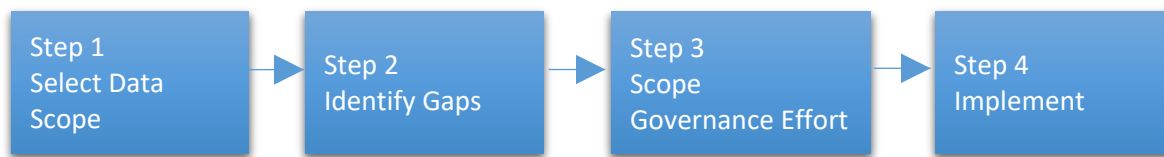


## Data Assessment, Governance Scoping and Initiation Guide

### Overview

Caltrans has defined a standard data governance process that can be pursued by any business unit that has an interest in improving data quality, documentation, interoperability and access. A set of templates and guidance resources are available to support these efforts and promote consistency in how we document and share information about Caltrans data resources. The process is illustrated and described below. It is best carried out by a working group that includes people who are knowledgeable about how the data of concern are stored and managed and any issues related to data quality, consistency and access.



### Step 1. Select Data ScopeX

**Identify what data you would like to include in the effort.** You may want to select an entire database, a set of related databases, or a set of data elements that appear across multiple systems you would like to include in a new data mart or report series. Be careful not to take on too much at one time.

Use the criteria in Table 1 to see if the data you have selected meets criteria for Corporate Data at Caltrans. If it does, meet with the Caltrans Geospatial Data Officer (GDO) to discuss whether to pursue formal designation of the data. Data governance support provided by the Enterprise Data Governance Program (CTDATA) is prioritized for Corporate Data.

If you are considering multiple data governance efforts but don't have the resources to pursue all of them, work through the questions in Table 2 to help you identify which effort should be pursued first. The more "Yes" responses, the higher the priority.

### Step 2. Identify Gaps

**Assess the data to identify gaps.** Conduct an assessment of the data to identify what issues might need to be addressed within the scope of a data governance effort. Use the assessment and scoping tool in Appendix A.

### Step 3. Scope Governance Effort

**Create a workplan for a data governance effort.** Use the assessment and scoping tool in Appendix A to identify activities to close gaps in the data. This workplan should consider actions in the following categories:

- Roles
- Data Documentation
- Data Value

- Data Quality
- Data Standardization, Consistency, and Interoperability
- Data Classification
- Data Protection and Preservation

Walk through your draft workplan with the GDO. This provides an opportunity for you to clarify what is involved for each activity and get connected with tools or other resources that may be helpful. It also enables the GDO to identify any possible synergies between your effort and other data governance efforts. You may also be requested to provide a brief presentation to the Enterprise Data and Geospatial Governance Board.

Line up resources needed to carry out your workplan. Identify internal staff that can spend time on the effort and (if appropriate) line up supplemental resources.

#### Step 4. Implement

If your governance initiative will be making decisions impacting multiple business units across Caltrans, it is a good idea to establish a Technical Advisory Committee (TAC) to ensure coordination and alignment across affected parties.

A charter for this group should clearly identify the scope of the effort, the membership and its roles, and expectations with respect to meetings and other time commitments. A sample TAC charter can be found here [\(future\)](#). You may wish to consult with your Enterprise Data Steward (EDS) on TAC membership, and include them in an advisory role.

Move forward with carrying out your initiative. Please provide quarterly status updates to the GDO/Enterprise Data and Geospatial Governance Board.

As part of the effort, create a Data Documentation Package. The Data Documentation Package (DDP) provides a single “home” to organize and link to all of the data governance deliverables that were created as part of your initiatives.

## Prioritization of Candidate Data Governance Efforts

**Table 1. Corporate Data Criteria**

Criteria	Question to Ask	Response	For Further Information, See:
a. Master or Reference Data	Does it include master or reference data?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<ul style="list-style-type: none"> <li>• <a href="#">Caltrans Corporate Data Identification Document</a></li> <li>• Caltrans Data Documentation Package – Appendix B</li> </ul>
b. Required by Control Agencies	Is it required for reporting to federal or state-level control agencies?	<input type="checkbox"/> Yes <input type="checkbox"/> No	
c. Shared across Caltrans	Is it currently shared across multiple Caltrans business units?	<input type="checkbox"/> Yes <input type="checkbox"/> No	
d. Shared externally	Is it currently shared with external partners or with the public?	<input type="checkbox"/> Yes <input type="checkbox"/> No	
e. Potential for further internal sharing	Is it potentially of value to other Caltrans business units that may not be using it currently?	<input type="checkbox"/> Yes <input type="checkbox"/> No	
f. Potential for future external sharing	Is it potentially of value to agency partners and/or the traveling public?	<input type="checkbox"/> Yes <input type="checkbox"/> No	
g. Purchased/ Licensed Data	Is it produced as the result of state grants, contracts, or cooperative agreements?	<input type="checkbox"/> Yes <input type="checkbox"/> No	

**Table 2. Data Governance Initiative Prioritization**

#	Item	Question	Response
1.	Corporate Data	Does this dataset meet criteria for corporate data?	<input type="checkbox"/> Yes <input type="checkbox"/> No
2.	Scope	Is the dataset statewide in scope?	<input type="checkbox"/> Yes <input type="checkbox"/> No
3.	Emergency Operations	Is the dataset needed for emergency operations?	<input type="checkbox"/> Yes <input type="checkbox"/> No
4.	Frequency of Use	Is the data accessed/used on a daily or weekly basis?	<input type="checkbox"/> Yes <input type="checkbox"/> No
5.	Frequency of Updating	Is the dataset updated on a regular basis to meet a stated ongoing business need?	<input type="checkbox"/> Yes <input type="checkbox"/> No
6.	Suitability for Use	Are there known issues with the dataset that impact its suitability for its intended uses	<input type="checkbox"/> Yes <input type="checkbox"/> No
7.	Manual Processes	Are the data in this dataset maintained primarily through manual entry?	<input type="checkbox"/> Yes <input type="checkbox"/> No
8.	Data Consistency	If this is a time series dataset, have there been changes to data collection methods or data definitions that have impacted consistency of data over time?	<input type="checkbox"/> Yes <input type="checkbox"/> No
9.	Access	Is or should the dataset be broadly shared within Caltrans?	<input type="checkbox"/> Yes <input type="checkbox"/> No
10.	External access	Is or should the dataset be made available to non-Caltrans employees?	<input type="checkbox"/> Yes <input type="checkbox"/> No
11.	Location referencing	If the dataset incorporates location, does it use a location referencing method not compatible with the official LRS maintained by the Division of Research, Innovation and System Information?	<input type="checkbox"/> Yes <input type="checkbox"/> No
12.	Data duplication	Does this dataset fully or partially overlap with the contents of another Caltrans dataset?	<input type="checkbox"/> Yes <input type="checkbox"/> No
Count # of Yes Responses:			

## Appendix A. Data Assessment and Scoping Tool

Once you have identified the dataset(s) to be included in your effort, use the tables below to scope your data governance effort.

The first step to take in a data assessment is to see whether the data you have selected in your scope meets criteria for Corporate Data. A full gap assessment is recommended for Corporate Data.

Each item asks a Yes/No question that provides an “as-is” assessment of your dataset(s). If the answer to each question is “Yes”, there is no need for further action – roles are clearly defined; the data are well documented, and there is no pressing need for improvements to data quality, access or consistency. Otherwise, a data governance effort may be warranted to bring your data to a better “To Be” governed state. Each “No” answer represents a gap that can be closed through a data governance effort. Actions to consider for closing the gaps and available resources to help in carrying out the action are listed.

Not all of the gaps need be addressed through a single effort – actions should be prioritized based on the business needs. However, the items marked with asterisks are *\*strongly recommended\** for all Corporate Datasets.

Note that this tool was developed primarily to support governance of Corporate Data. However, even if your dataset(s) don’t meet criteria for Corporate Data, it is still beneficial to walk through the steps outlined in the table and use the resources provided to improve data stewardship, data quality and documentation.

1. Roles and Responsibilities – Items checked “No” are Gaps			Actions to Consider for Closing Gaps	Resources Available
a. Enterprise Data Steward	Has an Enterprise Data Steward been identified?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> *Work with the Geospatial Data Officer to have stewards designated and management buy-in.	<ul style="list-style-type: none"> <li>• <a href="#">Caltrans Data Governance Roles and Responsibilities</a></li> <li>• Data Steward Listing</li> </ul>
b. Business Data Stewards	Have Business Data Steward(s) been identified?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> *Work with assigned Enterprise Data Steward to identify the steward(s). Provide assignments back to the Geospatial Data Officer.	
c. Custodians	Have Data Custodian(s) been identified?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> *Work with assigned Business Data Steward(s) to identify the custodian(s). Provide assignments back to the Geospatial Data Officer.	
Comments:				

2. Data Documentation – Items checked “No” are Gaps			Actions to Consider for Closing Gaps	Resources Available
a. Data Catalog	If the data meets any of the criteria for Corporate Data, are there one or more entries for this data in the Caltrans Data Catalog?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> *Prepare a data catalog entry and provide to the Geospatial Data Officer	<ul style="list-style-type: none"> <li>• Data Documentation Package – Appendix A, Guidance for Data Catalog</li> <li>• Data Catalog Template</li> </ul>
b. Dataset Metadata	Is there metadata for the dataset(s) available in the Caltrans dataset metadata catalog?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> *Compile any existing dataset-level metadata, and use to populate the Caltrans Dataset Metadata Template. <input type="checkbox"/> *Fill in the remaining items, and provide to the Geospatial Data Officer.	<ul style="list-style-type: none"> <li>• Data Documentation Package – Appendix E, Guidance for Dataset Metadata</li> <li>• Dataset Metadata Template</li> </ul>
c. Data Dictionary	Is there a data dictionary describing each of the data elements in the Caltrans data dictionary catalog?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> *Compile any existing data dictionary information and use to populate the Caltrans Data Dictionary Template. <input type="checkbox"/> *Fill in the remaining items, and provide to the Geospatial Data Officer.	<ul style="list-style-type: none"> <li>• Data Documentation Package – Appendix G, Guidance for Creating a Data Dictionary</li> <li>• Data Dictionary Template</li> <li>• Dataedo Tool (work with Enterprise Data and Geospatial Governance Program staff to utilize this tool)</li> </ul>

2. Data Documentation – Items checked “No” are Gaps			Actions to Consider for Closing Gaps	Resources Available
d. Corporate Data Element Sources	For data that has been designated as Corporate Data, have the Corporate Data Element Sources been identified and categorized?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> *Complete the Corporate Data Element Sources Template for each corporate data element	<ul style="list-style-type: none"> <li>Data Documentation Package – Appendix B, Guidance for Documenting Corporate Data Element Sources</li> <li>Caltrans Corporate Data Element Sources Template</li> </ul>
e. Entity-Relationship Diagram	Is there an Entity-Relationship (E/R) diagram showing how the different physical database tables are related to one another?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Work with the data custodian to create an E/R diagram	<ul style="list-style-type: none"> <li>Dataedo Tool (work with Enterprise Data and Geospatial Governance Program staff to utilize this tool)</li> <li><a href="#">Sample E/R diagram 1</a></li> <li><a href="#">Sample E/R diagram 2</a></li> </ul>
f. Context Diagram	Is there a high level context diagram available that illustrates the systems that store the data, how data flows into and out of that system, and the sources and recipients of the data?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> *Create a high level context diagram showing data flows.	<ul style="list-style-type: none"> <li>Data Documentation Package – Appendix D, Data Diagrams</li> <li><a href="#">Sample context diagram</a></li> </ul>
g. Lineage Diagrams	Are there more detailed lineage diagrams showing how given data elements flow across different systems?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Consider creating detailed data lineage diagrams.	<ul style="list-style-type: none"> <li>Data Documentation Package – Appendix D, Data Diagrams</li> <li><a href="#">Sample lineage diagram 1</a></li> <li><a href="#">Sample lineage diagram 2</a></li> </ul>



2. Data Documentation – Items checked “No” are Gaps			Actions to Consider for Closing Gaps	Resources Available
h. Business Glossary	Have terms and definitions for key entities and attributes for this dataset been identified for inclusion in the Caltrans business glossary?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Scope a task to identify potential glossary terms and gain agreement on their definitions.	<ul style="list-style-type: none"> <li>• Data Documentation Package – Appendix F, Guidance for Defining Business Glossary Terms</li> <li>• Caltrans Business Glossary Template</li> </ul>
Comments:				

3. Understanding of Data Value – Items checked “No” are Gaps			Actions to Consider for Closing Gaps	Resources Available
a. Users and Uses	Is there any documentation of how this data is being used (e.g. list of primary users, diagram of how data flows to different reports, data value maps, etc.)	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Consider creating a data value map that links data sources, products and users/uses or conducting a data value assessment	<ul style="list-style-type: none"> <li>• Data Documentation Package – Appendix D, Data Diagrams</li> <li>• <a href="#">NCHRP Data Value Assessment Tool</a></li> <li>• Sample data value diagram (future)</li> </ul>
b. Business Function Mapping	Have the Caltrans business functions been identified that this data supports?	<input type="checkbox"/> Yes <input type="checkbox"/> No		
c. Data Processing Schedule	If there are reporting deadlines – internally or externally imposed that drive the data collection and updating schedule, have these been documented?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Consider mapping the data collection and processing activities and responsibilities (if not already completed)	<ul style="list-style-type: none"> <li>• <a href="#">Sample data processing diagram</a></li> </ul>
Comments:				

4. Data Quality – Items checked “No” are Gaps			Actions to Consider for Closing Gaps	Resources Available
a. Business Rules	Have business rules for data validation been documented for each of the data elements?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> *Create business rules using the Caltrans business rules template	<ul style="list-style-type: none"> <li>Data Documentation Package – Appendix H, Guidance for Specifying Business Rules</li> <li>Caltrans Business Rules Catalog Template</li> </ul>
b. Data Quality Management Plan	Is data quality being actively managed for this dataset – i.e. are there well-defined of quality control and assurance steps applied to the data and are these documented in a data quality management plan (or equivalent)?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> *Create and implement a Data Quality Management Plan	<ul style="list-style-type: none"> <li>Caltrans Data Quality Management Plan Template</li> </ul>
c. Data Profiling - Completeness	Has data completeness been assessed: e.g. percentage of records with missing values, percentage of target scope covered?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Scope a data profiling task to obtain quantitative measures of quality and identify issues such as duplicate records, missing records, or invalid records.	<ul style="list-style-type: none"> <li><a href="#">Data Profiling Tutorial</a></li> </ul>
d. Data Profiling - Validity	Has data validity been assessed: e.g. number or percent of records with invalid values?	<input type="checkbox"/> Yes <input type="checkbox"/> No		
e. Data Profiling - Currency	Has data currency been assessed: e.g. date of last update for data records?	<input type="checkbox"/> Yes <input type="checkbox"/> No		

f. Data Usability	Has data usability been assessed: e.g. reported user ratings of how easy it is to obtain and transform the data for use?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Scope a task to convene a user focus group or conduct a user survey the better understand user perspectives.	<ul style="list-style-type: none"> <li>• <a href="#">NCHRP Data Value Assessment Tool</a></li> </ul>
<p>Comments:</p>				

5. Data Standardization, Consistency, and Interoperability – Items checked “No” are Gaps			Actions to Consider for Closing Gaps	Resources Available
a. Location Referencing	If this dataset includes location references, are these consistent with the Caltrans Linear Referencing System (LRS), and are processes in place to enable historical data to be located on the current LRS?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> NA	<input type="checkbox"/> *Scope a task to explore options for updating location references in the dataset.	<ul style="list-style-type: none"> <li>• <a href="#">Caltrans Linear Referencing System Branch</a></li> </ul>
b. Authoritative Sources	Is there a shared understanding or agreement about what the “official” authoritative source system is for the data elements in the dataset?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> *Scope a task to gain agreement on and document authoritative sources.	
c. Duplicative Data	Is there reasonable confidence that the data in this dataset are not duplicated across various Caltrans systems – other than to provide an authorized reporting source?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> *Scope a task to identify and suggest strategies to reduce data duplication.	
d. Interoperability	If elements from this dataset are found in multiple systems, has there been a review of opportunities to improve consistency or streamline how these elements are shared across systems?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> NA	<input type="checkbox"/> *Scope a task to explore improvements to interoperability – e.g. using data integration tools, data warehouse solutions, or data services.	

5. Data Standardization, Consistency, and Interoperability – Items checked “No” are Gaps			Actions to Consider for Closing Gaps	Resources Available
e. Applicable Standards	Has this dataset been assessed for conformance to existing Caltrans or national data standards?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> *Scope a task to identify and assess conformance to data standards.	
f. Future Standardization	Is there reasonable confidence that there would be no benefit to further standardization of definitions, formats, or lists of values of the data elements in this dataset?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> *Scope a task to develop Caltrans data standards for one or more elements.	
Comments: <div></div>				

6. Data Classification – Items checked “No” are Gaps			Actions to Consider for Closing Gaps	Resources Available
a. Identification of Sensitive and Confidential Data	Have the data been reviewed to identify any sensitive or confidential data elements (e.g. personally identifiable information or PII)?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> *Review and classify data elements based on established California Department of Technology Information Security Program Management Standard <input type="checkbox"/> Scope a task for data de-identification	<ul style="list-style-type: none"> <li>• <a href="#">California Department of Technology Information Security Program Management Standard SIMM 5305-A</a></li> <li>• Caltrans information security guidance (future)</li> <li>• <a href="#">CHHS Deidentification Guidelines</a></li> </ul>
b. Documentation of Sensitive and Confidential Data	Have data element classifications been documented in a data dictionary?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> *Record data element classifications in the Caltrans Data Dictionary Template	<ul style="list-style-type: none"> <li>• Data Documentation Package – Appendix G, Guidance for Creating a Data Dictionary</li> <li>• Data Dictionary Template</li> </ul>
c. Identification of Public, Restricted and Non Public Data	Have the data been assessed to determine whether they are public, restricted or non-public?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> *Review and classify datasets based on whether they are public, restricted or non-public	<ul style="list-style-type: none"> <li>• Data Documentation Package – Appendix A, Guidance for Data Catalog and Appendix E, Guidance for Dataset Metadata</li> <li>• Caltrans Data Catalog Template</li> <li>• Caltrans Dataset Metadata Template</li> </ul>

6. Data Classification – Items checked “No” are Gaps			Actions to Consider for Closing Gaps	Resources Available
d. Documentation of Public, Restricted and Non Public Data	Have dataset classifications of public, restricted, and non-public been documented in dataset metadata?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> *Record dataset classifications in the Caltrans Data Catalog and Dataset Metadata Templates	<ul style="list-style-type: none"> <li>• Data Documentation Package – Appendix A, Guidance for Data Catalog and Appendix E, Guidance for Dataset Metadata</li> <li>• Caltrans Data Catalog Template</li> <li>• Caltrans Dataset Metadata Template</li> </ul>
<p>Comments:</p>				



7. Data Access, Protection and Preservation – Items checked “No” are Gaps			Actions to Consider for Closing Gaps	Resources Available
a. Data Access	Is this dataset generally available and accessible to people who need it or may find it valuable – both internally and externally?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Scope a task to improve internal and/or data accessibility	<ul style="list-style-type: none"> <li>• <a href="#">California Open Data Handbook</a></li> </ul>
b. Sensitive Data Protection	If the data are sensitive, are sufficient protections in place to control access to the data?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> *Scope a task to review and revise how data access permissions are granted and revoked.	
c. Backups	Are the data currently being backed up?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> *Work with Caltrans Information Technology to implement backup procedures.	
d. Verification of Backups	Have backups been tested and verified by a data steward (or other business user representative) over the past 12 months to ensure that data can be restored properly?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Work with Caltrans Information Technology to schedule a data restore test including business representative involvement.	

7. Data Access, Protection and Preservation – Items checked “No” are Gaps			Actions to Consider for Closing Gaps	Resources Available
e. Data Storage Requirements	Is there an understanding of data storage requirements over the next 5 years and a plan in place to meet these requirements?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Scope a task to estimate future storage requirements <input type="checkbox"/> Scope a task to develop a data storage plan identifying where the data are to be stored, and how storage is to be resourced.	
f. Data Update Plan	Is there an established and resourced data update cycle or approach to meet anticipated business needs?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Scope a task to develop a data updating strategy that addresses business needs.	
g. Data Retention	Has a data retention schedule been established and validated by business users?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> *Scope a task to identify and revise (as needed) the applicable data retention schedule	
h. Data Archiving Plan	Is there a resourced plan for data archiving?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Scope a task to create a data archiving plan to ensure sufficient access to the data to meet business needs through the end of the retention period.	

7. Data Access, Protection and Preservation – Items checked “No” are Gaps	Actions to Consider for Closing Gaps	Resources Available
<p>Comments:</p>		