Data Catalog Guidelines

Document No: CDO-003
Updated: December 8, 2024
Issued by: Chief Data Officer

1.0 Purpose

The purpose of the Data Catalog Guidelines is to establish common guidelines for data catalog management across State of Hawaii agencies. Through effective data catalog management, state agencies can promote transparency and data sharing, improve data governance and compliance, and enable evidence-based decision-making across the State of Hawaii agencies.

2.0 Authority

Hawaii Revised Statutes (HRS)[1] Section §27-44, established within the Office of Enterprise Technology Services, in the Department of Accounting and General Services, and under the authority of the Chief Information Officer, the Chief Data Officer to develop, implement, and manage statewide data policies, procedures, standards, and guidelines. HRS §27-44 also established a Data Task Force to assist the Chief Data Officer in developing the State's data policies, procedures, and standards.

3.0 Scope

3.1 State Agencies

The Data Catalog Guidelines apply to all state agencies.

The Data Catalog Guidelines provide high level guidelines. Each agency shall develop additional policies and guidelines as necessary according to relevant federal and state laws and regulations, both at the data asset level and at the individual field level, to ensure compliance in its operations. When a conflict exists between the Data Catalog Guidelines and an agency's policy, the more restrictive policy will take precedence.

3.2 Definitions

[1] HRS §27-44. https://www.capitol.hawaii.gov/hrscurrent/Vol01_Ch0001-0042F/HRS0027/HRS_0027-0044.htm

A data catalog is a detailed inventory of all data assets (and information about those data assets) in an organization, designed to help data professionals quickly find and understand data efficiently for any business purposes.

3.3 Covered Use

The Data Catalog Guidelines apply to all data assets handled by state agencies. This includes, but is not limited to systems in the cloud, on premises, and/or on local drives.

4.0 Information Statement

Each agency is responsible for creating and maintaining its data catalog and metadata. The following are the recommended minimum requirements:

- **Ownership:**

Each agency shall identify a data steward for each data assets. A data steward is an agency official with statutory or operational authority for specified information and responsibility for establishing the controls for its generation, collection, processing, dissemination, and disposal.[2]

Responsibilities of data steward for each data asset shall include:

1. Define access control to ensure compliance.
2. Maintain quality of the data.
3. Ensure data integrity and security.

- **Completeness:**

The data catalog shall include all relevant data assets for each specific business use case including spatial data, regardless of format, location, or storage method (databases, spreadsheets, files, cloud storage, etc.). This includes databases, spreadsheets, files, cloud storage, and any other potential data sources and storage locations.

- **Metadata:**

The catalog shall include up-to-date metadata information about each data asset. This metadata serves as essential information about the data, fostering interoperability and data sharing.

- **Automation:**

Automated tools and workflows are strongly recommended whenever possible for data catalog creation and updates to improve efficiency and accuracy.

[2] National Institute of Standards and Technology (NIST) Glossary. https://csrc.nist.gov/glossary/term/information_steward

5.0 Compliance

The Data Catalog Guidelines shall take effect upon publication. The Chief Data Office may amend at any time; compliance with guidelines is strongly recommended.

6.0 Contact Information

Submit all inquiries and requests for future enhancements to the Chief Data Officer in the Office of Enterprise Technology Services, Department of Accounting and General Services, at data@hawaii.gov.

Additional data related policies and guidelines documents can be found at data.hawaii.gov.

7.0 Definitions of Key Terms

All terms shall have the meanings found in the Data & AI Glossary (under Glossaries on https://data.hawaii.gov/).

- **Data Catalog:** Data Catalog refers to an organized inventory of data assets in the organization. It uses metadata to help organizations manage their data. It also helps data professionals collect, organize, access, and enrich metadata to support data discovery and governance.[3]
- **Metadata:** Information describing the characteristics of data including, for example, structural metadata describing data structures (e.g., data format, syntax, and semantics) and descriptive metadata describing data contents (e.g., information security labels).[4]
- **Data Asset:** Data Asset refers to any entity that is comprised of data. For example, a database is a data asset that is comprised of data records. A data asset may be a system or application output file, database, document, or web page.[5]
- **Data Steward:** A data steward is an agency official with statutory or operational authority for specified information and responsibility for establishing the controls for its generation, collection, processing, dissemination, and disposal.[6]
- **Data Governance:** Data Governance refers to setting direction on data use through prioritization and decision making, and ensuring alignment with agreed-on direction and objectives.[7]
- **Spatial data:** Spatial data, also known as geospatial data, refers to information that

[3] Oracle. https://www.oracle.com/big-data/data-catalog/what-is-a-data-catalog/
[4] National Institute of Standards and Technology (NIST) Glossary. https://csrc.nist.gov/glossary/term/metadata
[5] National Institute of Standards and Technology (NIST) Glossary. https://csrc.nist.gov/glossary/term/data_asset
[6] National Institute of Standards and Technology (NIST) Glossary. https://csrc.nist.gov/glossary/term/information_steward
[7] Information Systems Audit and Control Association (ISACA). https://www.isaca.org/resources/glossary

explicitly describes the location, shape, and relationships of geographic features and phenomena. It can exist in various formats and dimensions, including:[8]

- Points: Representing discrete locations (e.g., addresses, landmarks).
- Lines: Representing linear features (e.g., roads, rivers, boundaries).
- Polygons: Representing areas with defined extents (e.g., buildings, countries).
- Rasters: Representing continuous data on the earth's surface using grid cells (e.g., elevation, temperature, windspeed, precipitation). Higher-resolution data has smaller grid cells and represents data with greater precision.
- Images: Capturing spatial information through pixels (e.g., satellite imagery, aerial photographs).
- Lidar: Refers to a type of remote sensing data that uses lasers, commonly used to create high-resolution elevation models.
- 3D models: Representing objects and features in three dimensions.

8.0 Revision History

Date	Description of Change
December 16,2024	Approved by the State Data Task Force

[8] Types of Spatial Data. https://geographicbook.com/types-of-spatial-data/