> “BEING ABLE TO DEMONSTRATE EFFECTIVE OWNERSHIP AND STEWARDSHIP OF DATA ELEMENTS IS KEY TO DATA GOVERNANCE.”

– NATHAN SNYDER
PARTNER AT BRICKENDON CONSULTING

SECTION 1

DATA GOVERNANCE

1. DATA GOVERNANCE OVERVIEW

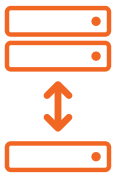
"Why do we care?" about data governance and some key concepts to get started.

Data Strategy & Policy for Good Data Governance – Section 1.1

- Data strategy & why it's needed (1.1.1 – 1.1.2)
- Examples of data strategy (1.1.3 – 1. 1.4)
- Data policy & why it's needed (1.1.5)

Key Aspects of Data Governance – Section 1.2

- Identify what's collected (1.2.1)
- Create data inventories (1.2.2)
- Assigning roles and responsibilities (1.2.3)
- Classify data (1.2.4)
- Define data access (1.2.5)
- Ensure data quality (1.2.6)
- Data retention & destruction (1.2.7)
- Preserve data (1.2.8)
- Use data efficiently (1.2.9)
- Ensure data oversight (1.2.10)



Key Considerations for Data Sharing – Section 1.3

- Data sharing & integration (1.3.1)
- Risks & benefits of data sharing (1.3.2)
- Role of data professionals in data sharing (1.3.3)
- Types of data sharing agreements (1.3.4 - 1.3.5)
- Responding to data requests (1.3.6)
- Data de-identification (1.3.7)
- Prepare a successful data request (1.3.8)

Understanding Open Data – Section 1.4

- Open data & its guiding principles (1.4.1)
- Move to open data (1.4.2)
- Draft your open data policy (1.4.3)

Check out the resource repository at the end of the section.

1. DATA GOVERNANCE

WHY DO WE CARE?

Cities and communities generate data by monitoring an array of activities, including pedestrian and vehicle traffic monitoring, waste and air quality management, and many other use cases. Data collection and management forms the basis of smart city applications. As such, a city or a community's success depends on how well it uses its data to improve residents' and visitors' experiences.

Data Governance is often used as a catch-all term to define how data are collected, stored, protected, used, and shared. For our discussion, we define data governance as the implementation and enforcement of a set of policies and practices to manage and use data so that cities and communities can extract maximum value from it to provide a better quality of life to its residents and organizations.

Data governance's goal is to break information silos, harmonize data systems and practices, and create a data-driven organization. As cities ramp up the utilization of data to make decisions regarding delivering public goods and services, it becomes critical to institutionalize data practices, policies, and roles that guide how the data is managed and shared.

Good data governance is essential for creating well-functioning, sustainable, and innovative smart cities and communities. In this section, we discuss the following:

1. Establishing a Data Strategy and a Model Policy for Good Data Governance
2. Key Aspects of Data Governance
3. Key Considerations for Data Sharing
4. Understanding Open Data
5. Data Governance Resource Repository for Cities and Communities

KEY DEFINITIONS:[1]

- **Administrative Data** – Data which is derived from the operation of administrative systems (e.g., data collected by government agencies for the purposes of registration, transaction and record keeping, which is then used for business and statistical purposes).
- **Dataset** – A particular collection of data, curated for a specific purpose or classification. Data Assets – Data collected and/or sourced and stored by an organization.
- **Data Classification** - A way to group a set of related categories in a meaningful, systematic, and standard format based on their sensitivity level.
- **Data Management** – The practice of collecting, keeping, and using data securely, efficiently, and cost-effectively with the goal to maximizing benefits to the organization.
- **Data Sources** – A place or system or service where data are obtained.
- **Metadata** – It is data that provides information about one or more aspects of the data.

1 Oracle. (n.d.). Data Management. Retrieved April 2021, from https://www.oracle.com/database/what-is-data-management/#link1 ; Government of New Zealand. (n.d.). Data Capability Framework Guide. Data.Govt.Nz. Retrieved April 2021, from https://www.data.govt.nz/manage-data/data-capability-framework/data-capability-framework-guide/#Glossary ; NIST. (n.d.). Computer Security Resource Center. Retrieved April 2021, from https://csrc.nist.gov/glossary.

1.1. ESTABLISHING A DATA STRATEGY AND A MODEL DATA POLICY FOR GOOD DATA GOVERNANCE

1.1.1 WHAT IS A DATA STRATEGY?

A data strategy is a comprehensive vision document that guides the city or community's goals, policies, practices, and principles. It outlines how to optimally use data to maximize the value of the city or community and the people it serves.

1.1.2 WHY SHOULD CITIES AND COMMUNITIES HAVE A DATA STRATEGY?[2]

Regardless of the size of the city or a community, a formalized and well documented data strategy creates the foundation of a shared vision and can help unlock the strategic benefits of using data to make better decisions that can efficiently deliver public goods and services.

TIP: CHECKLIST FOR CREATING A DATA STRATEGY

- Set your vision (Refer to pg. 7 of Data Governance Playbook)
- Assess your data maturity (Refer to pg. 9 of Data Governance Playbook)
- Establish data principles (Refer to Federal Data Strategy principles or DOD's principles or FAIR principles for data management)
- Define your goals
- Implement timelines, metrics, and data practices (Refer to pg. 13 of Data Governance Playbook)

1.1.3 WHAT ARE SOME GOOD EXAMPLES OF A DATA STRATEGY?

Some prominent data strategies from federal agencies, state and local governments include:

- **Federal Data Strategy** – includes a 10-year roadmap for federal agencies. It has four components that provide guidance for federal data use and management. The document includes a mission statement, 10 timeless principles, 40 practices to operationalize the principles, and 20 action steps to implement the practices.
- **Department of Defense (DOD)'s Data Strategy** – outlines DOD's vision, focus areas, and eight guiding principles for all their data efforts. The strategy also identifies four essential capabilities needed to achieve the agency's seven goals which include making the data visible, accessible, understandable, linked, trustworthy, interoperable, and secure.
- **California's Data Strategy** – outlines the mission and vision for the state of California. It further provides three goals and 10 objectives to empower state agencies to use data to make better decisions.

2 Government Chief Data Steward. (2018, December). Data Strategy and Roadmap for New Zealand [Slides]. Government of New Zealand. https://www.data.govt.nz/assets/Uploads/data-strategy-and-roadmap-dec-18.pdf

1.1.4 WHAT IS A DATA POLICY? HOW IS IT DIFFERENT FROM DATA STRATEGY?

A data policy outlines the basic principles, tasks, and procedures regarding proper data handling and data lifecycle management. It also attempts to define the roles and responsibilities of data management to help cities and communities think through their data policies and staffing.

QUESTIONS TO CONSIDER

- What outcomes are you trying to achieve?
- How will the data help to meet organizational needs?
- How are you going to track, monitor, and assess the implementation of the strategy?
- How will you communicate the strategy with the staff and stakeholders?

TABLE 1. DIFFERENCE BETWEEN DATA STRATEGY AND DATA POLICIES

DATA STRATEGY	DATA POLICIES
Data Strategy provides an overarching vision and plan for the entire organization.	Data Policies provide more granular details about how the data will be handled across the organization.
It is not legally binding.	It can be legally binding if properly and formally adopted.

QUESTIONS TO CONSIDER

- Who should be involved in the process of creating a data policy?
- How will the policies be communicated to all stakeholders?
- If policies are not legally binding, how will they be enforced?

1.1.5 WHY A DATA POLICY IS NEEDED?

A set of institutionalized data policies help to create: (i) consistent access to data across agencies; (ii) consistent storage format, structure, and vocabulary; (iii) clarity of roles and responsibilities; (iv) shared and open flow of information; (v) clarity about what data are collected and by whom; and (vi) proper attention to privacy, access, and retention.

TIP

Refer to US Ignite's Data Standards and Policies database for municipal data policies.

Also, refer to the recently published white paper on Smart City Data Governance Policies for recommendations on how to create policies for digital transformation and data sharing needs for the future.

1.2. KEY ASPECTS OF DATA GOVERNANCE

You cannot govern data when you don't know what data you have and who has access to it. Therefore, the bare minimum of data governance is to identify what your organization is collecting, identifying who is responsible for storing it or owning it, and giving them specific responsibilities and guidelines.

1.2.1 HOW CAN CITIES AND COMMUNITIES IDENTIFY WHAT DATA ARE BEING COLLECTED?[3]

One of the very first steps toward good data governance is identifying what is being collected. This means that cities and communities should create a data inventory or a data catalogue. This process can be broken into three steps:

DEFINITION

A data inventory is a list of datasets that provides descriptions of the type of data, their source, frequency, unit of measurement, and other useful information. They are closely related to data dictionaries.

FIGURE 1. CREATING A DATA INVENTORY

Source: Modified based on DataSF.

1.2.2 WHY SHOULD CITIES AND COMMUNITIES CREATE DATA INVENTORIES?[4]

A data inventory offers the following benefits:

- They ease the discovery process and make it easy to locate and use data as and when needed.
- They eliminate redundancies and duplication of efforts by identifying what is collected.
- They improve data quality as well as decision making.

TIP

- Looking for a data dictionary template? Refer to this data dictionary template by U.S. Department of Agriculture or refer to this data inventory template by DataSF.
- Refer to pg. 6 of DataSF guidebook and refer to this step by step guidance to create a data inventory depending on the complexity level of your preference (see pg. 4). A one page summary of the data inventory process can be found here. Alternatively, refer to this white paper on how to create a data inventory or you can also refer to guidance for creating data dictionaries from the Government of New Zealand.

3 DataSF. (n.d.). Data Coordinator Guidance. Retrieved April 2021, from https://datasf.org/resources/data-inventory-guidance/
4 Ibid.

1.2.3 WHAT ROLES AND RESPONSIBILITIES SHOULD BE ASSIGNED?[5]

Several key roles help maintain good data practices. These roles represent a set of responsibilities that individuals must address while managing data, particularly while sharing data between teams or with other cities and communities. Individuals in smaller cities and communities may take on the mantle of more than one role but thoughtfully defining how each role is addressed is critical for robust data governance. Have at least one individual with technical and analytical data skills and one with an understanding of the business side of data use.

TIP

Refer to CDO Playbook by Deloitte Insights and Beeck Center.

TABLE 2. ROLE AND RESPONSIBILITIES FOR DATA GOVERNANCE

ROLE	DESCRIPTION OF RESPONSIBILITIES
Chief Data Officer (CDO)	Designated by a municipal executive, the CDO is accountable for the overall implementation and reporting of the data strategy and policies.
Data Steward	Data Stewards are in charge of individual databases, datasets, or information systems. In general, a data steward has business knowledge of the data and can answer questions about the data itself.
Data Custodian	Data Custodians assist with the technical implementation of individual databases, datasets, or information systems. Not all systems or data sources will have a data custodian.
Data Owner	The Data Owner is the ultimate holder of rights to data within the data store. In mixed data environments this may be several individuals. In cases of derived data, a clear chain of ownership should be established for data artifacts, data surrogates, and any synthesized data.
Data Producer	The Data Producers may be individuals, organizations, or even devices where data originates for the datastore.
Data Manager	The Data Manager, sometimes referred to as the "Data Executive", is responsible for ensuring an effective data plan, ensuring the clear identification of the other major roles, and what agreements and specific policies should be in place to manage data effectively.

Source: DataSF Guidebook.

Defining roles and responsibility is not an end itself. It is important to train employees regularly to keep them up to speed with the skills required to carry their responsibilities efficiently. For guidance on what skills are required and how to assess these skills, refer to the data skills catalog and the playbook on assessing data skills included in the Federal Data Strategy.

QUESTIONS TO CONSIDER

- How will the roles be formalized?
- What organizational changes are required to identify and establish these roles?
- How will the employees receive training for their roles?

5 Ibid.

1.2.4 HOW CAN CITIES AND COMMUNITIES CLASSIFY DATA?[6]

Data can be classified as having risk along two broad dimensions: operational and privacy. Operational risks impact business processes, products and services, and may carry a significant liability or cost with breaches or mishandling. Privacy risks include direct or indirect impacts on individuals or other organizations through identity theft, economic loss, humiliation or discrimination.

Data classification helps to understand how the information needs to be saved, used, protected and more importantly, decide who should and should not have access to it. Once a data inventory is created and roles and responsibilities are assigned, use the inventory to classify data as:

- **Public/ open (least sensitive)** – data that can be freely shared with everyone; alternatively, data that has negligible or insignificant impact on the city, community, and any individual if breached.
- **Private** – data that cannot be publicly shared but can be used by governments that have a low impact on the city, community, and any individual if breached.
- **Sensitive/Confidential** – data that is protected by law (for instance, health data), that if breached can lead to significant reputational and economic damage to the city or community, as well as individuals.
- **Highly Sensitive/Highly Confidential** – data that, if breached can severely impact a city or community's ability to perform its statutory functions.

TIP

Refer to D.C.'s data classification here.

TIP: TOP FIVE THINGS CITIES AND CONNECTED COMMUNITIES CAN DO TO GET STARTED:

1. Create a data inventory
2. Assign roles and responsibilities
3. Classify your data.
4. Identify who should and should not have access to the different classes of data.
5. Brainstorm how you can use data to meet your organizational need.

1.2.5 HOW CAN CITIES AND COMMUNITIES DEFINE DATA ACCESS?[7]

Data access is typically described in terms of permissions to access and permissions to work with data. Depending on the classification of data, cities and communities should decide who gets access to which levels of data. Use the principle of least privilege to determine who gets access and how much control they have over the data or system. The principle states that any person or system should have the least access to data that will allow them/it to fully carry out their tasks and responsibilities. Learn more about the principle and review some examples here.

CAUTION

Classifying data can be tricky as some datasets may be deemed open by federal or state laws such as voter registration data. However, there is always a risk that even after removing all elements of personally identifiable information it can be traced back to an individual.

6 US Ignite (n.d.). Model Data Governance Policy. (Internal policy document); DigitalGuardian. (2016). The Definitive Guide to Data Classification [Slides]. Infosec. https://infosecpartners.com/wp-content/uploads/2017/02/The-Definitive-Guide-to-Data-Classification.pdf

7 Ibid.

Data access permissions are typically described in terms of what functional actions users with data access may take. Typically, these fall into four main categories:

- **Read-Only** – Users have access to view but not modify or change any data.
- **Limited Read-Only** – Users have access to read portions of the data from derived views or other restricted system tables.
- **Modify Data** – Users have permission to edit or modify existing data but not create new data.
- **Creation** – Users have permission to create new records or data fields within the datastore.

QUESTIONS TO CONSIDER

- Is the data complete?
- Are there any errors or missing values in the dataset?
- What is its unit and itsfrequency? Is it consistent across data points?
- Are there any inconsistencies across data sources?

1.2.6 HOW CAN CITIES AND COMMUNITIES ENSURE THE QUALITY OF DATA?[8]

Data quality ensures the overall health, utility, and trustworthiness of managed data. It refers to a series of attributes that ensure data can be effectively used. This includes formats, validation, encodings, accuracy, completeness, consistency, and standards that affect how reliably and easily data can be manipulated and used. It can be defined as:

TIP

Refer to this data quality module by preparecenter.org

Data Quality = Completeness of Data x Validity of Data x Timeliness of Data

Regular data quality checks can help organizations understand their data sources and potential sources of error in their data and develop a mitigation strategy accordingly.

1.2.7 WHAT IS DATA RETENTION AND DESTRUCTION?

An important aspect of data governance is to define data retention and destruction policies. Depending on the nature and legal requirements around data, cities and communities should decide how long it is justifiable to hold on to data (refer to the resource repository for resources on data retention guidelines).

Provide a clear guideline and timeline on how and when the data should be destroyed and disposed of, consistent with any applicable federal or state laws at the end of the retention period. In the case of data shared with a third party, it is absolutely critical to validate that the data has actually been deleted as per the terms and conditions of the data sharing agreement. Ask for proof to verify data deletion by the third party.

TIP

Refer to NIST's guideline for Media Sanitization. It provides guidance for four levels of sanitization – refers to removing information such that recovery is not possible – including Simple Disposal, Clearing, Purging, and Destroying.

8 DAMA UK Working Group. (2013, October). The Six Primary Dimensions for Data Quality Assessment. DAMA UK. https://damauk.wildapricot.org/resources/Documents/DAMA%20UK%20DQ%20Dimensions%20White%20Paper2020.pdf; Dasy Center. (n.d.). Data Governance Toolkit: Data Quality. Retrieved April 2021, from https://dasycenter.org/data-governance-toolkit/data-quality/

1.2.8 HOW CAN CITIES AND COMMUNITIES PRESERVE DATA IN CASE OF AN EQUIPMENT FAILURE, NATURAL DISASTER, OR CYBERATTACK?[9]

Floods, fires, viruses, ransomware, and hacks can destroy digital data or render it inaccessible. Periodic data backups of all data and systems should be mandatory. Data backup creates a duplicate copy of the data in a secondary location. As such, data backup is critical for preserving data in an event of equipment failure, natural disaster, or a cyberattack (read the Cybersecurity section of this guide to learn how to protect your data, networks and systems from a cyberattack).

Cities and community leaders should decide what type of backup they want to do, how often they want to do backups and how they want to do backups. The Resilient Organization guide by TechSoup is a three-part guide - Disaster Preparedness, Disaster Recovery, and Staff Preparedness – that provides best practices and steps to prepare and plan for natural disasters and cyberattacks. The three-part guide is written for non-profits but can come in handy for cities and communities that are just getting started. Refer to the guide based on what you need to know:

QUESTIONS TO CONSIDER

- How much data to backup and how often?
- What devices and applications should be used for backup?
- How would you ensure the security of sensitive backup data?
- How will you keep track of the backup records?

TABLE 3. OVERVIEW OF TECHSOUP GUIDES

GUIDE	WHAT CAN YOU EXPECT TO LEARN?
Disaster Preparedness	• Why is disaster preparedness important and how can we assess where you stand and what you need to do next? • How can we create a disaster preparedness plan, principles of technology disaster planning and backup solutions and strategies? • How can we inform and train your staff, and handle employee transition?
Disaster Recovery	• How can we activate our recovery plan and reestablish internal and external communications? • How can we recover data, systems, and equipment?
Staff Preparedness	• What communication tools can be used? • How should we develop and periodically review the disaster management plan? • How can we prepare an emergency supply kit?

Source: Author based on TechSoup's Resilient Organization Guide.

9 TechSoup. (2020). Disaster Planning and Recovery Guide. https://www.techsoup.org/disaster-planning-and-recovery

1.2.9 HOW CAN CITIES AND COMMUNITIES USE DATA EFFICIENTLY?[10]

The first step to thinking about data use is defining the problem that the organization is trying to solve. There are several frameworks and principles (see resource repository) that provide guidance on how organizations can use data to drive maximum social value. Key principles to keep in mind for safe and effective use of data are:

- **Deliver Clear Public Value** – The use of data should deliver clear public value and benefit to municipal management and/or its residents, businesses, and other organizations. It should be used to meet the needs of the constituents and deliver public services efficiently and cost-effectively.
- **Ensure Data are Fit for Purpose** – Cities and communities may end up hoarding data that is not needed. Data stewards and managers should clearly define how the data serves a purpose and solves a problem.
- **Focus on People** – It is not the data but the people that cities and communities should aspire to protect. They should actively consider ways to prevent the misuse of information that can harm their residents.
- **Maintain Transparency** – As cities and communities ramp up the use of data, the key to maintaining trust is by engaging with stakeholders and communicating clearly how the data are being used to make decisions and how it impacts the lives of residents.
- **Forge Strategic Partnerships** – If your city or community is limited in its ability to use data because of lack of data skills, an alternative is to forge partnerships with nearby cities, communities, academic institutions, local non-profits and independent consultants and utilize their expertise. The Commonwealth of Virginia has a partnership with their technical college system where information system and computer science students intern for different state agencies in Virginia and build applications/platforms for the state as part of their coursework.
- **Understand Limitations** – Data can be a powerful tool for cities and communities. It, however, has its own biases and can lead to discriminatory or unbiased outcomes if not assessed properly. Refer to the Equity section of this guide to learn how to operationalize data use to achieve your equity goals.
- **Retain Human Oversight** – Given that biased data or data sources can lead to inequitable or discriminatory outcomes there is a need to avoid overreliance and dependence on data by adding a human in the loop to oversee final decision making.

QUESTIONS TO CONSIDER

- What is the problem that the organization is trying to solve?
- What data are needed to solve the problem at hand?
- How long does the data need to be retained?
- How will it be disposed after use?
- Who should be informed how the data are being used and how will the organization communicate the information?
- How and who to report to if there are errors or inconsistencies in data?

TIP

Refer to Federal governments Data Ethics Framework for guidance on ethical data use.

10 Government of New Zealand. (2018, May). Principles for the safe and effective use of data and analytics. https://www.stats.govt.nz/assets/Uploads/Data-leadership-fact-sheets/Principles-safe-and-effective-data-and-analytics-May-2018.pdf; New South Wales Government. (n.d.). Module 8: Data-driven Culture. Retrieved April 2021, from https://data.nsw.gov.au/data-governance-toolkit-0/module-8-data-driven-cultu

- **Establish Mechanisms for Reporting Errors and Inconsistencies** – Every employee at some point might be playing the role of a data steward. As employees work with city or community data, they should know "how" and "to whom" to report inconsistencies or errors in a dataset, who should have access, and their duty to report any loss or breach.
- **Foster a Data-Driven Culture** – Focus on fostering a work culture that values data as an organizational asset. Provide resources and learning opportunities to all employees. Design performance metrics or incentive structures to reward data-driven values and behavior. A change in work culture may be required to move the needle and introduce fundamental changes in how employees deal with data daily. Failure to do so is likely to result in data silos, data quality degradation, and sub-optimal data utilization.

1.2.10 HOW CAN CITIES AND CONNECTED COMMUNITIES ENSURE DATA OVERSIGHT?

While implementing a data strategy and a data policy are great steps towards establishing data governance, cities and communities need a data oversight board or a data advisory group to oversee data governance and decision making within the organization. This group should constitute a group of multi-disciplinary leaders from across the organization. It can also include external stakeholders to make the group more representative and inclusive. This group should be responsible for, but not limited to, defining success metrics, coming up with a communication plan, evaluating data-driven decision making, and advocating for a data-driven culture.

TIP

The oversight board should have the following representation:

- The CDO or equivalent
- Chief Information Officer or equivalent
- Data stewards and custodians
- Data analysts from different departments
- A member from city clerk's office
- City Manager or an administration officer
- A professor or an independent consultant with expertise in data science
- One or two members of the public with data skills

1.3. KEY CONSIDERATIONS FOR DATA SHARING

1.3.1 WHAT IS DATA SHARING AND INTEGRATION?

Data sharing is the practice of providing partners within and outside the organization access to information or data to facilitate learning and collaboration on shared priorities. It can multiply the value that cities and connected communities can extract from data. Data integration is a form of data sharing which refers to "the joining or merging of data based on common data fields. These data fields can include personal identifiers, such as name, birth date, social security number, or a common encrypted "unique ID" that is used to link or join records at the individual level."[11]

11 Actionable Intelligence for Social Policy. (2020, May). Introduction to Data Sharing & Integration. https://www.aisp.upenn.edu/wp-content/uploads/2020/06/AISP-Intro-.pdf

1.3.2 WHAT ARE THE RISKS AND BENEFITS OF DATA SHARING?[12]

Data sharing and integration has significant benefits as well as risks:

TABLE 4. RISKS AND BENEFITS OF DATA SHARING

BENEFITS	RISKS
Holistic view of data and information.	In the absence of proper safeguards, there is always a high risk of security breach when data are transferred.
Scalable data to make better decisions.	Since the data are shared and used by partners who may not be the original source of data there is a risk of data being misinterpreted especially in the absence of metadata.
It allows for data reuse that can significantly cut cost and save time spent to otherwise collect data to answer questions around implementation and evaluation of public services.	Sharing data can perpetuate and replicate structural racism and may misrepresent economically disadvantaged people who have historically been the target of discrimination.

[6] Source: Modified based on AISP.

1.3.3 WHAT ROLES DO DATA PROFESSIONALS PLAY IN DATA SHARING?

TABLE 5. DATA ROLES IN DATA SHARING AND WITHIN ORGANIZATION

	ROLE IN DATA SHARING	ROLE WITHIN ORGANIZATION
Chief Data Officer or equivalent	Oversees the process and defines the terms for data sharing agreement.	The overall leader for data governance.
Data Owners	Accountable for the quality and security of the data and holds decision-making authority regarding access and use.	Typically, agency leadership with signatory authority.
Data Steward	Responsible for the governance of data, including metadata. Support established processes and policies for access and use.	Typically, the subject matter experts and data analysts that work with data.
Data Custodian	Responsible for the technology used to store and transport data.	Typically, an IT person or team.

Source: Modified based on AISP.

12 Ibid.

1.3.4 WHAT ARE DATA SHARING AGREEMENTS?

When sharing data between individuals or organizations, it is important to have an explicit agreement that outlines acceptable policies for use, privacy, handling, breach or loss reporting, and other concerns to set the appropriate expectations, and ensure proper handling. Refer to the Contracts for Data Collaboration library for a list of data agreement examples by sector.

Elements of a good data sharing agreement address:

- **Limited Data Set Definition** – This details the collections, types, and formats of data to be shared. This not only helps maintain an understanding of what is visible to all parties but serves as an important check on whether data is being inadvertently shared.
- **Safeguards for Handling, Access and Storage** – This includes expected measures for preventing breaches, exposure, and basic access to data. It may include requirements for encryption at rest and during transfer, securing endpoints, restricting access and other important security or integrity procedures.
- **Passthrough Requirements** – This should explicitly require disclosure of any additional agencies or individuals who may obtain access, including subcontractors. It should detail the need for any additional notification procedures and processes (including training) to ensure they abide and agree to all requirements of the data sharing agreement, and explicitly define additional agreements required.
- **Data Risk Assessment** – Assessment of risk of exposure of data, what the impact might be and to whom. See data classification for different data based on their operational and privacy risk levels.
- **Confidentiality and Privacy** – This should detail any requirements to protect individuals or any Personally Identifying Information (PII) within the dataset. This may range from a simple agreement not to share or expose data, to an agreement that requires the modification of data to de-identify individuals and what level they must do it. Refer to the Privacy section of this guide to learn more about privacy.

QUESTIONS TO CONSIDER

- Why do we need to share the data?
- Who are the stakeholders involved and who should be informed about data sharing?
- What type of data are being shared?
- With whom is the data being shared?
- How will it be shared?
- Is it legal to share the data?
- Is the data sharing ethical?

TIP

Refer to this playbook on how to draft a successful Memorandums of Understanding and Data-Sharing Agreements.

1.3.5 WHAT ARE THE DIFFERENT TYPES OF DATA SHARING AGREEMENTS?[13]

TABLE 6. TYPES OF DATA AGREEMENTS

AGREEMENT TYPE	BEST FIT FOR	SPECIFICS
Memorandum of Understanding (MoU)	• Ongoing data transfers with consistent and formalized parameters. • When the basis of a relationship is grant funding or a service contract.	• Roles and responsibilities of involved groups. • Why agreement is required. • Terms and conditions of partnership.
Data Use Agreement (DUA)/ Data Use Licenses (DUL)	Individual data sharing transactions.	• Parameters for data transfer. • Access information. • Intended data use. • Time parameters for data use. • How the requester should destroy the data.
Enterprise Memorandum of Understanding(E-MOU)	Long term agreement signed by multiple parties (for instance government agency to government agency) to facilitate multiple data sharing requests.	• Describe parties involved. • Set up governance boards. • Define codified request procedures. • Responsibilities of data stewards and data requesters.
Data Sharing Agreements (DSA)	Long term data sharing relationships that involve multiple transfers with different parameters.	• Identify the involved parties. • Terms and conditions for the partnership. • They can stand independently or be an addendum to an MOU or E-MOU. • May also define a process for authorizing data requests along with requirements for storing, protecting, and disposing of shared data.
Business Associate Agreement (BAA)	Personal Health Information.	• Each parties' responsibilities.
Statement of Work (SOW)	To provide a detailed overview of the project in all its dimensions.	• Information about vendors and contractors who are bidding to work on the project. • Timeline & Deliverables. • Scope of Work. • Budget.
Non-Disclosure Agreement (NDA)	Binding contract between two or more parties that prevents sensitive information from being shared with any others.	• Parties involved. • Laws governing the NDA. • Information that is being declared as confidential.

Source: Modified based on Skylight's Data Sharing Playbook.

13 Skylight. (n.d.). Data Sharing Playbook. Retrieved April 2021, from https://skylight.digital/work/toolkits/data-sharing-playbook/responding-to-data-requests/

1.3.6 HOW SHOULD CITIES AND COMMUNITIES RESPOND TO DATA REQUESTS?[14]

Cities and connected communities can have data sharing relationships within their organization, with other government organizations, with external companies/vendors or with the public. Which agreement an organization uses depends on the nature of sharing and parties involved. Some agreements can also be used together, for instance Enterprise Memorandum of Understanding (E-MOU), Data Sharing Agreement (DSA) and Data Use Agreement (DUA) are often signed together. The organization must work closely with its legal team to ascertain if a data agreement is needed in the first place. For instance, sharing open access data does not require an agreement. If an agreement is needed, they must decide which agreement is best suited to protect the interest of the organization and the people they serve.

TIP

Consider setting up a data request process to streamline how your organization responds to data requests. Steps to get started:

- Set up a request form/questionnaire.
- Create and publish a data dictionary.
- Data request fee (if applicable)

1.3.7 HOW CAN WE APPROACH DATA DE-IDENTIFICATION/ANONYMIZATION FOR RESPONDING TO DATA SHARING REQUESTS?

De-identification/anonymization refers to the process of removing all personal identifiers from data. Lack of legal frameworks around de-identification of data, inconsistency in understanding and lack of clarity on what constitutes as a personally identifiable information makes de-identification all the more difficult and challenging. The following are the primary steps to approach de-identification/anonymization of data:

- **Identify Who will Lead the Process** – Identify who the request should be directed to and who will work closely with the requester to analyze how the data needs to be de-identified. Typically, the data steward or the data owner of the dataset is in the best position to understand the request and direct it through its due process.
- **Who Else Should be Involved in the Process** – Establish a process that involves permission from all the relevant stakeholders. These stakeholders include, but are not limited to, the CDO, legal team, communications team, data analysts, and the administrative staff handling the data. Approvals from all stakeholders should be documented before initiating the de-identification process.

QUESTIONS TO CONSIDER

- Who works with the requester to understand the de-identification needs of the data requested?
- Who should be involved in the request approval process?
- What's the cost involved in the de-identification process? Who pays for it?
- Is it acceptable to refuse requests on the grounds other than confidentiality of data?
- How will the organization document the provisions of de-identified data to the requester?
- Consider what other documents maybe needed (eg. MOU, DSA, etc.)?

14 Ibid.

- **Grounds for Approval** – Authorize the data stewards or owners to reject a request on the grounds of confidentiality, risks, time and cost, as well as the purposes of the requester.
- **Consider Centralized De-Identification Services** – Create a centralized de-identification service where a team of technical experts can take the lead on all de-identification requests. In the absence of in-house expertise, cities and communities must consider onboarding independent researchers or partnering with data science departments at local academic institutes.

1.3.8 HOW CAN WE PREPARE FOR A SUCCESSFUL DATA REQUEST?[15]

Cities and communities may also find themselves in a position where they have to request data from external parties (government organization or a vendor). That is when having a plan for creating a data request will come in handy. Consider including the following in your plan:

- **Problem Statement** – A description of the problem that you hope to solve with the data. Refer to Bardach's 'Define the Problem' section.[16]
- **Identify the Data** – This is where published data inventories/ dictionaries come in handy. Review the data inventory of the organization and identify the data you need to meet your objective.

TIP

Follow these steps to get organized before you approach a data owner with a data request.

- **Establish credibility** – Outline how the partnership with the organization will help you address the problem at hand. Provide information on how you plan to use, protect, share (internally), and store data.
- **Make the Case for Data Sharing** – Highlight how the data owner can benefit from data sharing. Identify and emphasize if there are any potential synergies from the collaboration.
- **Specify the Parameters of Data** – The more specific you are the easier it will be for others to process your request. Provide information on the specific date range, frequency, unit or specific filters such as age, gender, geography, etc.
- **Provide a Realistic Scope and Timeframe** – Give the data owner a reasonable timeframe to consider and complete your request. Keep in mind the time needed to draft and sign a data sharing agreement.

15 Ibid.
16 Eugene Bardach (2012), A Practical Guide for Policy Analysis – The Eightfold Path to More Effective Problem Solving, Fourth Edition, Sage, Los Angeles.

1.4. UNDERSTANDING OPEN DATA

1.4.1 WHAT IS OPEN DATA AND WHAT ARE ITS GUIDING PRINCIPLES?

Open data refers to data that can be freely used, reused, and distributed to the public. This implies that data should be both legally open – placed in public domain for use with minimal restrictions – as well as technically open – published in an electronic machine readable and non-proprietary format.[17]

TIP

Refer to the list of definitions for Open Data concepts.

There are several (such as the International Open Data Charter and Sunlight Foundation's 10 Principles for Opening Up Government Information) open data principles that form the foundation for open data access and sharing. Broadly speaking, open data are based on the principles of open by default, availability and access, reuse and redistribution, universal participation, comparable and interoperable, and inclusive development and innovation.

1.4.2 HOW CAN CITIES AND COMMUNITIES START TO OPEN UP DATA?[18]

- **Identify Data Coordinators** – The first step to building an open data program is identifying the data champions across different departments who are skilled and, in a position, to make a commitment to open data. Cities and connected communities should designate "Data Coordinators" for each department who will act as the main point of contact and accountability for open data in their department. Data Coordinators will take the lead in developing inventories for the department, establishing timelines for publishing, implementing privacy, data licensing, metadata and other data practices.
- **Start Simple and Small** – There is no harm in starting small and simple. Even publishing just one dataset or subset of a larger dataset is a great start. Cities and communities should focus on small wins of high value and build on them and multiply wins over time.
- **Choose your Dataset(s)** – This is one of the most critical steps in opening up data. Having a pre-existing data inventory can make this step relatively easy. Once the organization has a data inventory they should:
    - Consult with all stakeholders (in-house data professional, data governance board members, and the public) to understand which dataset(s) will create the most social value if published.

17 World Bank. (n.d.). Open Data Essentials. Retrieved April 2021, from http://opendatatoolkit.worldbank.org/en/essentials.html
18 Open Knowledge Foundation. (n.d.). Open Data Handbook. Retrieved April 2021, from https://opendatahandbook.org/guide/en/how-to-open-up-data/

- Undertake a risk-benefit analysis to weigh the benefits of releasing the data against the privacy risks it may pose to the organization or an individual. Open data are likely to conflict with individual privacy as organizations mature and add more data to their open data platforms, and therefore each dataset is assessed to ensure that it has been de-identified before being shared openly. Assessments such as the benefit-risk analysis undertaken by Future of Privacy Forum to evaluate City of Seattle's Open Data Risk Assessment are good examples for cities and communities to follow.
- Other than the risk, cities and communities may also consider data which are the easiest to release.
- Open data has become a cornerstone for the smart city's movement. Cities and communities can always learn from the roll out of datasets of cities that have a well-established open data program (refer to resource repository).

TIP

Refer to this de-identification protocol for open data.

- **Apply an Open License** – A critical step to make data legally open is to specify a license. The Federal Open Data Policy states: "Agencies must apply open licenses, in consultation with the best practices found in Project Open Data, to information as it is collected or created so that if data are made public there are no restrictions on copying, publishing, distributing, transmitting, adapting, or otherwise using the information for non-commercial or for commercial purposes." Examples of open licenses and dictation can be found here.
- **Make the Data Technically Open** – The data should be available in an electronic machine-readable format. Cities and communities can host data on their website or via third party sites.
- **Post the Applicable Open License and Any Appropriate Metadata and Disclaimers** – Appropriately describing the dataset, its content, metadata, applicable licensing, and any legal disclaimers is important. If the data are not real-time, the snapshot time should be included.
- **Make it Available and Discoverable** – Post a data inventory on the data portal that users can browse through or preview without downloading the dataset. Be thoughtful about the user-friendly format(s) in which the data can be downloaded to maximize the utility that users can get from the dataset.

TIP

- Refer to common license types for datasets here.
- Local libraries and academic institutions' digital libraries, arXiv.org, and GitHub are open-access repositories that cities and connected communities can use to host their data. For instance, City of Boston launched its Open Data to Open Knowledge project in collaboration with Boston Public Library.

- **Engage and Get Feedback Early and Often** – Engage with potential and actual users in the early stages of development of the program. Get feedback on dataset and data formats that will be of most value to them. Open data is an iterative process, and their feedback can help organizations make sure that the next round of iterations fill the gaps identified in the first round of datasets released.
- **Periodic reports on Progress in Implementing Data** – Lastly, the Chief Data Officer (CDO) or equivalent should provide periodic reports or updates on the progress and implementation of the open data program. The report should outline datasets opened overtime, plan for future data releases, and feedback and evidence from stakeholders on the usability and relevance of open data.

1.4.3 WHAT IS AN OPEN DATA POLICY AND HOW CAN CITIES AND CONNECTED COMMUNITIES DRAFT AN OPEN DATA POLICY?

An Open Data policy provides information on what data will be made public and how it will be made public. It is a testament to a city's commitment to transparency and innovation. The Open Data Policy Hub by the Sunlight Foundation provides a collection of 109 local and 12 state governments open data policies.

There are many resources available to help cities and communities craft their open data policies. A report by the National League of Cities analyzed open data policies of five cities and provided recommendations on how cities can achieve their goal of open data. The Sunlight Foundation provides detailed guidance on establishing open data and offers a policy generator tool to help cities and communities craft their own open data policy.

1.5. DATA GOVERNANCE RESOURCE REPOSITORY FOR CITIES AND COMMUNITIES

NO.	TITLE/ORGANIZATION	LEVEL	WHAT CAN YOU EXPECT TO LEARN?
		DATA STRATEGY AND FRAMEWORK	
1	The 5 Essential Components of a Data Strategy	Intermediate	Written for for-profit organizations, the report explains why a data strategy is important. It introduces the five key components of a good data strategy.
2	Data Management Strategy 2019-2022	All	A simple easy to read and visually appealing data strategy from the City of Dallas.
3	Oregon's Data Strategy 2021 – 2023	All	A draft of Oregon's data strategy was published for public comment in July 2020. The document captures the guiding principles, outcomes and a roadmap for action to achieve its data goals.
4	Data Ethics Framework (Government of U.K.)	All	An ethics framework for data use developed by the government of the United Kingdom. The framework includes the principles and self-assessment frameworks for actions.
5	Ethics & Algorithm Toolkit	Intermediate – Advanced	A toolkit developed for governments to assess the risk of data-based decision making.
6	The Data Equity Framework	All	This seven-step framework is easy to apply and doesn't require you to reinvest the way your team works. It helps you apply the equity lens by offering simple steps and checklists to apply the tool.
		DATA INVENTORY & RETENTION	
7	Retention Schedules for Local Governments: Five Things You Should Know	All	Read tips on retention schedule based on North Carolina's new General Records Schedule for Local Government Agencies.
8	Guide to the Inventory, Scheduling, and Disposition of Federal Records	All	This guide provides essential information, guidance, and tools necessary for Federal agency records managers to establish, manage, and operate an effective records disposition program within their agencies.
		DATA SHARING	
9	Sharing Data for Social Impact: Guidebook to Establishing Responsible Governance Practices	Beginner – Intermediate	A comprehensive guide that captures the different aspects of data sharing including: (i) defining and understanding the data being shared, understanding ethical implications of data sharing; (ii) defining operations, formalizing best practices, drafting data sharing agreements; and (iii) driving impact, monitoring and assessing privacy and security concerns, and processes to improve data quality.

NO.	TITLE/ORGANIZATION	LEVEL	WHAT CAN YOU EXPECT TO LEARN?
10	Smart Cities Data Sharing Framework	Intermediate – Advanced	The report discusses the role of data sharing in smart cities, business opportunities, trends in industry approaches, introduces a data framework and concludes with recommendations for city planner on next steps for supporting city's data evolution plan.
11	State of Connecticut Data Sharing Playbook	All	A great resource for local government to get started with data sharing. It also provides examples and further recommended readings on the different steps involved in data sharing.
12	California Health and Human Services (CHHS) Data Sharing Framework	All	Refer to the example of process flow, legal agreement, form and instructions provided by CHHS.
13	Using and Sharing Data/ National Association of Counties	All	A great resource for examples, workshops, reports and toolkit for data sharing and use.
DATA DE-IDENTIFICATION			
14	Data De-Identification Guidelines	Advanced	This resource from CHHS provides guidance on de-identification methods, examples, reporting types and legal frameworks.
OPEN DATA			
15	A De-identification Protocol for Open Data	All	The blog published by International Association for Privacy Professionals (IAPP) provides five step guidance with examples to data de-identification for open data.
16	City of Seattle Open Data Policy	All	A good example of open data policy if you are looking for examples to draft your own data policy.
17	District of Columbia (D.C.) Data Policy	All	D.C. has a comprehensive policy with its open data polices baked into its data policies.
18	Open Data	All	A great resource that provides background on open data and a model policy including examples and additional resources.
19	Chief Data Officer's Annual Report (D.C.)	All	A good example of a report by the CDO on the progress of the open data program.
20	City of Seattle Open Data Risk Assessment	All	A risk assessment of Seattle's open data program undertaken by the Future of Privacy Forum (FPF). The report discusses the different privacy risks tied to open data and provides a model for risk-benefit analysis for opening up data.

NO.	TITLE/ORGANIZATION	LEVEL	WHAT CAN YOU EXPECT TO LEARN?
21	City Open Data Policies	All	A great report from the National League of Cities provides an analysis of the open data policies of five cities – Chicago, Austin, Seattle, Boston, and Amsterdam. It concludes with recommendations based on the lessons learned from the implementation of the policy in the five cities.
22	Project Open Data Metadata Schema	Intermediate – Advanced	Provides guidelines and resources for data standards, open data as well as examples of data field and types.
23	OpenDataPhilly	All	The City of Philadelphia's open data program. Refer to the list of datasets offered by the city and the different formats in which it is offered.
24	Open Government	All	See the list of cities and counties with open data programs.
25	Open Data Privacy	All	The report is divided into four parts/chapters: (i) introduces the concepts and practices for doing a risk-benefit analysis of open data; (ii) outlines a lifecycle approach to managing privacy in open data; (iii) emphasizes role of internal control; and (iv) describes the role of public engagement.
26	Standard Open Data Licensing	All	A list of resources by DataSF on licensing open data. It provides recommendations on data licensing, inventory of licenses across cities and states, and a practical guide to licensing open data.