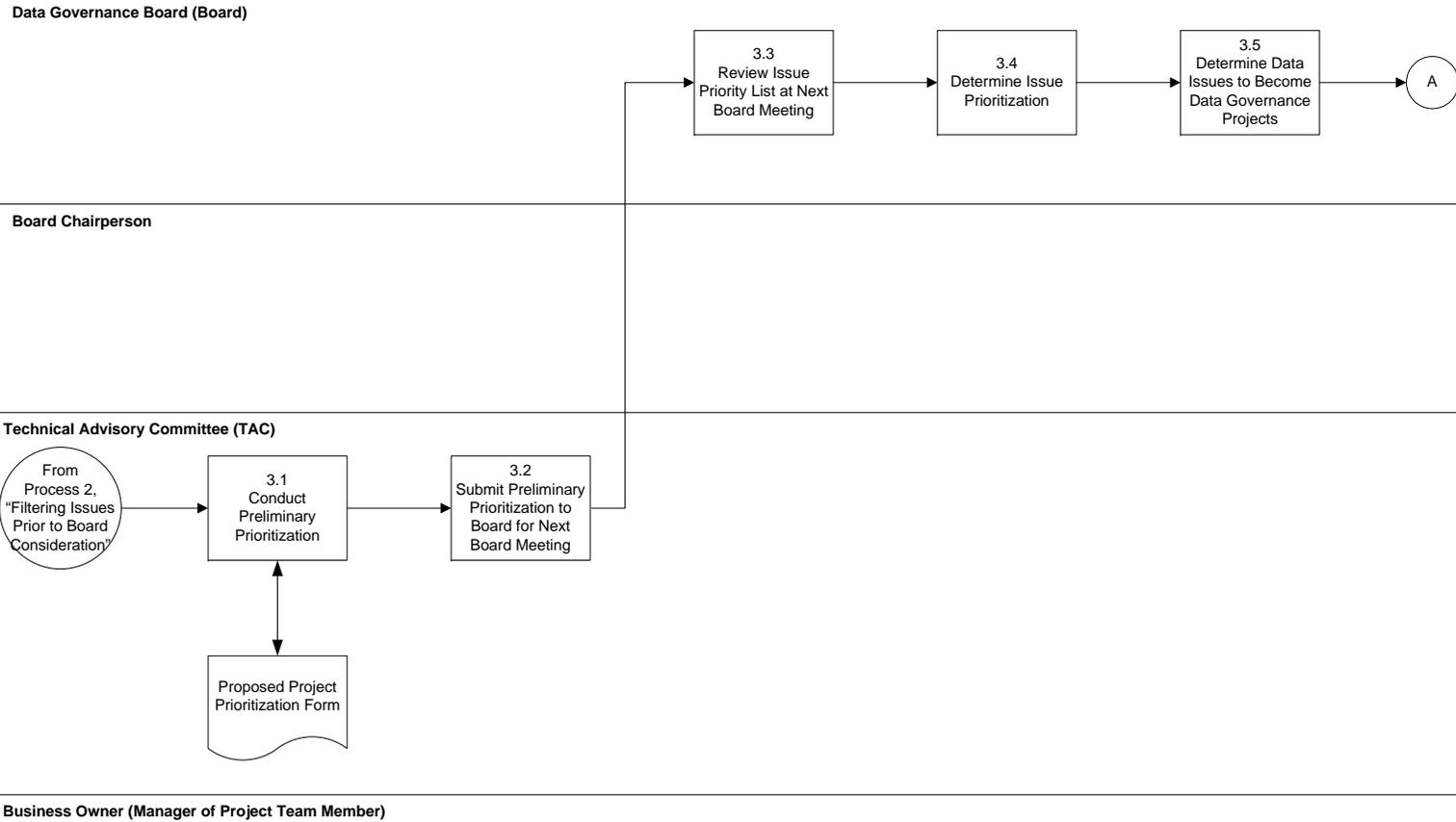
Description		For all data issues identified in Process 1, “Submitting Data Governance Issues for Consideration,” the TAC will determine whether or not each issue should be referred to the Board, should be resolved by the TAC, or should be returned to the originator without further action. The TAC will have the delegated authority to resolve lower level data issues without bringing them forward to the Board for consideration.			
Assumptions		<ul style="list-style-type: none"> • Data issues have been submitted to the TAC with Data Issue Forms • Some data issues will be brought to the Board for consideration • The TAC will have delegated authority to resolve lower level data issues • Some issues will not require action and will be returned to the originator 			
Process Trigger		Data issues are submitted to the TAC by Data Custodians with completed Data Issue Forms			
Completion Indicator		The TAC prepares the data issue for Board consideration, resolves lower level data issues, or returns the data issue to the originator without action.			
#	Activity	Description	Role	Input	Output
2.1	(From Process 1, “Submitting Data Governance Issues for Consideration”) Review Data Issue at Next TAC Meeting	The TAC reviews data issues that are submitted by Data Custodians.	TAC	Data issue and completed Data Issue Form	None
2.2	Decision: Refer Issue to Board?	The TAC determines whether to refer the issue to the Board; if: <ul style="list-style-type: none"> • <u>Yes</u>: Refer the Issue to the Board for their consideration (go to Step 2.3), or; • <u>No</u>: Decide whether or not to take action through the TAC (go to Step 2.4) 	TAC	None	Decision
2.3	Forward Data Issue Form to Board (Go to Process 3, “Prioritizing Data Governance Issues for Action”)	The TAC forwards the Data Issue Form to the Board for consideration at the next Board meeting.	TAC	Data issue and completed Data Issue Form	Data Issue Form provided to Board

2.4	Decision: Take Action on Issue thru TAC?	The TAC determines whether to take action on the issues through the TAC; if: <ul style="list-style-type: none"> • <u>Yes</u>: Take action on the data issue because the issue is within the level of authority delegated to the TAC (go to Step 2.8), or; • <u>No</u>: Determine that no action needs to be taken regarding the data issue (go to Step 2.5) 	TAC	Issue will not be presented to the Board	Decision
2.5	Inform Data Custodian of Decision	The TAC sends notification to the Data Custodian who originated the data issue that the issue will not be considered for action by the Board or the TAC and is being returned without action.	TAC	None	Notification of decision
2.6	Receive TAC Decision and Inform Business Owner	The Data Custodian who originated the Data Issue receives the TAC notification and informs the relevant Business Owner of the TAC decision.	Data Custodian	Notification from TAC	Information provided to Business Owner
2.7	Confer with Data Custodian and TAC if necessary (End of Process)	The Business Owner reviews the TAC decision with the Data Custodian and, if necessary, discusses the decision with the TAC to reach consensus.	Business Owner	Information from Data Custodian	Discussions with Data Custodian and TAC, if necessary
2.8	Meet with Issue Originator and Discuss Issue	The TAC (or representative of the TAC) meets with the Data Custodian and others who originally identified the data issue to discuss how the issue may be resolved by the TAC without going through the Board.	TAC	Data issue	Discussion of potential alternative solutions
2.9	Recommend Solution to Data Issue	The TAC considers alternative solutions to the data issue and recommends a solution for consideration by the originating Data Custodian and Business Owner.	TAC	Potential alternative solutions	Recommend solution to data issue
2.10	Review Recommended Solution and Inform Business Owner	The Data Custodian receives the TAC's recommended solution to the data issue and informs their Business Owner of the TAC's recommendation, along with any analysis from the Data Custodian.	Data Custodian	TAC recommended solution	Data Custodian analysis

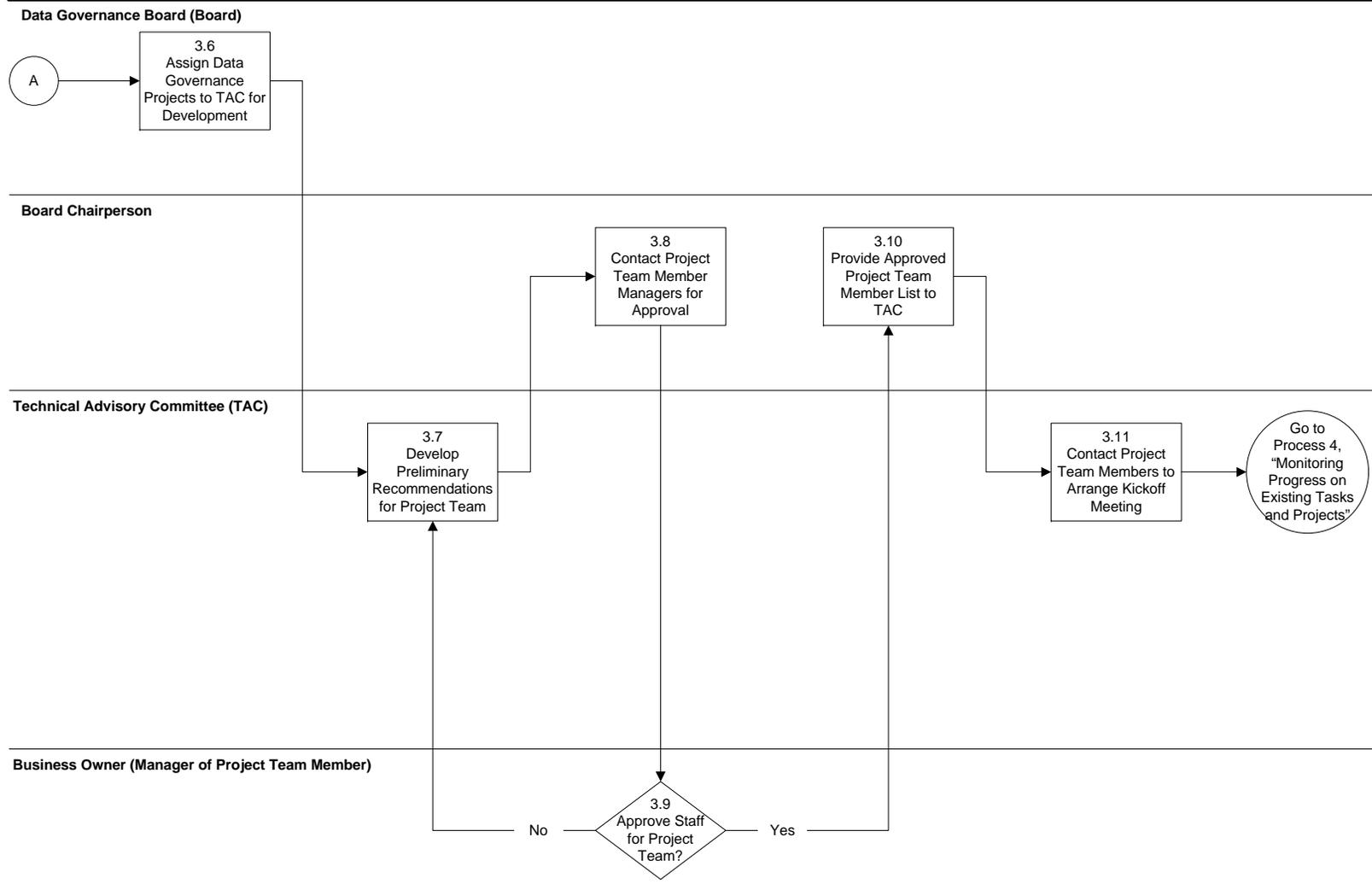
2.11	Confer with Data Custodian and TAC, if necessary (End of Process)	The Business Owner reviews the TAC recommendation with the Data Custodian and, if necessary, discusses the recommendation with the TAC to reach consensus.	Business Owner	Information from Data Custodian	Discussions with TAC, if necessary
------	--	--	----------------	---------------------------------	------------------------------------

Process 3: Prioritizing Data Governance Issues for Action – Process Map

Process 3: Prioritizing Data Governance Issues for Action



Transportation System Data Business Plan - Data Governance Processes
Process 3: Prioritizing Data Governance Issues for Action



Process 3: Prioritizing Data Governance Issues for Action – Process Narrative

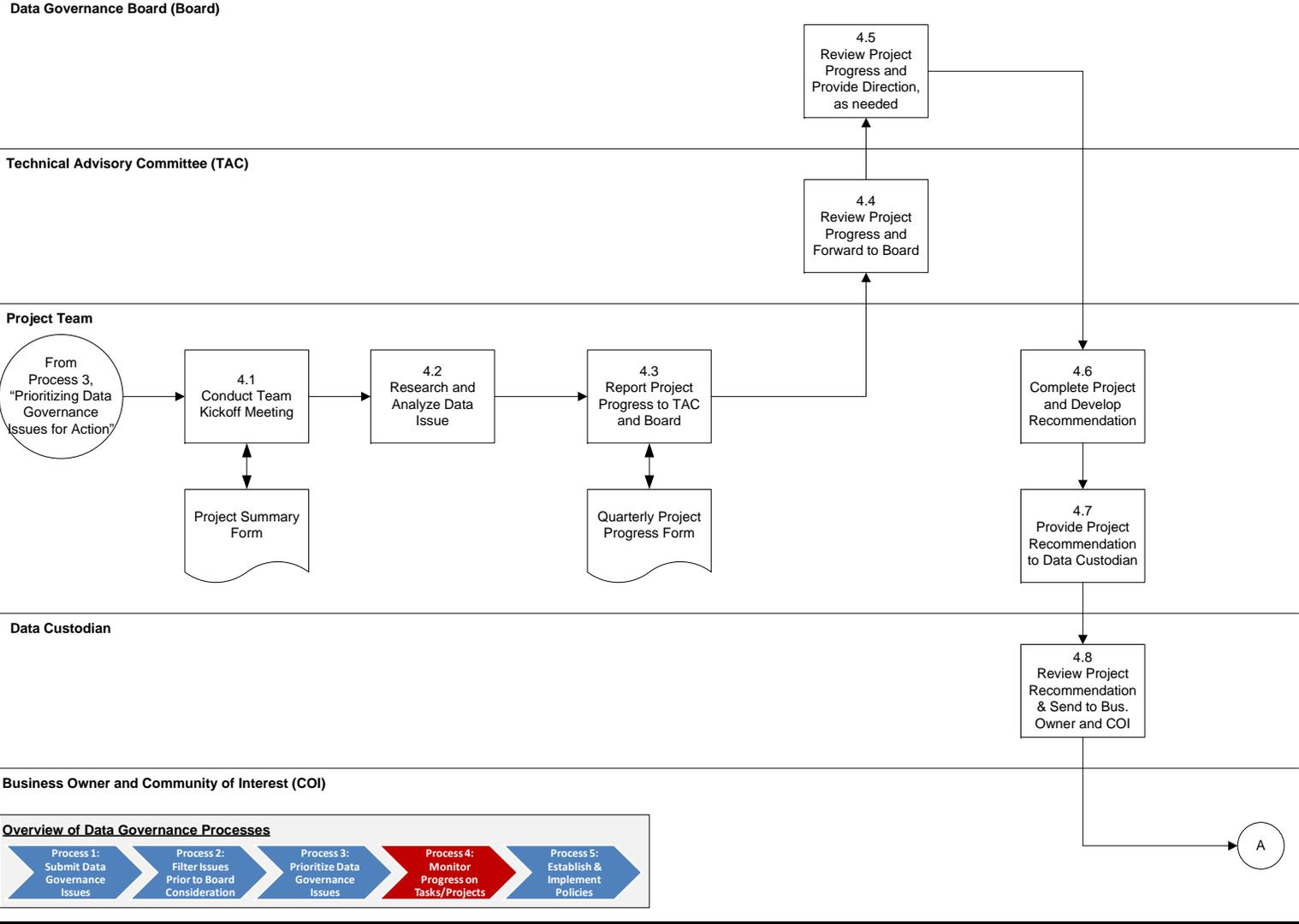
Description	After the TAC reviews the submitted Data Issue Forms and determines that certain data issues will be presented to the Board for consideration, the TAC will conduct a preliminary prioritization of data issues to be presented to the Board. The TAC will complete a Proposed Project Prioritization Form to guide the Board's discussion at the next Board meeting. The Board will determine the final issue prioritization, which issues will become data governance projects, and direct the TAC to begin establishing project teams. The proposed project teams will be reviewed by the Board Chairperson and any impacted Business Owners before the approved project team membership will be confirmed. The project team will then be contacted to kickoff the project.				
Assumptions	<ul style="list-style-type: none"> The TAC has reviewed the data issues submitted by Data Custodians and determined which should be forwarded to the Board. The TAC has resolved lower level issues or returned issues to their originator without action. 				
Process Trigger	The TAC has determined which data issues require Board consideration.				
Completion Indicator	Project teams have been approved by the impacted Business Owners and Board Chairperson.				
#	Activity	Description	Role	Input	Output
3.1	(From Process 2, "Filtering Issues Prior to Board Consideration") Conduct Preliminary Prioritization	The TAC performs a preliminary prioritization of the data issues submitted by Data Custodians and reviewed by the TAC. The TAC completes the Proposed Project Prioritization Form containing the TAC's recommended prioritization of data issues that will become proposed projects to be presented to the Board.	TAC	Data issues for Board consideration	Proposed Project Prioritization Form
3.2	Submit Preliminary Prioritization to Board for Next Board meeting	The TAC provides its recommendations for project prioritization (contained in the Proposed Project Prioritization Form) to the Board for their review in preparation for the next Board meeting. Project prioritization will be a regular agenda item at Board meetings.	TAC	Proposed Project Prioritization Form	None

3.3	Review Issue Priority List at Next Board Meeting	The Board reviews the TAC's recommendations on the Proposed Project Prioritization Form and discusses the TAC's recommendations in preparation for developing the Board's own prioritization list.	Board	Proposed Project Prioritization Form	None
3.4	Determine Issue Prioritization	The Board determines the final proposed project prioritization.	Board	None	Reprioritized proposed project list
3.5	Determine Data Issues to Become Data Governance Projects	<p>The Board determines which of the projects on the final proposed project prioritization list will become data governance projects.</p> <p>Because data governance resources will be limited, particularly in the near term, it will be necessary for the Board to carefully select those projects that will become projects. These projects must be doable, address high visibility data issues, and contribute to the track record of success that can be used to increase the support for data governance at Caltrans.</p>	Board	Reprioritized proposed project list	Projects selected
3.6	Assign Data Governance Projects to TAC for Development	The Board assigns the TAC to begin developing project teams for each project the Board has created.	Board	Projects selected	Direction to TAC to develop project teams
3.7	Develop Preliminary Recommendations for project team	The TAC develops recommendations for subject matter experts to become part of each project team. The TAC forwards its recommendations to the Board Chairperson for review.	TAC	Direction from Board	Preliminary project team membership proposal
3.8	Contact Project Team Member Managers for Approval	The Board Chairperson reviews the TAC recommendations for project team membership and requests approval of each proposed team member's manager.	Board Chairperson	Preliminary project team membership proposal from TAC	Requests to managers for staff participation on project teams

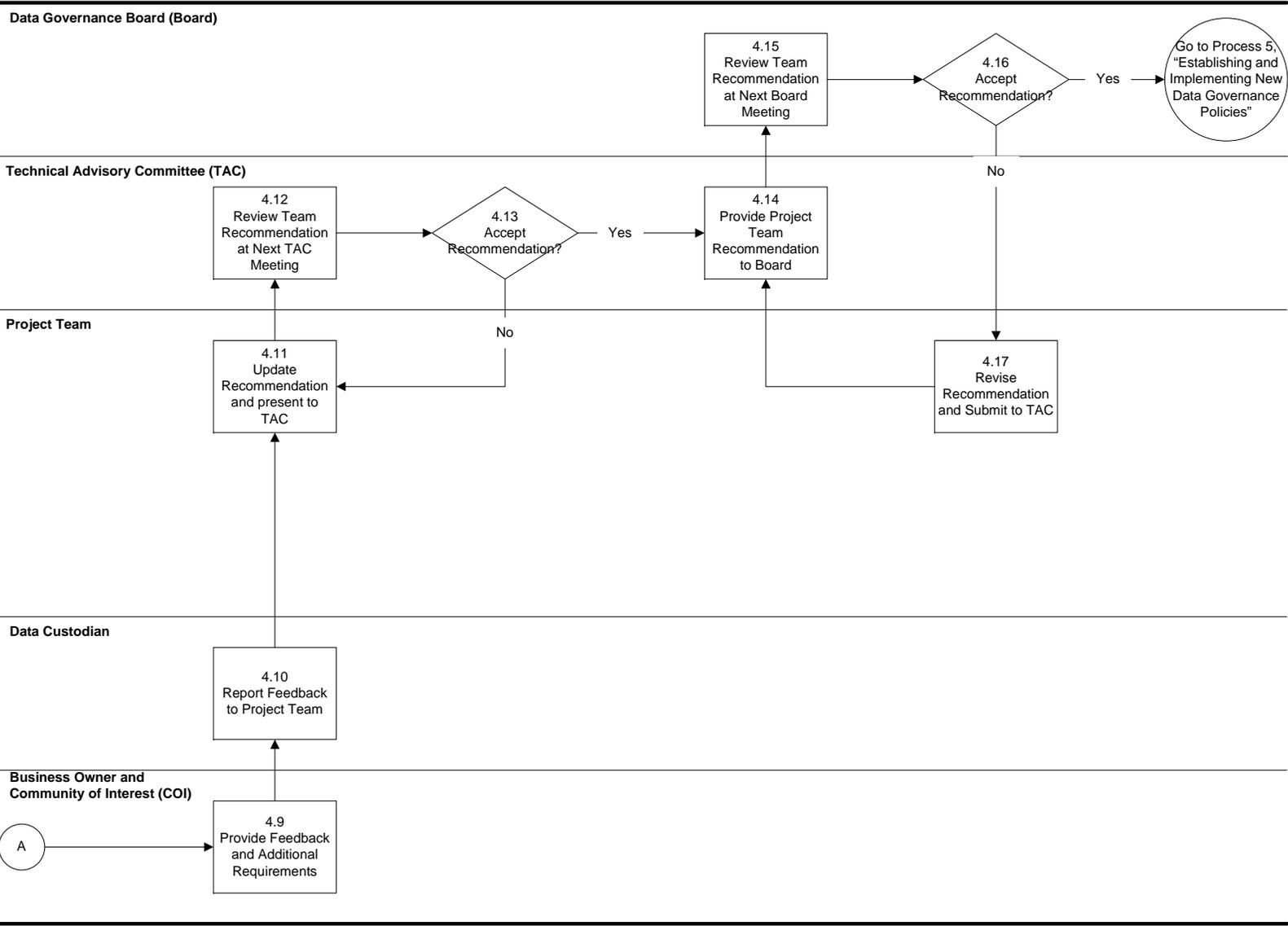
3.9	Decision: Approve Staff for project team?	<p>Each manager of a project team member decides whether or not they can allow their staff to participate on the data governance project team; if:</p> <ul style="list-style-type: none"> • <u>Yes</u>: The proposed project team member can participate on the project team (go to Step 3.10) • <u>No</u>: The proposed project team member is not available to participate on the project team (return to Step 3.7 to develop a recommended replacement for the unapproved project team member) 	Business Owners (managers of proposed project team members)	Requests to managers for staff participation on project teams	Decision to allow staff to participate or not
3.10	Provide Approved project team Member List to TAC	The Board Chairperson, after receiving approval from the impacted managers, informs the TAC of the approved project team list.	Board Chairperson	Decisions from managers on project team participation	Approved project team list
3.11	<p>Contact Project Team Members to Arrange Kickoff Meeting</p> <p>Go to Process 4, “Monitoring Progress on Existing Tasks and Projects”</p>	The TAC contacts all project team members to schedule a project kickoff meeting and provides the team members with the project background and goals.	TAC	Approved project team list	Schedule project kickoff

Process 4: Monitoring Progress on Existing Tasks and Projects – Process Map

Process 4: Monitoring Progress on Existing Tasks and Projects



Transportation System Data Business Plan - Data Governance Processes
 Process 4: Monitoring Progress on Existing Tasks and Projects



Process 4: Monitoring Progress on Existing Tasks and Projects – Process Narrative

Description		Following the creation of the project team, the team will analyze the data issue and periodically report progress to the TAC and the Board. The project team will complete their analysis and provide it to the Data Custodian, Business Owners, and Communities of Interest for review and feedback. The project team will update their recommendation and submit it to the TAC. If accepted by the TAC, the recommendation will be forwarded to the Board for consideration. If accepted by the Board, it may establish a new policy (see Process 5 “Establishing and Implementing New Data Governance Policies”).			
Assumptions		<ul style="list-style-type: none"> • The data issue has been determined to be a priority by the Board and a project team has been created. • There are subject matter experts in Caltrans to address the data issue. 			
Process Trigger		A project team is established to analyze the data issue.			
Completion Indicator		Project team recommendations are accepted by the TAC and Board.			
#	Activity	Description	Role	Input	Output
4.1	(From Process 3, “Prioritizing Data Governance Issues for Action”) Conduct Team Kickoff Meeting	Following the creation of the project team, a project kickoff meeting is held. The project team completes a Project Summary Form to provide the essential background and purpose of the project.	Project Team	Project Summary Form	Kickoff meeting held
4.2	Research and Analyze Data Issue	The project team researches and analyzes all necessary aspects of the data issue.	Project Team	None	Research and analysis of data issue
4.3	Report Project Progress to TAC and Board	The project team reports project progress to the TAC and Board quarterly throughout the course of the project. A Quarterly Project Progress Form will be completed each time progress is reported from the project team. <i>This process step will be repeated until the project is completed.</i>	Project Team	None	Quarterly Project Progress Form

4.4	Review Project Progress and Forward to Board	The TAC reviews the project team's Quarterly Project Progress Form and forwards a summary of the project's progress to the Board for their review. <i>This process step will be repeated until the project is completed.</i>	TAC	Quarterly Project Progress Form	Project summary for Board review
4.5	Review Project Progress and Provide Direction, as needed	The Board reviews the summary of project progress at each Board meeting and provides any additional direction to the TAC and project team, if necessary. <i>This process step will be repeated until the project is completed.</i>	Board	Project summary for Board review	Additional direction to TAC and project team, if necessary
4.6	Complete Project and Develop Recommendation	The project team completes their analysis and develops a recommendation based on their findings.	Project Team	Feedback from TAC and Board	None
4.7	Provide Project Recommendation to Data Custodian	The project team provides their recommendation to the impacted Data Custodian(s) for their review relative to their business area. NOTE: The project team's recommendation needs to be reviewed by the impacted Data Custodians, Business Owners, and Communities of Interest so that their feedback can be considered and incorporated before the recommendation goes forward to the TAC and Board. This is completed in steps 4.8 through 4.12.	Project Team	None	Recommended Alternative
4.8	Review Project Recommendation and Send to Business Owner and Community of Interest (COI)	The Data Custodian(s) reviews the project team's recommendation for impacts to their business area and forwards their analysis to their Business Owner(s) and Communities of Interest	Data Custodian	Recommended Alternative	Analysis of Impact

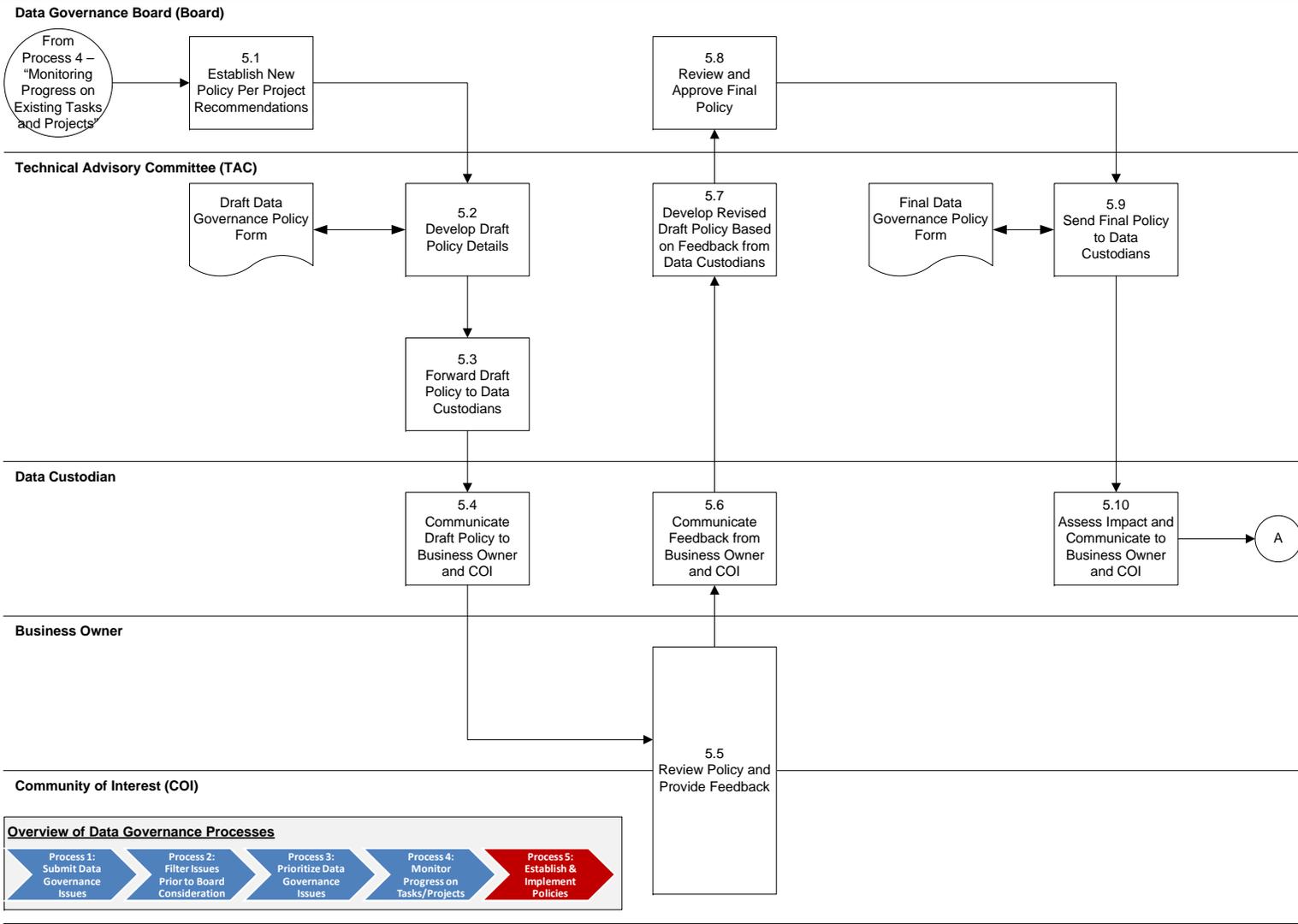
4.9	Provide Feedback and Additional Requirements	The impacted Business Owner(s) and related Community of Interest review the project team's recommendation and the Data Custodian analysis and provide their feedback and requirements to the Data Custodian to pass on to the project team.	Business Owner and Communities of Interest	Project team Recommended Alternative and Data Custodian Analysis	Feedback and Requirements
4.10	Report Feedback to project team	The Data Custodian(s) receives feedback and requirements from the Business Owner(s) and Community of Interest and report this information to the project team for their consideration.	Data Custodian	Feedback and Requirements from Business Owner and Communities of Interest	Provide Information to project team
4.11	Update Recommendation and Present to TAC	The project team updates their recommendation per the feedback and requirements received from the Data Custodians and present the updated recommendation to the TAC.	Project Team	Feedback and Requirements from Business Owner and Communities of Interest	Updated Recommended Alternative
4.12	Review Team Recommendation at Next TAC Meeting	The TAC reviews the project team recommendation at the next TAC meeting.	TAC	Recommended Alternative	None
4.13	Decision: Accept Recommendation?	<p>The TAC determines whether or not to submit the project team recommendation to the Board or to send it back to the project team for reconsideration; if:</p> <ul style="list-style-type: none"> • <u>Yes</u>: Accept project team recommendation (go to Step 4.14) • <u>No</u>: The project team is asked to revise their recommendation per comments from the TAC and resubmit the revised recommendation to the TAC (go to Step 4.11). <p><i>NOTE: Process steps 4.11, 4.12 and 4.13 will be repeated until the TAC accepts the project team recommendation.</i></p>	TAC	None	Decision and TAC comments

4.14	Provide project team Recommendation to Board	The TAC summarizes the project team recommendation and provides it to the Board at the next Board meeting.	TAC	Project Recommended Alternative	Summary of Recommended Alternative
4.15	Review Team Recommendation at Next Board Meeting	The Board reviews the TAC summary of the project team recommendation at the Board's next meeting.	Board	Summary of Recommended Alternative	None
4.16	Decision: Accept recommendation?	<p>The Board decides whether or not to accept the project team recommendation; if:</p> <ul style="list-style-type: none"> • <u>Yes</u>: The project team's recommendation is implemented. If a new data governance policy, procedure, or guideline will be established, go to Process 5, "Establishing and Implementing New Data Governance Policies." • <u>No</u>: Return recommendation to project team with Board's comments and revisions (go to step 4.17) 	Board	None	Decision
4.17	Revise Recommendation and Submit to TAC	<p>The project team revises their recommendation per comments from the Board and resubmits the revised recommendation to the TAC (go to Step 4.14).</p> <p><i>NOTE: Process steps 4.14 through 4.17 will be repeated until the Board accepts the project team recommendation.</i></p>	Project Team	Board Decision	Revised Recommended Alternative

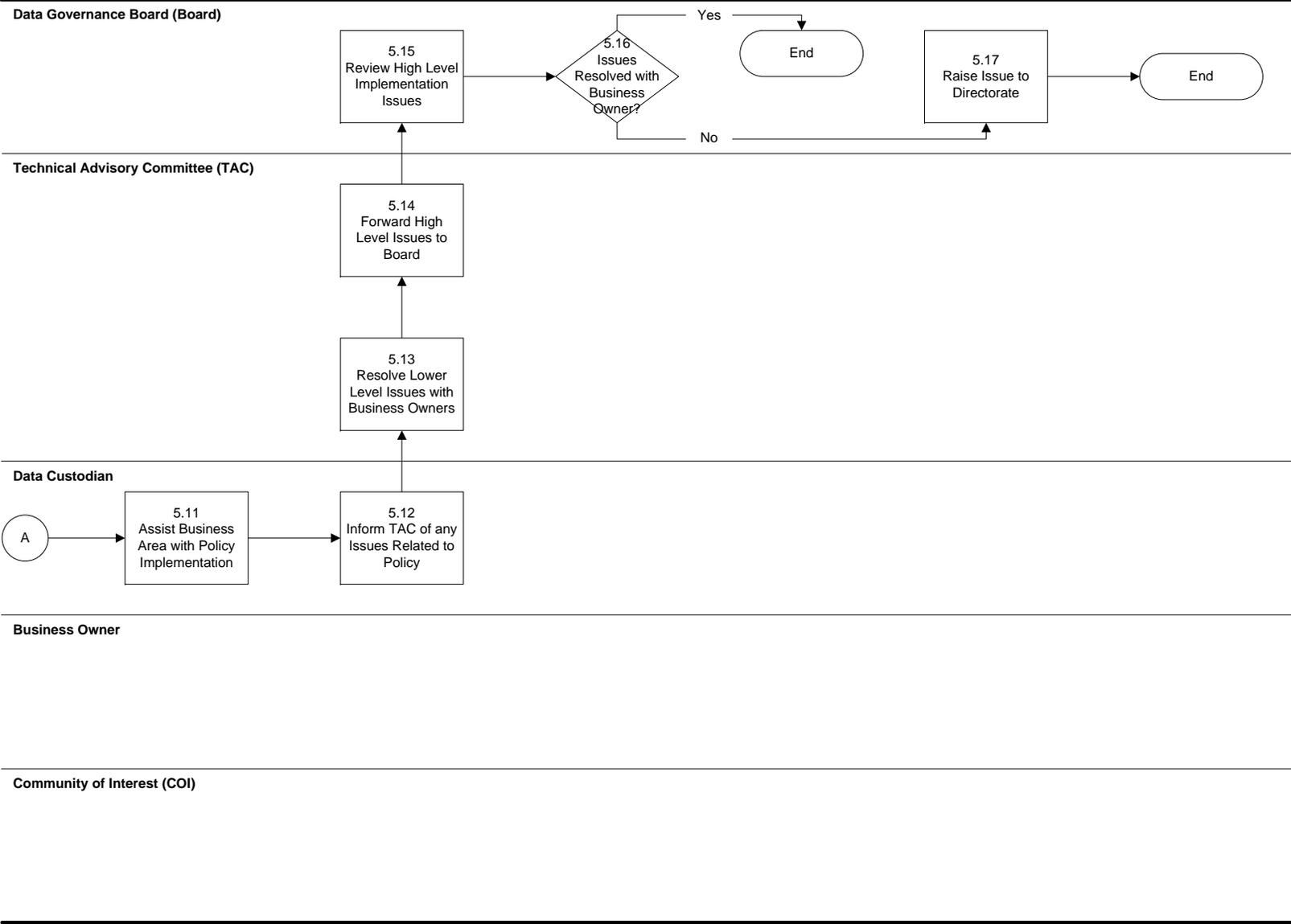
Process 5: Establishing and Implementing New Data Governance Policies – Process Map

Transportation System Data Business Plan - Data Governance Processes

Process 5: Establishing/Implementing New Data Governance Policies



Transportation System Data Business Plan - Data Governance Processes
Process 5: Establishing/Implementing New Data Governance Policies



Process 5: Establishing and Implementing New Data Governance Policies – Process Narrative

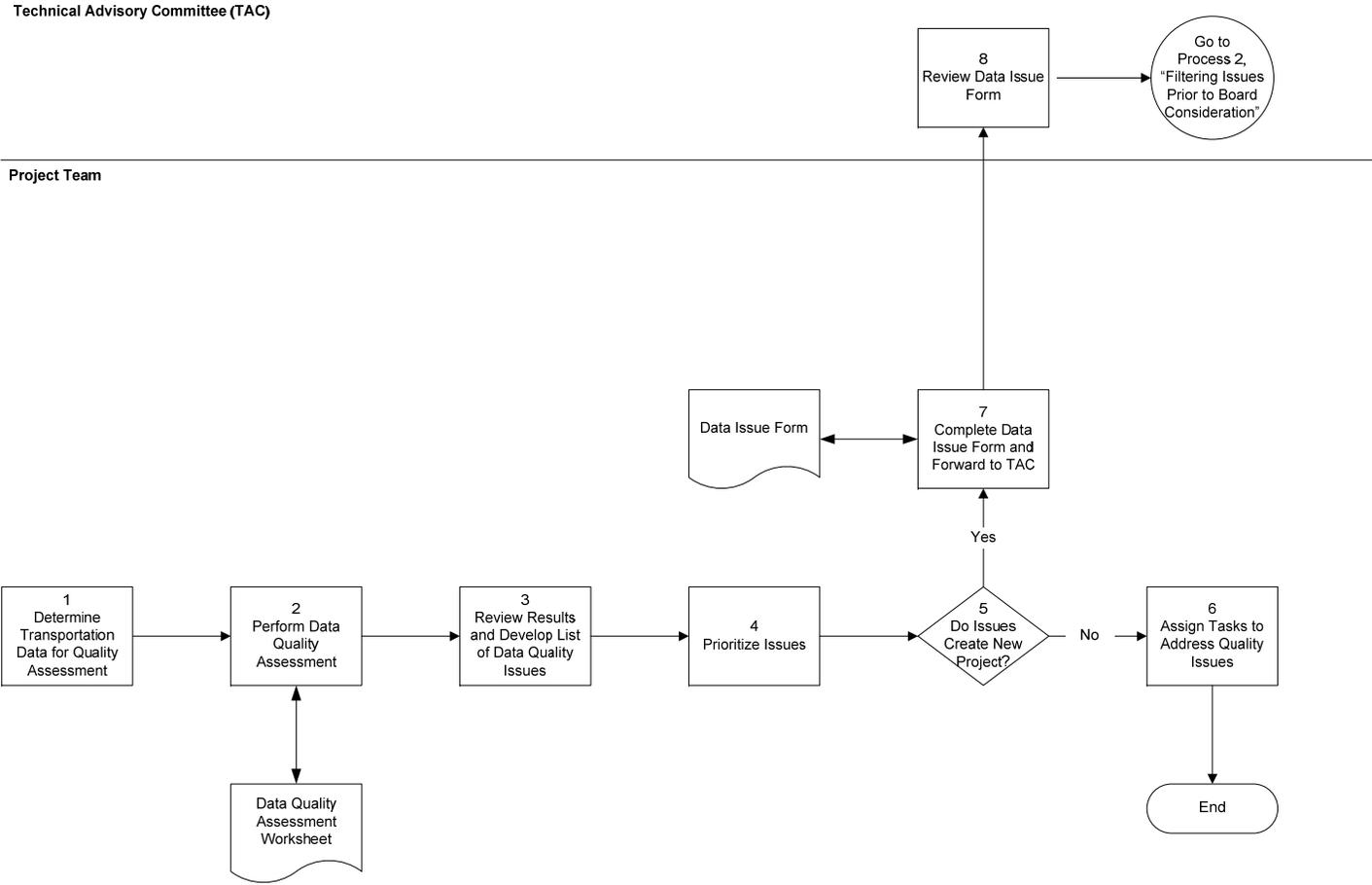
Description		Following the establishment of a new data governance policy, procedure or guideline, there will be implementation and monitoring of the policy, procedure, or guideline, including the development of details and instructions and assistance with implementation. Low-level compliance issues will be resolved by Data Custodians and the TAC. High level compliance issues will be resolved by the Board.			
Assumptions		A new policy, procedure or guideline has been created from a completed data governance project.			
Process Trigger		A data governance project creates a new policy, procedure, or guideline.			
Completion Indicator		Details of new policy, procedure, or guideline have been developed and implemented with issues being resolved at the appropriate level of authority.			
#	Activity	Description	Role	Input	Output
5.1	(From Process 4, “Monitoring Progress on Existing Tasks and Projects”) Establish New Policy Per Project Recommendations	The Board establishes the broad outline of a new policy, procedure, or guideline per the approved recommendations from a data governance project team.	Board	Approved Project	New Policy
5.2	Develop Draft Policy Details	The TAC develops policy and procedure details based on the policy direction established by the Board. The TAC completes a draft Data Governance Policy Form that contains the details of the policy and how it will be implemented.	TAC	New Policy	Draft Data Governance Policy Form
5.3	Forward Draft Policy to Data Custodians	The TAC provides the draft policy information to the impacted Data Custodians for their review.	TAC	None	Draft Data Governance Policy Form
5.4	Communicate Draft Policy to Business Owner and Communities of Interest	The Data Custodians communicate the new policy to their Business Owners and Communities of Interest for their review and feedback.	Data Custodian	Draft Data Governance Policy Form	Communicate to Business Owner, Communities of Interest

5.5	Review Policy and Provide Feedback	Impacted Business Owners and Communities of Interest review the draft policy detail and provide their feedback	Business Owner, Communities of Interest	Draft Data Governance Policy Form	Feedback on draft policy
5.6	Communicate Feedback from Business Owner and Communities of Interest	The Data Custodians communicate the feedback received from the Business Owners and Communities of Interest to the TAC for their consideration in developing the draft policy.	Data Custodians	Feedback on draft policy from Business Owner, Communities of Interest	Feedback on draft policy to TAC
5.7	Develop Revised Draft Policy Based on Feedback from Data Custodians	The TAC receives feedback from Data Custodians and revises the draft data governance policy to incorporate the feedback, as appropriate. The revised draft data governance policy is forwarded to the Board for their review and approval.	TAC	Feedback on draft policy	Revised draft policy
5.8	Review and Approve Final Policy	The Board reviews the draft data governance policy and revises it as needed. Following the adoption of any Board revisions, a final data governance policy is approved by the Board. The TAC is directed to develop the final policy using the Data Governance Policy Form and to communicate the new policy as appropriate.	Board	Revised draft policy	Approved final policy
5.9	Send Final Policy to Data Custodians	The TAC completes the final Data Governance Policy Form and sends it to all impacted Data Custodians to communicate through their business areas.	TAC	Approved final policy	Completed Final Data Governance Policy Form

5.10	Assess Impact and Communicate to Business Owner and Communities of Interest	<p>The Data Custodian receives the final Data Governance Policy Form and assesses the impact of the new policy to their business area.</p> <p>The Data Custodian informs their Business Owner and Community of Interest of the new policy and any impacts.</p>	Data Custodian	Final Data Governance Policy	Communicate Impact
5.11	Assist Business Area with Policy Implementation	The Data Custodians assist their business areas to implement the newly developed policy.	Data Custodian	New policy and procedures	Assist with implementation
5.12	Inform TAC of any Issues Related to Policy	The Data Custodian informs the TAC of any issues that they have confronted during implementation of the new data governance policy.	Data Custodian	Implementation issues in business area	Inform TAC of implementation issues
5.13	Resolve Lower Level issues with Business Owners	The TAC discusses all lower-level implementation issues with Business Owners (or Data Custodians, as appropriate) and resolves those issues that can be agreed upon.	TAC	Implementation issues	Resolution of lower level issues
5.14	Forward High Level Issues to Board	The TAC forwards any higher level implementation issues to the Board for review and resolution.	TAC	Implementation issues	Communicate high level issues to Board
5.15	Review High Level Implementation Issues	The Board reviews all high-level implementation issues that could not be resolved by the Data Custodians or TAC.	Board	High level implementation issues	None
5.16	Decision: Issues resolved with Business Owners?	<p>The Board resolves as many high level issues as possible with the Business Owners; if:</p> <ul style="list-style-type: none"> • <u>Yes</u>: Issues resolved (End of process) • <u>No</u>: Elevate issues (Go to Step 5.17). 	Board	None	Issue resolved or elevated
5.17	Raise Issue to Directorate (End of process)	If the Board cannot resolve an issue with the Business Owners, these issues will be raised to Caltrans' Directorate for resolution.	Board	Unresolved implementation issue	Issues raised to Directorate

Appendix V Data Quality Assessment Process Map & Narrative

Transportation System Data Business Plan - Data Quality Assessment



Description	During the course of a data governance project, the project team may need to perform a data quality assessment that focuses on data quality issues with important impacts on the business needs of users of the data. If a data quality assessment is necessary, the project team will determine which data will be assessed and will then perform the assessment using the “Data Quality Assessment Worksheet” (see Appendix J). The assessment will result in a list of data quality issues that address business needs of importance to the business functions represented on the project. The issues are prioritized and tasks are assigned to resolve the data quality issues. Some issues may be complex or sensitive and require the establishment of a separate project.				
Assumptions	A data governance project has been established and is underway				
Process Trigger	The project team determines the need for a data quality assessment				
Completion Indicator	Data quality issues are resolved or new data governance projects are identified				
#	Activity	Description	Role	Input	Output
1.	Determine Transportation Data for Quality Assessment	The project team determines that transportation system data contained in their project scope requires a data quality assessment to identify data quality issues or problems.	Project Team	Project Scope	None
2.	Perform Data Quality Assessment	The project team performs the data quality assessment to identify any data quality issues or problems. The “Data Quality Assessment Worksheet” is completed to guide the assessment.	Project Team	None	Data Quality Assessment Worksheet
3.	Review Results and Develop List of Data Quality Issues	The project team reviews the results of the data quality assessment and lists all data quality issues that resulted from the assessment.	Project Team	Data quality assessment	List of data quality issues
4.	Prioritize Issues	The project team prioritizes the data quality issues list in order of importance.	Project Team	List of data quality issues	Prioritized list of issues

5.	Decision: Do Issues Create a New Project?	<p>Some issues identified during the data quality assessment may be of such complexity, importance, or sensitivity that they will need to become separate data governance projects outside the scope of the current project.</p> <ul style="list-style-type: none"> • <u>Yes</u>: A new data governance project will be submitted for consideration (go to Step 7). • <u>No</u>: No new data governance projects are identified (go to Step 6). 	Project Team	List of data quality issues	Decision whether any issues may be separate data governance projects
6.	Assign Project Tasks to Address Quality Issues (end of process)	<p>The project team assigns tasks to members of the project team to resolve data quality issues.</p> <p><i>Note: These tasks are resolved as part of the ongoing project and will be reported and addressed with other project tasks.</i></p>	Project Team	Prioritized list of issues	Tasks assigned to project team
7.	Complete Data Issue Form and Forward to TAC	The project team completes the “Data Issue Form” and forwards it to the Technical Advisory Committee (TAC) for review.	Project Team	Data quality issue	Completed Data Issue Form
8.	Review Data Issue Form (Go to Process 2, “Filtering Issues Prior to Board Consideration”)	The TAC receives and reviews the “Data Issue Form” and performs the necessary data governance processes.	TAC	Completed Data Issue Form	None