

California Department of Transportation Enterprise Data and Geospatial Governance Practice	NUMBER: DGP-03	DATE ISSUED: AUGUST 11, 2021
SUBJECT: Data Dictionary Standard	REFERENCES: Caltrans Data Documentation Package DDP-8 Caltrans Data Dictionary Template CA Open Data Handbook	
BACKGROUND AND PURPOSE: A data dictionary provides information about the different data elements or fields in a dataset or database table. It serves several different purposes: <ul style="list-style-type: none"> • Capture and preserve information about data elements to maintain and sustain a common understanding of the data • Help future data users to understand the meaning of each data element • Articulate business requirements to be used by database developers to design new databases or modify existing databases • Communicate current database design characteristics to application and report developers <p>The purpose of this practice document is to identify the data dictionary standard for Caltrans corporate datasets and database tables that must be maintained and included when a dataset is shared with data consumers or database structure is shared with authorized recipients.</p>		
DEFINITIONS: Data consumer – Any recipient of a dataset. 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. Please visit the following webpage for full details: https://datagovernance.onramp.dot.ca.gov/corporate-data		
RESOURCES: Caltrans Data Documentation webpage. This webpage includes the Data Documentation Package and DDP-8 Caltrans Data Dictionary Template. https://datagovernance.onramp.dot.ca.gov/data-documentation CA Open Data Handbook. The online document explains the requirements around publishing open data and includes documentation on the state open data data dictionary standard. https://handbook.data.ca.gov/		

PRACTICE:

A data dictionary shall be completed and maintained for all Caltrans corporate datasets and databases. The data dictionary elements that must be completed are described in the Data Documentation Package Appendix G and DDP-8 Caltrans Data Dictionary Template. Individual data dictionary elements are identified as required, if applicable or optional.

Geographic Information System (GIS) format datasets (shapefiles, file geodatabase feature classes, web feature services, etc.)

GIS format datasets shall use the ISO 19139 GML 3.2 metadata specification for storing data dictionary details. The data dictionary-specific metadata elements are specified in Data Documentation Package Appendix G and shall be directly managed as part of the dataset itself using a metadata editor such as Esri's ArcCatalog. Where it is impossible to manage the data dictionary-specific metadata in this fashion, the data dictionary must be provided as an ISO 19139 GML 3.2 compliant XML file or as a CSV export from the DDP-8 template.

Non-GIS format datasets (CSV, XLS, XML, JSON, etc.)

Non-GIS format datasets shall use DDP-8 for recording the dataset data dictionary and exported as a CSV file.

Databases (Oracle, MS SQL Server, Enterprise Geodatabase)

Database tables shall be documented using DDP-8 for recording the data dictionary and exported as a CSV file.

All data dictionaries shall be maintained over the entirety of the dataset or database lifecycle and shall be delivered with all corporate datasets to data consumers or, as needed, to staff or consultants perform system interoperability efforts.

THIRD-PARTY SOURCED DATA:

For externally sourced data where the responsible entity is unable or unwilling to create a compliant data dictionary but for which Caltrans reuses, reshapes, and/or enhances the dataset, the responsible Business Data Steward must ensure that a compliant data dictionary is created.

TIME FRAME:

This practice document shall take effect immediately and existing published corporate datasets shall have compliant data dictionaries created within 30 days and all corporate databases shall have compliant data dictionaries created within 90 days. If additional time is required for compliance of existing published corporate datasets, business areas may request a time extension via email to the Geospatial Data Officer.

Newly completed corporate datasets will have a data dictionary created as part of the overall dataset creation process.

Newly created or expanded databases will have a data dictionary created as part of the development process of creating or expanding the database.

IMPLEMENTATION:

Caltrans Business Data Stewards and Data Custodians should lead efforts to create data dictionaries for all existing corporate datasets and database tables. They will ensure all new datasets and database tables created have data dictionaries created concurrently. When datasets are shared to the open data portal, emailed, or distributed through webpages or any other means, the associated data dictionary shall be provided.

ROLES AND RESPONSIBILITIES:

Enterprise Data Stewards, District Enterprise Data Governance Liaisons, Business Data Stewards, Data Custodians and other identified staff shall lead efforts to modify or augment existing practices as necessary to implement the requirements of this Practice document.

Division Chiefs and Deputy District Directors shall, where applicable, direct and support staff to create, maintain and distribute data dictionaries for their corporate datasets and databases.

Consultants working on projects and/or are under contract with Caltrans shall follow the requirements of this Practice document.

When entering into agreements or contracts with an external entity, the requirements of this Practice document should be addressed as part of the negotiation process.

APPROVAL:

Chad Baker

Date: August 11, 2021

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