DATA GOVERNANCE

Salt Lake County
Data Governance Working Group

DATA GOVERNANCE GUIDEBOOK

June 2022

Table of Contents

Background

Purpose of Data Governance Guide

The purpose of this guidebook is to provide guidance to Data Coordinators in Salt Lake County government. Data Coordinators should use this guidebook to help them in their role. We'll update this guide as the roles and responsibilities of the Data Coordinators evolve and as we learn more. We appreciate your patience and willingness to remain flexible as we embark on this endeavor together.

Data Governance Working Group

What is Data Governance Working Group?

Data Governance Working Group was created under the Technology Advisory Board (TAB) and the GIS Steering Committee to ensure that appropriate mechanisms are in place to establish a culture of operational excellence that recognizes and supports institutional data as an asset of the County.

Who is in Data Governance Working Group?

Data Governance Working Group is staffed by Office of Data & Innovation. The members include representatives from the Council, Mayor's portfolio, Elected Offices and subject matter experts from Information Technology, Contracts & Procurement, Records Management and Archives, and Risk Management.

What Does Data Governance Working Group Do?

Data Governance Working Group is responsible for advising TAB and GIS Steering Committee on strategic guidance and implementation of the data governance program, prioritization for the data governance projects and initiatives, approval of organization-wide data policies and standards, as well as enabling ongoing support, understanding and awareness of the data governance program. The Group is also responsible for providing training and guidance to the County Data Coordinators.

Why We Need Data Governance

The term "Data is the new gold" describes the value and liability associated with the data. Salt Lake County agencies are generating new data every single day including but not limited to payment records, payroll, field survey, clients and patron information, registrations, property and tax records, GIS spatial data, and many countless other functions throughout the County contribute to the huge data.

The main role of data governance is to ensure that the data quality remains high throughout the complete lifecycle of the data and the controls which are implemented are in line with the organizations business objectives. It is important that information is used both effectively and efficiently and is in line with the County's intentions. Data governance identifies who can

take what action, as a result of what data, in which situations, and using what methods.

Data Governance Value Propositions

Salt Lake County Data Governance Working Group has identified six (6) value propositions to guide the County's data governance efforts. Here the Data Governance value propositions:

- Data as a Strategic Asset
- Data Catalog/Library
- Data Security
- Data Integrity
- Data Accessibility & Sharing
- Data Risk & Liability

Data as a Strategic Asset

Definition: Data and content of all types are assets with all the characteristics of any other asset. Therefore, they should be managed, secured, and accounted for as other material or financial assets.

Value Proposition: Advancing the understanding of data as a strategic asset will allow the implementation of and adherence to effective data governance policies and processes

Data Catalog / Library

Definition: A detailed and comprehensive data inventory that makes data appropriately discoverable.

Value Proposition: Knowing what information is available and how it can be related allows for new insights, knowledge, and wisdom.

Data Security

Definition: Data Security means protecting digital data from unauthorized access and unwanted actions.

Value Proposition: Knowing what data is restricted or protected allows implementation of proper security measures to protect the privacy and confidentiality of the data.

Data Integrity

Definition: Data integrity is the overall accuracy, completeness, relevance, timeliness, and consistency of data and metadata.

Value Proposition: Maintaining data integrity allows users to trust data and generate reliable insights. Strong data integrity guidelines support validity, stability, searchability, and recoverability.

Data Accessibility & Sharing

Definition: Data access is the on-demand, authorized ability to retrieve, modify, copy, or move data from IT systems based on organizational roles or responsibilities. Data sharing means sending data, receiving data, or both to advance shared objectives according to certain terms & conditions.

Value Proposition: Proper data access and ability to share with internal and external partners allows County agencies to leverage the power of data to advance greater public good.

Data Risk & Liability

Definition: The risk in information means there is a financial liability inherent in all data or content that is based on regulatory and ethical misuse or mismanagement.

Value Proposition: Recognizing risks and liability associated with data will promote appropriate measures for data collection, retention, disposal, and security to protect county from litigation and malicious ransomware attacks.

Data Coordinators' Roles and Responsibilities

Salt Lake County's Information Technology Standards on Data Classification and Protection states that, "County Agency Management shall designate a County Agency Data Coordinator who will be assigned to work with the Office of Data Innovation." Data Coordinators roles and responsibilities include:

- Act as single Point of Contact for Data Gov Working Group
- Serve as a liaison with Data Gov Working Group on issues related to data gov.
- Attend trainings & workshops on data management
- Assist with Applications inventory
- Assist with Data Catalog inventory
- Assist with implementing privacy, data licensing, metadata and other standards and practices
- Provide feedback regarding data management initiatives to Data Gov Working Group

Continued…

Table 1: Data Coordinators Roles & Responsibilities*

Role	Qualifications	Time-Commitment	Tasks
Act as single Point of Contact for Data Gov Working Group	Basic Understanding of agencies operations including IT, Programs, Fiscal & Personnel.	Low	As needed Attend occasional meetings
Serve as a liaison with Data Gov Working Group on issues related to data gov.	Familiarity with agency's IT systems and databases	Low	On going Attend occasional meetings and participate in discussion regarding data gov issues
Attend trainings & workshops on data management	Basic understanding of agency IT operations	Low	Participate in trainings & workshops to understand their roles and resources required for systems/data inventory.
Assist with applications inventory	Familiarity with agency's systems and databases	Moderate	Provide list of systems information maintained by their agency. May need to talk to system administrators and users.
Assist with Data Catalog inventory	Close working relations with the agency systems and database administrators.	High	Work closely with their agency systems/database administrators to provide details about the database, its content, identify classification and priorities.
Attend trainings & workshops on data management	Basic understanding of agency IT operations	Low	Participate in trainings & workshops to understand their roles and resources required for systems/data inventory.
Assist with implementing privacy, data licensing, metadata and other standards and practices	Be a data champion!	Low	Ongoing
Provide feedback regarding data management initiatives to Data Gov Working Group	Share best practices and ideas!	Low	Ongoing

*Roles and responsibilities may evolve as new initiatives and projects are developed. These are intended to give a general idea of what to expect as a Data Coordinator.

Data Coordinators are at the forefront of supporting and implementing data governance at Salt Lake County. They work with Data Gov Working Group to establish procedures for the responsible data management and governance. Data Governance Working Group will provide guidance and training to Data Coordinators.

Data Inventory

What is Data Inventory?

A data inventory is a fully described record of the data assets maintained by an organization. The inventory records basic information about a data asset including its name, contents, update frequency, use license, owner/maintainer, privacy considerations, data source, and other relevant details. The details about a dataset are known as metadata.

Why Conduct an Inventory of our Data?

One of the data governance value propositions is to treat data as a strategic asset of Salt Lake County which is "data and content of all types are assets with all the characteristics of any other asset. Therefore, they should be managed, secured, and accounted for as other material or financial assets" followed by Data Catalog / Library that states, "A detailed and comprehensive data inventory that makes data appropriately discoverable" to "know what information is available and how it can be related allows for new insights, knowledge, and wisdom."

Providing an accessible data inventory will make the County employees' jobs easier when they need information from another department - they will know what exists and how to find it. The same benefits apply to the public regarding its search for County information. Having a complete inventory is also important when determining which datasets to release publicly. It's not feasible to release all County's public datasets at once, so decisionmakers need a prioritization strategy. The data inventory can be used to prioritize the release of data according to strategic priorities, public interest, etc.

Data Inventory Framework

Data Governance Working Group is responsible to design and establish data inventory framework and processes to collect data inventory and maintain the data. Data Coordinators are responsible for conducting and submitting data inventory data through coordination with the agency staff.

The data inventory framework will include the following four steps:

1. Identification of data sources (applications) and completion of the SLCo Application Inventory Survey
2. Identification of data sets within these data sources and completion of the SLCo Dataset Inventory Survey
3. Review process and gap analysis
4. Develop an accessible data catalog

It is important to ensure data inventory accuracy and reliability. Data Governance Working Group will establish a periodic review process to updates the survey data.

Table 2 provide an overview of the Data Inventory Framework.

Table 2: Data Inventory Framework

	1 Data Sources Inventory (Applications)	2 Dataset Inventory (Data Catalog)	3 Gap Analysis	4 Data Catalog
Definitions	Data Source: Technology or system that stores data, including databases, named spreadsheets, information systems, business applications, etc.	Dataset: Contents of a single database table, worksheet, or defined view; data is provided as a single combination of unique rows (or records) and corresponding columns (or fields) describing that row. (Building permits dataset contains all records for a given timeframe)	Gap Analysis: A process of identifying missing data sources and datasets.	Data Catalog: A data catalog maintains an inventory of data assets through the discovery, description, and organization of datasets. The catalog provides context to enable data analysts, data scientists, data stewards, and other data consumers to find and understand a relevant dataset for the purpose of extracting business value. Gartner - December 2017
What is it	What are your authoritative data sources?	What are all the single datasets you can pull from the data sources?	What information is missing?	A data catalog informs customers about that available data sets and metadata around a topic and assists users in locating it quickly.
Examples	Microsoft Office, ESRI GIS, PeopleSoft, SharePoint, Adobe, VueWorks, EZMRx etc.	Payment records, Field survey, Employee list, Address points, Building permits, Marriage license, Property records etc.	What information is missing? Standardization of terms, cleanup etc.	An accessible online portal to view applications and dataset inventory.
Responsibility	Data Coordinators	Data Coordinators Database User / Administrators SME Group	Subject Matter Experts / IT	Enterprise Architect / IT

Step 1: Complete SLCo Application Inventory Survey

SLCo Applications Inventory survey provides a high-level overview of applications and sources of data. Your data may be housed in a variety of places from information systems or databases to shared drives and folders. This can also include 3rd party vendors and data hosted on vendor systems. Step 1 is about identifying the major data sources in your department.

Questions to help identify and discover data sources:

- What information systems does your department use?
- What databases does your department use?
- What applications capture information or are used in your business processes?
- Are some data resources kept in spreadsheets (on shared or individual drives)?
- What services does your department deliver, and how is information related to those services stored?

For each of the data sources, you will be asked to provide the following details for each of your department's Data Sources:

Agency Name	Data Coordinator Information	Application Name
Brief Description of Application Data	Application Administrator	Application Vendor Name
Application Type	DR Continuity of Business Tier	Application Data Type
Applicable Privacy Standard	Additional Comments	

SLCo Application Inventory Survey
(Please use Appendix A to get familiar with the terms used in this survey)

Survey Link: https://app.smartsheet.com/b/form/d458dcc017b3475c935adce5097634be

- ***Agency Name:*** Select your agency from the drop down.
- ***Designated Data Coordinator Information:*** Provide Name & Email Address of the designated data coordinator.
- ***Application Name:*** What is the name of the data source / application?
- ***Brief Description of Application Data***: Briefly describe the types of data that are processed by the system/application. What goes into and comes out of the system/application
- ***Application Administrator***: Who manages this application in your agency including managing upgrades, setting up access, and providing support to users?
- ***Vendor Name***: Who did you buy this application from?
- ***Application Type***: Select one of the following.
  - SLCO/IT Hosted
  - SLCO/Agency Hosted
  - Software as a Service (SaaS)
  - Installed on a Local Computer
  - ERP System (e.g., PeopleSoft)
  - IT Cloud
- ***Disaster Recovery Continuity of Business Tier:*** Select an appropriate Tier from the list:
  - Tier 1 - Service Restored in Hours

- o Tier 2 - Service Restored in Days
- o Tier 3 - Service Restored in Weeks
- o Tier 4 - Service Restored in Months
- o Tier 5 - Agency provided DR/COB plan
- o Not Sure
- *Application Data Type:* Select an appropriate data type from the list:
  - o Public Data
  - o Protected Data
  - o Restricted Data
  - o Not Sure
- *Applicable Privacy Standards:* Select applicable privacy standard(s) from the list (select all that apply)
  - o 42 CFR Part 2
  - o Communicable Disease Rule
  - o PII - Personally Identifiable Information
  - o PCI DSS - Payment Card Industry Data Security Standard
  - o HIPPA - Health Insurance Portability and Accountability Act
  - o CJIS - Criminal Justice Information
  - o Not Sure
  - o Other
- *Additional Comments*: Do you have any other comments about this Data Source?

Next Steps: Once an agency submits the SLCo Application Inventory Survey, the Data Gov Subject Matter Expert (SME) group will review the data for accuracy, reliability, and clarification. Additional changes may be made to the survey data based on the review.

Step 2: Complete SLCo Data Catalog Questionnaire

Each of your data sources should have an associated dataset(s). Some data sources may be datasets themselves. The purpose of Step 2 is to allow Data Coordinators to spend time with Data Administrators in their departments brainstorming datasets to be included on the SLCo Data Catalog Questionnaire. Agencies may ask, "what should be included as a dataset?" We've defined a dataset as the contents of a single database table, worksheet, or defined view. For example, if your data source is comprised of tables, listing out the tables may be a good place to start.

To help brainstorm, use the questions below:

- What data is critical for your agency's continuity of operations?
- What data populates your monthly or quarterly reports?
- What data does your department use for internal performance and trend analysis?
- What data is reported to federal, state, or local agencies?

- What data demonstrate value-added to the County?
- What data do other departments ask for?
- What data do the public ask for?

Caution: Don't exclude any datasets based on privacy or confidentiality concerns! Our goal is to have a holistic picture of our data. We do not plan to publish the dataset but rather inform the stakeholders of data availability.

SLCo Dataset Catalog Questionnaire
(Please use Appendix A to get familiar with the terms used in this survey)

Survey Link: https://app.smartsheet.com/b/form/edb28795b6d24414bc902114bfb66cb6

The details for each dataset are similar to those requested in the Applications Inventory but include a bit more detail. For each of the dataset, you will be asked to provide the following details for each of your department's datasets:

County Agency	Agency Data Coordinator	Data Steward	Application Name	Vendor / Developer
Dataset Name	Dataset Description	Master Copy	Data Source	Data Format / Type
Primary Field	Raw or Enrich Data?	Update Frequency	API Type	Retention Schedule
Disaster Recovery Tier	Data Category	Privacy Data Standards	Dataset Location	Dataset Users

- *Agency Name:* Select your agency from the drop down.
- *Designated Data Coordinator Information:* Provide Name & Email Address of the designated data coordinator.
- *Data Administrator:* Who maintains this dataset? Who should be contacted with questions related to this dataset?
- *Application Name:* What is the name of the data source / application?
- *Vendor Name*: Who did you buy this application from?
- *Dataset Name*: The title or name of the dataset.
- *Dataset Description*: Provide a brief description of the contents of the dataset. What is its purpose? What kinds of information does it contain?
- *Master Copy*? Check the box if data is considered the original source of information or master copy of the datasheet.
- *Data Source*: Which Data Source Does This Dataset Come From (if applicable): Does

this dataset come from a larger database or system? If so, what system or database?

- *Available Data Formats*: In what formats can this data be exported (e.g., PDF, CSV, Access, KML, Word, etc.)?
- *Primary Field:* What is the unique primary field in the dataset?
- *Raw or Enrich Data?* Is the data the original (raw) data or is it data that has been pulled and enriched with additional information?
- *Update Frequency:* How often is the data within the dataset refreshed with new information?
    - Real-time / Feed
    - Daily
    - Weekly
    - Monthly
    - Yearly
    - On-Demand
    - Historic – Not updated
- **API:** does this dataset has an application programming interface (API)?
    - Rest
    - SOAP
    - Other
    - Unsure
- *Retention Schedule:* How long is the retention schedule for this data? If longer than 10 years, select permanent?
    - 1 year – 10 years
    - Permanent
    - Unsure
- *Disaster Recovery Continuity of Business Tier:* Select an appropriate Tier from the list:
    - Tier 1 - Service Restored in Hours
    - Tier 2 - Service Restored in Days
    - Tier 3 - Service Restored in Weeks
    - Tier 4 - Service Restored in Months
    - Tier 5 - Agency provided DR/COB plan
    - Unsure
- *Data Category:* Select an appropriate data type from the list:
    - Public Data
    - Protected Data
    - Restricted Data
    - Not Sure
- *Applicable Privacy Standards:* Select applicable privacy standard(s) from the list (select all that apply)
    - 42 CFR Part 2
    - Communicable Disease Rule
    - PII - Personally Identifiable Information

- o PCI DSS - Payment Card Industry Data Security Standard
- o HIPPA - Health Insurance Portability and Accountability Act
- o CJIS - Criminal Justice Information
- o Not Sure
- o Other
- *Dataset Location:* Where is the dataset located? i.e., SharePoint, Shared Drive, Cloud, Local Computer etc.
- *Dataset Users:* Who are the potential users of this dataset?
  - County Agency
  - Other Gov Agencies
  - Non-profit Organization
  - Educational Institutes
  - Public
  - Not sure
- *Additional Information / Comments:* Use this space to provide any additional information regarding this dataset.

Step 3: Review Process

The last step of the inventory process will consist of a detailed review and gap analysis of each department's submissions. Data Gov SMEs will meet with department Data Coordinators and Data Administrators to review the information and ensure that all data sources and datasets are accounted for, especially those that are directly related to major departmental services and/or operations. While we may never ask for the data in a given dataset, it is important for us to have a comprehensive understanding of all available County data. This will ensure that we are adequately prepared to use our data in new and exciting ways.

Next Steps: Data Catalog

A data catalog is a record of an organization's existing data. It is a library where an organizations' data is indexed, organized and stored. Most data catalogs contain data sources, data usage information, and data lineage that describes the origin of the data and how it changed to its final form. With a data catalog, organizations can centralize information so that they can identify what data they have, distinguish data based on its quality and source. - https://research.aimultiple.com/data-catalog/

Appendix A - Glossary of Terms

Term	Definition
Application	An application is any program, or group of programs, that is designed for the end user. Applications software (also called end-user programs) include such things as database programs, word processors, Web browsers and spreadsheets. Examples: PeopleSoft, ADP, ESRI GIS, Power BI, Microsoft Office (Excel, Word, Access).
Protected Data	Protected data is information that must be guarded due to proprietary, ethical, or privacy considerations, and must be protected from unauthorized access, modification, transmission, storage or other use
Public Data	Public data is information that may or must be open to the general public. Public data has no existing local, national or international legal restrictions on access or usage.
Restricted Data	Restricted data is information protected by federal or state statutes or regulations (e.g. HIPAA), County ordinance (e.g. Ordinance 2.81), contractual language (e.g. PCI-DSS), or licensed data and must be protected from unauthorized access, modification, transmission, storage or other use
Data Coordinators/Steward	Data Coordinators are designated for each agency as the main point of contact and liaison
Data Custodians	Work with the Data Governance Working Group on data governance and standards issues. [The ones who manage the infrastructure of the dataset]
42 CFR Part 2	Protection of patient records around substance abuse
CJIS - Criminal Justice Information	Criminal Justice Information (CJI) - Criminal Justice Information is the abstract term used to refer to all the FBI CJIS provided data necessary for law enforcement agencies to perform their mission and enforce the laws, including but not limited to biometrics, identity history, person, organization, property (when accompanied by any personally identifiable information), and case/incident history data. In addition, CJI refers to the FBI CJIS-provided data necessary for civil agencies to perform their mission, including, but not limited to data used to make hiring decisions. The following types of data are

	exempt from the protection levels required for CJI: transaction control type numbers (e.g., ORI, NIC, UCN, etc.) when not accompanied by information that reveals CJI or PII.
Communicable Disease Rule	Data protection around spreadable diseases
HIPPA - Health Insurance Portability and Accountability Act	Protection of personal health records
PCI DSS - Payment Card Industry Data Security Standard	Protection of cardholder data
PII - Personally Identifiable Information	Protection of individuals information that can be used to identify them.
Data Source	A data source may be the initial location where data is born or where physical information is first digitized, however even the most refined data may serve as a source, as long as another process accesses and utilizes it. Concretely, a data source may be a database, a flat file, live measurements from physical devices, scraped web data, or any of the myriad static and streaming data services which abound across the internet.
Data Format	The form that the data is presented in. Examples include PDF, CSV, GIS SHAPE FILE, SPREADSHEET, TEXT, IMAGE, etc.
Dataset	A data set (or dataset) is a collection of data. In the case of tabular data, a data set corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set in question.
Dataset Location	Where the dataset is stored or located if different from the data source.
Dataset Users	The users and/or stakeholders that use the dataset to perform analysis or other job functions.
Disaster Recovery Tier 1	Less than 24-hours RPO - Less than 1-hour RTO - Less than 24-hours
Disaster Recovery Tier r 2	Less than 1-day PRO - Less than 1-hour RTO - Less than 24-hours
Disaster Recovery Tier 3	Less than 7-days RPO - Less than 24-hours RTO - Less than 1-week

Disaster Recovery Tier 4	Less than 1-month RPO - Less than 24-hours RTO - Less than 1-month
Disaster Recovery Tier 5	Agency provided DR/COB plan
Enrich Data	Data that has been enhanced from its raw form. This can include but not limited to cleaning, supplementing, summarizing, aggregating, etc.
Primary Field	Metadata that uniquely identify a individual record.
Raw Data	Raw data, in data management, is the collection of information as gathered by the source before it has been further processed, cleaned or analyzed.
Record Copy	The single copy of a dataset, often the original, that is designated as the official copy for reference and preservation.
REST API	A REST API (also known as RESTful API) is an application programming interface (API or web API) that conforms to the constraints of REST architectural style and allows for interaction with RESTful web services. REST stands for representational state transfer and was created by computer scientist Roy Fielding.
Retention Schedule	A systematic plan establishing how long information must be kept for legal and operational requirements and the guidelines for how to dispose of it.
SOAP API	Simple Object Access Protocol (SOAP) is a message specification for exchanging information between systems and applications.
Update Frequency	How often the data will be updated eg daily, monthly, weekly, yearly, on demand, real time.
Vendor	The company or organization that the application is purchased from.

Appendix B – Data Classification and Protection

IT Standard - Data Security and Protection.pdf