

California Department of Transportation ENTERPRISE DATA AND GEOSPATIAL GOVERNANCE PRACTICE	NUMBER: DGP-05	REVISION: 1	DATE ISSUED: February 28, 2025
SUBJECT: Open Data	REFERENCES: The CA Open Data Handbook Technology Letter (TL 19-01) Creative Commons 4.0 Attribution		
BACKGROUND AND PURPOSE: This practice document defines and standardizes the process of identifying, vetting, and publishing Caltrans corporate data as Open Data in compliance with the California Department of Technology (CDT) Technology Letter 19-01 (TL 19-01). CDT, working in cooperation with the Government Operations Agency (GovOps), released TL 19-01 in March of 2019 to promote more accessible, discoverable, and usable data that impacts economic development and improves government services. Open data encourages informed policy decisions, performance planning, research, and scientific discoveries, and increased public participation in democratic dialogue. TL 19-01 established the CA Open Data Portal to improve collaboration, expand transparency, foster innovation, and increase effectiveness by unlocking the value of government data. While several Agencies/State entities host their own open data portals, CA Open Data Portal was designed specifically to host open data from more than one agency. The CA Open Data Portal will link to existing State portals so that all the State's open datasets can be searched from a central location. In addition, TL 19-01 updated definitions for "Metadata" and "Open Data" in Section 4819.2 of the State Administrative Manual (SAM) and added Sections 5160, 5160.1 and 5160.2 on the Open Data Policy. This summarized a set of changes to the SAM specific to implementation of practices to more broadly share California government datasets that went into effect July of 2020.			
DEFINITIONS: Business Data Steward – An individual with accountability for data within a defined business area/scope. Corporate Data – Data assets that have been prioritized for data governance based on their value to Caltrans and its customers and stakeholders, and where lack of attention to data standardization, documentation, quality management and coordination introduces inefficiencies and risks for the agency. Corporate Data falls into one or more of the following categories: 1. Master Data identifies and describes the people, places and things that are fundamental to agency activities and appear across multiple agency information systems, like Project Short Names. 2. Reference Data is permissible values for other data fields, like County codes and names.			

3. **Control Agency Required Data** is data that is required by law or directly impacts agency funding or resource allocation decisions like SB1 reports or Federal Highway Performance Monitoring System data.
4. **Shared Data** is used across multiple internal business units or is valuable to either Caltrans' partners or the public. An example of shared data is the Caltrans Fact Book.
5. **Purchased/Licensed Data** are produced as a result of state/federal grants, contracts, or cooperative agreements.

Data Custodian – An individual with physical custody of the data. Responsible for the technical infrastructure (hardware, software, networking), database administration and backup of a particular set of data.

Enterprise Data Steward – An individual with accountability for corporate and shared data within an established domain or program area (e.g. capital projects, roadway, traffic operations, design).

Open Data – data that can be freely accessed, used, modified, and shared by anyone for any purpose (<http://opendefinition.org/>). At Caltrans, Open Data is corporate data that:

1. has an identified source system of record (SSOR) or authoritative reporting source (ARS) in accordance with DGP-06 "Designating Source Systems of Record and Authoritative Reporting Sources of Corporate Data,"
2. has a designated Enterprise Data Steward,
3. does not contain sensitive, confidential, and/or personally identifiable information,
4. does not pose a risk to Caltrans or its employees if shared,
5. has been documented using required CTDATA standard data documentation templates,
6. and made available to the public.

Open data is published using a license, such as Creative Commons 4.0 attribution.

Data Coordinator – single point of contact for a government agency/state entity on the open data portal.

Metadata - Structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource.

Creative Commons License – a simple way for creators to share their work with others while keeping some control over how it's used. Instead of making everything completely free or fully restricted, a creative commons license lets creators decide if others can copy, share, or modify their work, often with certain conditions like giving credit or not using it for commercial purposes, helping people collaborate and build on each other's ideas more easily.

RESOURCES:

[Enterprise Data and Geospatial Governance \(EDGG\) Structure and Roles](#) – This internal webpage defines the EDGG structure and roles at Caltrans, including resources for the definitions of Business Data Steward, Data Custodian, and Enterprise Data Steward.

[CA Open Data Handbook](#) – This online document explains the requirements around publishing open data and includes documentation on the state open data metadata standard.

[Technology Letter \(TL 19-01\)](#) – This document explains open data which promotes more accessible, discoverable, and usable data that impacts economic development and improves government services.

[CA Open Data Portal](#) – site that has datasets, geospatial or not, already published as open data.

[Creative Commons License](#) – public copyright license that gives everyone from individual creators to large institutions a standardized way to grant the public permission to use their creative work under copyright law.

[CA Health and Human Services \(CHHS\) Data Playbook](#) – a toolkit for using data in a department.

[State Administrative Manual \(SAM\) – Open Data Policy 5160](#) – acts as the introduction to the Open Data Policy and provides the core mission, vision, and goals for open data in state government.

[SAM – Open Data Policy Requirements 5160.1](#) – provides more guidance on adhering to the Open Data Policy, charging each agency/state entity with specific requirements.

[SAM – Open Data Exceptions 5160.2](#) – outlines applicable exceptions to the Open Data Policy.

CA OPEN HANDBOOK REQUIREMENTS FOR OPEN DATA:

Each organization will designate a Data Coordinator. The Data Coordinator should be an individual who:

- Has authority equivalent to that of a Program Manager or higher.
- Can identify appropriate persons with comprehensive knowledge of data and resources in use by their organization.
- Assumes responsibility for their organization's compliance with the Open Data Handbook, as well as data de-identification guidelines, such as the CA Health and Human Services (CHHS) Data De-Identification Guidelines, and future directives which may be needed to support the open data portal.
- Ensures that guidelines and tools, established by governance in the Open Data Handbook, are applied to aggregate data for public reporting that:

- (1) adequately minimizes risk of re-identification, maximizing security and confidentiality and
- (2) produces meaningful information.

State agencies should create and maintain an enterprise data inventory. This inventory should indicate whether datasets listed may be made publicly available and whether they are currently available to the public. The inventory shall also list any datasets or harvestable files that can be made publicly available at the Agency/State entity's open data site or portal. This should be in a format that enables automatic aggregation by the CA Open Data Portal and other services, to the extent practicable.

The inventory list should include, to the extent permitted by law and existing terms and conditions, datasets that were produced because of legislative mandates, state grants, contracts, and cooperative agreements, excluding any data submitted primarily for the purpose of contract monitoring and administration. Where feasible, the list should include standard citation information, including the Data Coordinator's contact information and preferably using a persistent unique identifier.

Organizations should make an assessment to determine the priority for each dataset. Prioritizing initial and ongoing publication will entail balancing high value data with the data's level of readiness for publication. Each organization should create and provide schedules prioritizing data publication in accordance with the guidelines set forth herein.

Prioritization done in a timely manner is important, recognizing that it may take time to prepare high quality data (noting that data tables vary in complexity and, as such, can significantly vary in preparation time). Approvals for the prioritization plan and scheduling will come from the organization's executive leadership team.

PRACTICE:

Data Coordinator

The Data Coordinator for Caltrans is the Enterprise Data and Geospatial Governance Program (CTDATA) Manager. The Caltrans Geospatial Data Officer supports data coordination efforts as well.

Open Data Compliance

CTDATA is responsible for ensuring that Caltrans adheres to the Open Data Handbook. Enterprise Data Stewards, Business Data Stewards, and Data Custodians for Caltrans databases, systems, and datasets are responsible for adhering to this practice document and attached workflows in support of the Open Data Policy.

Classifying Open Data

Enterprise Data Stewards, Business Data Stewards, and/or Data Custodians are responsible for determining if data in a system or dataset is corporate data by completing the Data Classification Checklist, see Appendix A. If the checklist indicates that the data contains sensitive, confidential, and/or personally identifiable information, and/or sharing the data poses a risk to Caltrans or its employees, then a review by the Data Classification and Security Committee (DCSC) is necessary to classify the data. Records of the classification of the data will be kept in the Corporate Data Catalog.

The Data Classification and Security Committee (DCSC)

The DCSC is the governing body responsible for evaluating data or systems that are identified by enterprise or business data stewards as potentially restricted, non-public, or has use constraints that would make the data unable to be shared with the public in its current form. The DCSC reports to the Enterprise Data and Geospatial Governance Board and acts as the authority for deciding if data can be classified as public (open data), restricted, or non-public. The Enterprise and Business Data Stewards that are responsible for the dataset will meet with the DCSC to answer any questions and clarify their justifications for the answers on the Data Classification Checklist. See the "Supporting Documents" section of this document for a link to the DCSC charter.

Data De-Identification

If the DCSC rules that the data can be classified as "public (open data)" and if some data or information about the data is redacted, then the Enterprise Data Steward, Business Data Steward, and/or Data Custodian must follow data de-identification guidelines. Data de-identification guidelines at Caltrans follow the [CHHS De-identification Guidelines](#).

Open Data Inventory

Once the dataset has been classified, the enterprise data steward and/or business data steward shall create the required data documentation for datasets, including an entry for the dataset on the [Corporate Data Catalog](#). The Corporate Data Catalog is the application responsible for tracking the enterprise data inventory for Caltrans. Datasets listed within the Corporate Data Catalog shall include the following information:

1. The dataset's name.
2. If the dataset is produced as a result of legislative mandate, state grants, contracts or cooperative agreements.
3. Whether the dataset's public access level is classified as "public (open data)", "restricted", or "non-public". If the dataset is classified as restricted or non-public, the reason for said classification must be included in a separate field.
4. If the data has use constraints, and the details of said use constraints, if any.
5. The business area(s) that the data represents.
6. The type(s) of corporate data categories that the data falls into.
7. The responsible Enterprise Data Steward's contact information.
8. The responsible Business Data Steward's contact information.
9. A hyperlink to the required data documentation for the dataset on the Caltrans intranet.

If the dataset is classified as "public (open data)" then these additional fields must be completed:

10. What license has been applied to the dataset.
11. A hyperlink to the intranet website the dataset is being shared on, if one exists.
12. A hyperlink to the specific public website the data is being shared on, if one exists.

See Appendix C for a list of all fields collected and tracked by the Corporate Data Catalog.

Choosing a License for Open Data

All data that is classified as "public (open data)" shall have a license applied to it. The default license for Caltrans Open Data shall be [Creative Commons Zero \(CC0\) 1.0 Universal](https://creativecommons.org/publicdomain/zero/1.0/?ref=chooser-v1).¹ Alternatively, [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/?ref=chooser-v1)² may also be chosen. Other licenses may be used if approved by Caltrans Legal and the Data Classification and Security Committee.

Open Data Publishing Platforms

GIS datasets classified as open data shall be publicly posted to the CA State Geoportal. Both GIS and non-geospatial data (e.g., spreadsheets and other tabular data) classified as open data shall be publicly posted on the CA Open Data Portal.

Publishing Priority

CTDATA shall maintain a list of data available on the CA Open Data Portal using the Corporate Data Catalog, including its publishing priority and schedule. This list will be published annually on the Caltrans external website.

¹ <https://creativecommons.org/publicdomain/zero/1.0/?ref=chooser-v1>

² <https://creativecommons.org/licenses/by/4.0/?ref=chooser-v1>

SUPPORTING DOCUMENTS:

Appendix A – Open Data Publishing Process

Appendix B – Data Classification Checklist (This checklist is also available online³)

Appendix C – Data Dictionary: Corporate Data Catalog

Appendix D – Data Classification and Security Committee Charter

APPLICABILITY:

This practice document shall apply to all Caltrans corporate data identified as, or required to be, public (open data).

DATA LICENSE:

The default license for Caltrans Open Data shall be Creative Commons Attribution 4.0 International (CC-BY 4.0). More restrictive licenses may be used as needed, however, the most permissible license possible shall be assigned.

TIME FRAME:

This practice document shall take effect immediately.

The publication process, outlined in Appendix A, shall take effect upon signature and approval of this practice document.

The list of data available on the CA Open Data Portal shall be published annually on the Caltrans external website starting the first business day of October 2025.

IMPLEMENTATION:

CTDATA staff, in collaboration with the Enterprise Data Stewards Committee, shall lead efforts to create and maintain the inventory for all existing corporate datasets and implement the best open data practices for all datasets in progress or in planning. Priority will be given to data and datasets that:

1. Are publicly available on Caltrans external websites, including datasets that Caltrans is required by existing law to publish.

³ <https://apps.gov.powerapps.us/play/e/82153e68-a579-e26e-8d68-87f1ea3bf91b/a/dc017d2b-c507-4a16-ad02-cc517c3dadd7?tenantId=621b0a64-1740-43cc-8d88-4540d3487556&hint=aea5c84a-7cde-4a49-ac60-53ffaeec35c&sourcetime=1764110573312>

2. Have been identified through repeated California Public Records Act and/or Freedom of Information Act requests.
3. Have been identified through engagement with Caltrans' customers, partners, and the public.

CTDATA staff will work with the Public Information Office to create an annual open data survey to engage with our customers, partners, and the public with the goal of identifying data of value. The results of this survey will be presented to the Enterprise Data and Geospatial Governance Committee, Board, and Enterprise Data Stewards Committee. EDSC members will work with their peers and management to prioritize the publishing of data identified through public engagement surveys.

Enterprise and Business Data Stewards are responsible for implementing this practice document on the data, datasets, and systems they are responsible for.

ROLES AND RESPONSIBILITIES:

Enterprise Data Stewards and District Enterprise Data Governance Liaisons shall work to ensure competing datasets are not created by bringing potential items to the Enterprise Data Steward Committee for discussion and action.

Enterprise Data Stewards, District Enterprise Data Governance Liaisons, Business Data Stewards, Data Custodians, and other identified staff shall assist with efforts to modify their practices and update their systems to utilize the authoritative datasets in their data management and quality practices, where applicable and as described under *Implementation* of this document.

Division Chiefs and Deputy District Directors shall, where applicable, direct and support staff to develop and implement workplans for migrating from competing datasets and instead utilize the identified SSOR and ARS.

APPROVAL:

Chad Baker

Digitally signed by Chad Baker

Date: 2025.12.09 11:48:56

-08'00'

Date: 12/09/2025

Chair, Enterprise Data and Geospatial Governance Board or Designee
California Department of Transportation

California Department of Transportation
**ENTERPRISE DATA AND GEOSPATIAL
GOVERNANCE PRACTICE**

NUMBER:
DGP-05

DATE ISSUED:
February 28, 2025