Data Management Policy

Committee on Information Technology (COIT)

PURPOSE AND SCOPE

Managing the systems that collect, store, process, and share data is key for government to work efficiently and deliver services. This matters even more when operations span multiple departments. Good data management brings several benefits:

- Ensuring trustworthy and secure data
- Enabling seamless integration between departments
- Unlocking the ability to deploy services quickly
- Driving staff efficiency through modern tools
- Enhancing the City's ability to mitigate risk and ensure compliance monitoring
- Greatly reduce costs across data storage, processing, and beyond

This policy applies to all information resources operated by the City and County of San Francisco as well as its departments and commissions, elected officials, employees, consultants, city-funded nonprofits, and vendors working on behalf of the City and County of San Francisco. The above listed entities are required to comply with this policy.

POLICY STATEMENT

This policy establishes a framework for the proper management of data across the City. Departments must adopt this framework, which includes:

- Ongoing modernization of data systems
- Inventory and classification of data systems, and the data within them
- Assessment of interdepartmental data uses
- Considerations for non-public data access
- Making data accessible for public use

POLICY REQUIREMENTS

MODERN DATA SYSTEMS

A data system as defined in California Government Code §7922.630, is the foundation of good data management. The City should be intentional in modernizing its data systems, especially when procuring new systems or performing upgrades to existing systems.

If departments do not have their own standards, departments can utilize the data systems guide when evaluating or updating data systems.

The City's standard contracts, such as the P-64X series, should be updated to more closely meet the guide referenced above.

DATA SYSTEM AND DATA INVENTORIES

Managing data starts with knowing the type of data under the control of the City. In order to do so, departments must:

- **Complete the Data Systems and Data Inventory** per procedures and templates set by the Chief Data Officer (CDO). This process should happen no less than once a year, the results of which will be public (See Dataset and System Inventory).
- **Classify All Data Systems and Data** based on the Data Classification Standard.

INTERDEPARTMENTAL DATA

Data often has a network effect – the more who use it, the more valuable it becomes. To facilitate easier access to interdepartmental data, a three-part check should be applied whenever data requests are received from another department:

1. There is a valid business reason to request data, such as supporting public services, internal operations, regulatory compliance, reporting, analysis, or decision-making;
2. Regulations allow the requesting department to access the data;
3. The requesting department can comply with data regulations.

All interdepartmental data must follow COIT Metadata Standards. It is also recommended that departments follow the Data Specifications Guide when sharing common fields such as Parcel or Address. Interdepartmental data stored on Department of Technology (DT) infrastructure is covered by the Data Custodian and Stewardship Policy. In all other cases, when necessary, ownership and custodianship should be defined in the data-sharing Memorandum of Understanding ("MOU").

ACCESSING NON-PUBLIC DATA

When managing and granting access to non-public data, departments must follow best practices to protect privacy and ensure the security of City data, datasets, replica datasets, derived datasets, and systems. To minimize risk, departments must follow: City and County of San Francisco Cybersecurity Requirements.

It is also recommended that departments:

- Utilized the Privacy or Security Toolkit when assessing risk,
- Follow all relevant data regulation
- Write MOUs,
- Avoid creating dataset replicas – instead granting direct access to a shared system when possible.

OPEN DATA

Data yields the most value when it is free to access, use and share – consistent with safety, privacy, and security considerations. Making open data available means it is published on the City's Open Data Portal (https://data.sfgov.org/ or successor site). All City departments must follow the Prioritization framework and Publishing process established by the CDO and the CDO must update the inventory when data is published. The default license for all Open Data is the Public Domain Dedication License (PDDL), though exceptions can be made.

EVOLUTION OF DATA MANAGEMENT

As technology evolves, so too must the data management policy. With regards to Artificial Intelligence, Departments must follow the rules and standards in Chapter 22J of the Administrative Code and future policies covering emerging technology.

ROLES AND RESPONSIBILITIES

Chief Data Officer (CDO)

The CDO is responsible for the following:

- Develop and oversee the process for creating, maintaining, and publishing the annual Data System and Dataset Inventory, including data classification;
- Develop processes and resources and support the publishing of open and internal data consistent with Citywide policies and standards
- Develop, with the City Attorney and input from departments, template agreements, supporting guidance, and a standard and timely review process for non-public data and information-sharing agreements across departments or with external partners;
- Develop, with the CCISO minimum data quality and protection standards;
- Define data architecture modeling standards and tools that departments will use;
- Engage with Data Coordinators, Stewards, and Custodians to strategically plan, and provide support and training for the publishing of data; and
- Lead & implement citywide data modernization via enabling consolidation and integration into a unified data platform to facilitate the objectives of the data management policy and the city's technology goals (done in partnership with the CIO, the Department of Technology, and Departments).

City Chief Information Security Officer (CCISO)

The CCISO is responsible for the following:

- Develop minimum data protection standards;
- Complete and publish the annual Open Data Program Risk Assessment; and,
- Advise departments on data management.

Data Coordinators

Department Data Coordinators are responsible for the following:

- Coordinate the annual Data system and Dataset Inventory process, including classification, in their department and per CDO procedures;
- Coordinate the department prioritization of data for publication;
- Coordinate publication of data on the open data portal per procedures and standards set by the CDO;
- Identify and record Interdepartmental data originating from their department; and,
- Alert CDO to any changes of the coordinator or stewards within their department.

Departmental Information Security Officer

Departmental Information Security Officers or their delegate are responsible for the following:

- Apply the 3-part check to assess Interdepartmental data requests when received;
- Make final decision on data access for Interdepartmental data requests.

Data Stewards

Department Data Stewards are responsible for the following:

- Work with the Department Data Coordinator to properly document data for which they are the steward according to the inventory requirements;
- Follow the requirements set out in the Data Classification standard and properly document the classification of data in the inventories; and
- Provide proper documentation of shared datasets to support the responsible use of data and metadata for all data, including unstructured documents, is accurately recorded and maintained.

Data Custodians

Department Data Custodians are responsible for the following:

- Work with the Data Coordinator and Data Stewards to provide appropriate documentation to support the data system inventory;
- Work with Data Stewards and Data Coordinator to support efforts as needed to make data available through the publishing process referenced in this Policy; and
- Adequately support their department's Data Stewards, Data Coordinators, and Information Security Officer in conducting their responsibilities in this Policy and in the Data Classification Standard.

Data Users

Data Users are responsible for the following:

- Responsibly use open data or data obtained from departments by reading or requesting documentation on data and applying analysis understanding any constraints on the data;
- Follow any constraints on use as specified in MOUs or other agreements where applicable;
- Be familiar with federal, state and local confidentiality or privacy laws pertaining to the data they collect, access, use, or maintain in conducting their work; and
- Adhere to data retention schedules and disposition procedures when handling data, and report any potential issues or discrepancies to the Data Steward.

City Departments

Department leadership and program management are responsible for the following:

- Assure that staff handling non-public data are sufficiently trained and aware of their duties with respect to securing and protecting private information including Data Stewards, Data Custodians and Data Users
- Make good faith effort to follow Interdepartmental data standards; and
- For those having Interdepartmental data or data standard candidates, work to meet expectations for management of Interdepartmental data and data standards as resources allow and by priority of value to the enterprise.

City Chief Information Officer (CIO)

The CIO and the Department of Technology are responsible for the following:

- Advise department leadership and program staff in technologies to support and build Interdepartmental data and data standards;
- Publish and promote standards and patterns on how to best implement the requirements for managing Interdepartmental data and data standards;
- Drive central procurement strategies to increase adoption of modern technology systems as described in Modern Data Systems and

- Partner and support citywide data modernization, in partnership with the CDO, via consolidation and integrations into a unified data platform to achieve the objectives of the data management policy and the City’s technology goals.

City Services Auditor (CSA)

CSA is responsible for the following:

- Audit this policy for compliance as needed

Office of Contract Administration (OCA)

OCA is responsible for the following:

- Explore updating the language in the Software as a Service (SAAS) contract to include stronger language related to Modern Data Systems

POLICY AUTHORITY

This policy is issued pursuant to the San Francisco Administrative Code, including but not limited to:

- **Chapter 22A** (Information and Communication Technology) which states,
    - “Considerations of both cost and the need for the transfer of information among the various departments in the most timely and useful form possible require a uniform policy and coordinated system for the use and acquisition of ICT technologies” and that,
    - COIT “shall review and approve recommendations for… ICT standards, policies and procedures to enable successful development, operation, maintenance, and support of the City's ICT.
- **Chapter 22D** (Open Data Policy) which requires the Chief Data Officer (CDO) to draft rules for the Open Data Policy and “help establish data standards within and outside the City.“

EXCEPTIONS

This policy does not preempt any Local, State or Federal regulations such as the San Francisco City Charter, the Administrative Code, Health Insurance Portability and Accountability Act (“HIPAA”), Criminal Justice Information Services Policy (“CJIS”), Homeless Management Information Systems ("HMIS") Data and Technical Standards, Sunshine Ordinance, Data Retention, etc. Further, this policy does not apply to data provided to the City by the State or Federal government, or by other municipal or regional jurisdictions.

METRICS FOR POLICY EFFECTIVENESS

To ensure the ongoing effectiveness of this Data Management Policy and to drive continuous improvement, the Office of the Chief Data Officer will publish the following metrics:

Inventory Metrics

- **Completion Rate:** Percentage of departments that have submitted a complete and up-to-date Data System and Dataset Inventory by the annual deadline.
    - Target: 100% compliance.

Interdepartmental Metrics

- **Participation Rate:** Percent of departments sharing data interdepartmentally or citywide
  - Target: 75%

Open Data Metrics

- **Publishing Rate:** Percentage of departments that are publishing data on the Open Data Portal.
  - Target: 100% compliance.

DEFINITIONS

Data System

Also referred to as an Enterprise System of Record, as defined in California Government Code §7922.630

Dataset

A dataset is defined as a structured or unstructured collection of digital information that is used in the delivery of public services, or that enables internal operations, regulatory compliance, reporting, analysis, or decision-making. For the purposes of this inventory, a dataset must be:

- **Official and authoritative.** Recognized as the source of truth for its purpose.
- **Production-level**. Used in finalized, downstream applications—not temporary, intermediary, or raw data tables used only during processing or development.

This includes structured data (e.g., databases, spreadsheets, digital maps, and CAD files) and select unstructured materials (e.g., internal knowledge bases, SOPs, photos, audio, videos, PDFs, or scanned files).

Dataset Replica

A dataset replica or replications is a copy of a dataset that is functionally the same as a dataset, but does not serve as the official and authoritative source of truth. It is commonly referred to as a: snapshot, clone, mirror, or backup.

Derived Dataset

A derived dataset is a dataset that is created by combining one or more source datasets or replica datasets to create a new asset. If a derived dataset is Official and Authoritative and Production-level, it can be treated as a new dataset with shared stewardship between the source dataset owners.

Interdepartmental Data

Interdepartmental Data is a dataset that meets a two-part definition:

- **Process dependencies**. At least one department is dependent on data from another department to conduct its work. This includes (but is not limited to) scenarios where it would be costly, redundant, or inefficient for the dependent to collect the data itself.
- **Authoritative Source**. The data is the authoritative source. Authoritative sources can be codified in federal, state or local law or through administrative policy or practice.

Non-public Data

Any dataset not classified as Level 1 (Public) using the COIT data classification standards.

Open Data

Open Data means that the dataset or replica dataset is:

- Free to download, use, and share
- Discoverable and accessible to all over the internet
- Published to minimize time between the creation and dissemination of the data
- Documented
- Provided under terms that permit re-use, redistribution, and the mixing with other datasets
- Provided in an open format that is machine-readable and accessible via API