Data Management Policy

Committee on Information Technology

PURPOSE AND SCOPE

Data is a key asset in meeting the demands of a 21st century government. Proper data management can add value to the work of the City, including:

- Improvements in data consistency and quality
- Faster, easier access to data
- Better controls and security
- Data sharing and interoperability between datasets
- Integrated data across departments
- Data analytics and more advanced data science

To deliver better outcomes, data must be proactively managed and maintained much like our capital and financial assets.

This policy applies to all information resources operated by the City and County of San Francisco and its departments, and commissions. Elected officials, employees, consultants, and vendors working on behalf of the City and County of San Francisco are required to comply with this policy.

POLICY STATEMENT

This policy establishes a framework for the management of data as an asset across the City. Departments **must** adopt this framework and the requirements below to support the ongoing, proactive management of data, which includes:

1. The identification and classification of data in **database and dataset inventories**
2. The processes and policies for appropriately sharing **open and confidential data**
3. An approach to identify **interdepartmental data and interdepartmental data standards** and actively manage those data, databases, and standards

Together, these represent the requirements for managing data as an asset. Increasing Citywide understanding of our data assets and the purposes they serve will allow the City to proactively plan for and meet the needs of an ever-changing city.

COIT Policy Dates

Approved: January 17, 2019
Next Anticipated Policy Update: June 2020

POLICY REQUIREMENTS

1.0 Database and Dataset Inventories

Managing data as an asset starts with knowing the range and type of data under the control of the City, which supports:

- improved cyber and information security,
- improved understanding of the criticality of data access to support the Citywide DPR3 policy
- proactive publication of data as appropriate, and enhanced use of shared data through the identification of data critical to improved operations

San Francisco Administrative Code Chapter 22D establishes the requirement to publish "a catalogue of the Department's data that can be made public, including both raw data sets and application programming interfaces (API's)"

The following requirements build on Chapter 22D and form a foundation for good data management:

1. **Collection of Database and Dataset Inventory.** Departments must create and maintain on an annual basis a Database Inventory and Dataset Inventory[1] per procedures and templates set by the Chief Data Officer (CDO) and with guidance from the City Attorney's Office.
2. **Classification of Databases and Datasets.** Departments must classify databases and datasets per procedures set by the CDO and City Chief Information Security Officer (CCISO) in the Data Classification Standard. Data classification is an indicator for aiding in the proper management and security of data in use, in transit and at rest. It is not meant to place any additional restrictions over and above what is required by law, or administrative policy. The classifications as submitted per the Data Classification Standard can be updated at any time if classifications don't reflect the nature of data. This can happen either through departmental review or in the course of review by the CDO or CCISO during other processes described in this policy.
3. **Publication of Database and Dataset Inventory.** The CDO must publish the Database and Dataset Inventories on the City's open data portal (or successor site) and publish updated inventories no less than annually by July 1st. The public inventories are inclusive of datasets regardless of classification except where local, State or Federal laws specify otherwise.

The Data Coordinator Guidebooks, incorporated by reference, provide more detailed direction on the database and data inventory process.

[1] Required by SF Admin. Code 22D.2(c)(4)(A)

2.0 Open Data

The greatest value from City data is realized when anyone is free to access, use and share datasets consistent with safety, privacy, and security. "Open Data" means that the dataset is:

- Available as a whole to all at no cost and discoverable and accessible 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, and
- Provided in an open format that is machine-readable.

Making open data available means it is published on the City's open data portal (https://data.sfgov.org/ or successor site) consistent with this definition.

2.1 Publishing Prioritization and Plans

Departments must prioritize data for publication and develop publishing plans as follows[2]:

1. **Publishing Prioritization.** Departments must prioritize datasets for publication per procedures set by the CDO and no less than annually as part of the Data Inventory process. In general, those datasets that have the highest public value and can easily be published should be prioritized.
2. **Publishing Plans.** Departments must submit a Publishing Plan per procedures set by the CDO and no less than annually. The Publishing Plan will describe the Department's commitments with respect to publishing data on the City's open data portal during the publishing plan timeframe. Datasets will be reviewed by the CCISO for privacy considerations prior to publishing.

The Data Coordinator Guidebooks, incorporated by reference, provide detailed direction on the process for prioritizing data for publication and for developing publishing plans.

2.2 Publishing Data

Departments must publish data per procedures set by the CDO and as follows:

1. **Publishing Portal.** Departments must submit datasets for publication to the Publishing Portal. The Submission Guidelines and Publishing Guidelines, incorporated by reference, provide more detailed direction on the publishing process.
2. **Dataset Documentation (Metadata).** Departments must document data prior to publication per the City's Metadata Standard.
3. **Data Standards.** Departments should publish data if possible, practical and available, per commonly used standards for that type of data.
4. **Non-Public Data.** Departments must use the Open Data Release Toolkit and work with the CDO to transform non-public data so that a relevant and appropriate view of that data (e.g. limited fields, aggregate data) can be published as Open Data. Non-public data

[2] Implements SF Admin. Code 22D.2(c)(3)

includes anything properly classified as Level 2 or higher per the Data Classification Standard.[3]

5. **Licensing Standard.** The default license for all Open Data is the Public Domain Dedication License (PDDL).[4] While this license is the default, other licenses may be used for specific datasets and as approved by the CDO.
6. **Recording.** The CDO must update the publishing status of datasets in the Dataset Inventory when a dataset becomes public.
7. **Language.** Data sets, including metadata, are not required to be published in additional languages beyond those used at the source. The City may opt to implemented automated language translation of data sets in the future.

3.0 Confidential Data

Data and information sharing increases our ability to improve service outcomes through more accurate evaluation of policy options, improved stewardship of taxpayer dollars, and more coordinated and personalized delivery of public services.

The City and County of San Francisco is committed to inter-agency information and data sharing as a standard practice. At the same time, it is essential to maintain public trust that confidential information is safe and secure via appropriate, strong, and effective safeguards and compliance with applicable privacy rules.

1. **Minimum data sharing, protection standards and privacy.** The City Chief Information Security Officer (CCISO) and CDO must establish minimum data sharing and protection standards using a risk based approach consistent with the Citywide Data Classification Standard and the needs and requirements of interdepartmental data and information sharing.
2. **Data sharing agreements and resources.** The City Attorney's Office, and with input from departments, must provide template agreements, supporting guidance and a standard and timely review process for confidential data and information sharing agreements across departments or with external partners.
3. **Departmental processes.** Departments must protect confidential data per applicable law and regulations. Departments should work to maximize the value of data and data sharing for public good purposes while appropriately managing risk. In this context, Departments should actively pursue interdepartmental data sharing for reasons such as program evaluation, research, analysis, care coordination, operations, or other public good purposes and per citywide standards and resources for data sharing and protection.

[3] Supports SF Admin. Code 22D.2(c)(4)(G)

[4] License terms available here: http://opendatacommons.org/licenses/pddl/1.0/.

4.0 Management of Interdepartmental Data

Proper interdepartmental data management can unlock value and control costs for the City. Where data can serve multiple programs, initiatives or client groups across more than one department, the City should adopt data management practices.

Data management means that data identified as interdepartmental is actively managed to:

1. **Support integrations.** Data should be accessible in a manner that users can develop self-service integrations.
2. **Minimize redundancy and errors.** Data should be easily accessible to minimize the production of derivative data products that conflict with the authoritative source.
3. **Ensure data quality.** Data should be monitored and managed to maintain data quality appropriate to the requirements of its use.
4. **Scale with the needs of users.** Data stewards should solicit, vet, prioritize and implement changes that meet the needs of the users over time and as resources allow.
5. **Properly control access.** Access to data should be managed commensurate with risks.
6. **Be clearly documented.** To promote responsible data use, data must be documented in a tool that promotes proper maintenance and accessibility of documentation.

Done well, properly managed data should bring down the overall costs to the organization by reducing errors and duplication of effort, while also increasing effective use of interdepartmental data.

In pursuit of adopting appropriate data management practices, this policy defines:

1. **Interdepartmental data.** Criteria for interdepartmental data and related requirements.
2. **Interdepartmental data standards.** Criteria for interdepartmental data standards and related requirements.
3. **Implementation.** How the City should ensure interdepartmental data and standards are managed per this policy.

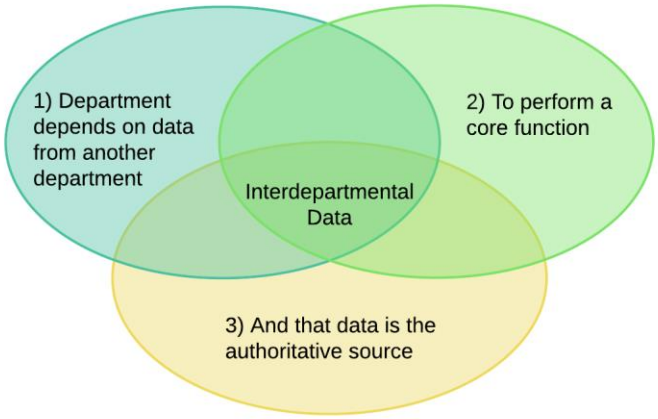
4.1 Interdepartmental Data

Definition / Scope. Interdepartmental Data is data that has substantial value when used across more than one department. Properly managed, interdepartmental data should also unlock efficiencies across departments. Data can be characterized as interdepartmental data when it meets a three-part definition:

1. **Process dependencies.** At least one department is solely dependent on data from another department to conduct its work.
2. **City and County core functions.** The data supports work that is integral to the operations and management of core functions for the City and County. A core function is a service, program or activity authorized by statute, regulation, the Administrative Code or other authorizing authority. It does not include functions in support of core functions such as research or planning.

3. **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.

Per this three-part definition, not all departments will have interdepartmental data under their stewardship.



4.2 Interdepartmental Data standards

Definition / Scope. Interdepartmental Data standards are those standards that are valuable when applied across more than one department and can include:

- lists of permissible values like those used in a lookup or as an identifier or code,
- formats, and
- methods for data collection, sharing and/or reporting

Standards can be applied to common attributes used across the City, including but not limited to parcel identification, demographics, department names and codes, sexual orientation and gender identity.

Standards allow easier integration of data across the City and enable consistent and comparable reporting across City agencies.

Per this criteria, not all departments will have Interdepartmental Data standards under their stewardship, but all departments are expected to use Interdepartmental Data standards to the extent reasonable.

Departments must use data standards as listed in the Data Standards Reference Handbook unless otherwise required.

4.3 Identification of and Requirements for Interdepartmental Data and Standards

Interdepartmental data should be identified, prioritized and aligned strategically over time. To that end the Committee on Information Technology and the CDO encourages and supports cross-department working groups to:

1. Identify Interdepartmental Data and Standards per this policy.
2. Rank and prioritize these based on the organizational value they can provide.

Based on the results on the working group, Departments managing Interdepartmental Data or Standards must work to implement the requirements in the Appendix.

5.0 ROLES AND RESPONSIBILITIES

5.1 Chief Data Officer. The CDO is responsible for the following:

- Develop and oversee the process for creating, maintaining and publishing the annual Database and Dataset Inventory, including data classification;
- Develop processes and resources and support the publishing of open 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 confidential data and information sharing agreements across departments or with external partners;
- Develop, with the City Chief Information Security Officer, minimum data sharing and protection standards;
- Define data architecture modeling standards and tools that will be utilized by departments; and
- Engage with Data Coordinators, Stewards and Custodians to strategically plan, and provide support and training for the publishing of data,

5.2 City Chief Information Security Officer. 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.

5.3 Data Coordinators. Department Data Coordinators are responsible for the following, and according to Administrative Code Chapter 22D:

- Work with the CDO to coordinate implementation of and oversee compliance with the Data Policy within their department;
- Coordinate the annual Database and Dataset Inventory process, including classification, in their Department and per CDO procedures;
- Coordinate the department prioritization of data for publication and the creation of department plans and timelines for publishing data; and
- Coordinate publication of data on the open data portal per procedures and standards set by the CDO.

5.4 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.

5.5 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 database 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, Cybersecurity Officer in conducting their responsibilities in this Policy and in the Data Classification Standard.

5.6 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.

5.7 Departments. Department leadership and program management are responsible for the following:

- Assure that staff handling confidential 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 identify Interdepartmental data and 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.

5.8 City Chief Information Officer. 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; and
- Help implement any technology and integrations for managing Interdepartmental data and data standards.

5.9 Controller City Services Auditor Audits. CSA Audits is responsible for the following:

- Audit this policy for compliance as needed.

REFERENCE

- Cybersecurity Policy
- DPR3 Policy
- Data Classification Standard
- Metadata Standard

APPENDIX: Requirements for Interdepartmental Data. Departments managing Interdepartmental Data should work to accomplish the following for that data:

1. **Establish appropriate change management and user feedback practices.** Establish feedback processes appropriate to the complexity of the data. These could include one or more of the following:
    a. Create data user groups with representation from major users of that data to elicit ongoing needs or solicit feedback when planning changes.
    b. Provide opt-in communications with users of data through a listserv or similar broadcast channel.
    c. Develop processes to report errors and reconcile those errors within a published target service level.
2. **Provide clear, accessible documentation.** At a minimum, document data per the Data Standard. Interdepartmental Data stewards are encouraged to supplement the minimum standard with additional information and explanations, including wikis or similar collaborative documentation.
3. **Provide means for accessing the data as appropriate to City Departments.**
    a. Develop an access control policy to determine who may be granted access to the data based on a risk assessment and consistent with appropriate privacy protections.
    b. Develop and manage a standard, clear and visible process for granting access to the data and consistent with the access control policy. Minimally make this process available through a discoverable URL.
    c. Make the data available on any citywide data sharing platforms, including the open data portal if the data can be made publicly available and consistent with the access control policy.
    d. Provide the data as a service preferably through a documented Application Programming Interface or other web service.
4. **Monitor and report on data.**
    a. Establish a data quality monitoring program and report on data quality metrics. The Data Quality resource collection provides guidance on establishing a data quality monitoring program.
    b. Establish and monitor service availability metrics and report on these in a consistent, preferably automated manner.

Requirements for Interdepartmental Data Standards. The following requirements apply to Interdepartmental Data standards:

1. Stewards for Interdepartmental Data standards must make the standard available as a service preferably through a documented API. Publishing the data on the open data portal meets this requirement.
2. Interdepartmental Data standard stewards are responsible for the management and quality control of their respective data standards.
3. Interdepartmental Data standard stewards must set clear, visible feedback mechanisms for reporting errors or issues with the data standard.