



# Data Governance Primer

**Administration for Community Living  
Office of Performance Evaluation**

**ACL Data Council**



## TABLE OF CONTENTS

|   |    |
|---|----|
| Introduction  | 1  |
| What is the ACL Data Governance Primer?   | 1  |
| Who Should Use It?  | 1  |
| How Should It Be Used?  | 3  |
| How Was the Primer Developed?   | 3  |
| Background  | 4  |
| What is Data Governance?  | 4  |
| What Is the Difference Between Data Governance,<br>Data Management, and Data Stewardship? | 4  |
| Why Does Data Governance Matter?  | 5  |
| Data Governance Framework   | 7  |
| Data Governance Participants and Roles  | 7  |
| Data governance leaders   | 7  |
| Data stewards   | 8  |
| Technical advisors  | 10 |
| Other data stakeholders   | 12 |
| Data Maturity Assessment  | 13 |
| ACL Governance Guidance   | 16 |
| Data principles   | 16 |
| Data quality  | 18 |
| Data lifecycle  | 22 |
| Metadata  | 29 |
| Data access   | 32 |
| Next Steps  | 34 |
| References  | 35 |
| Appendix A. Definitions   | 37 |
| Appendix B. Federal Government Data Maturity Model  | 40 |
| Appendix C. Application of the Federal Government Data Maturity Model to ACL              | 42 |

# Introduction

## What is the ACL Data Governance Primer?

This Administration for Community Living (ACL) Data Governance Primer (hereinafter “Primer”) is a practical tool to guide ACL’s data policies, standards, and practices. It is designed to strengthen ACL’s data management and to help ensure that the data ACL collects through its evaluation, grant reporting, and administrative performance measures are available, usable, consistent, and secure. The Primer includes general information about data governance and its elements as well as specific data examples from ACL (provided in green call-out boxes).

**A practical tool to guide ACL’s data policies, standards, and practices.**

## Who Should Use It?

All ACL staff can benefit from the Primer. Grant Officers are both data curators, receiving important information from grantees and ensuring its accuracy, and consumers, who use the data received to draft funding opportunity announcements. Managers and leadership use data to define and implement policy.

## EXAMPLE

---

Supporting families and family caregivers is at the heart of ACL's mission. Without the daily efforts by family members to provide support and assistance to individuals of any age with a long-term need or disability, the long-term services and supports system would not be able to meet the need. Family caregivers are the bedrock of this system and key partners in the healthcare delivery system. Twenty years ago, ACL began implementing the National Family Caregiver Support Program (NFCSP), the first federal recognition of the need to support family members and ensure they are able to remain in their roles for as long as possible. Data are vital to the success of any program where understanding its impact is necessary in order to grow and improve the program over time. The State Program Report (SPR) is the primary means by which ACL gathers performance data from states on the operation of the NFCSP and other programs and services funded by the Older Americans Act. The programmatic lead for the NFCSP relies on the data provided by the states via the SPR to provide insights into trends in service delivery, potential challenges that states might be experiencing, and a better understanding of the needs and preferences of the caregivers served by the program. Ensuring the SPR is relevant and useful for states is a critical component of good program management and oversight.

## How Should It Be Used?

The Primer should be used as

- a learning tool to become familiar with key federal data standards and to understand ACL's approach to data governance;
- a framework to guide what ACL programs do, by building data governance into program planning and design; and
- a practical tool to help ACL Centers determine where they are and where they want to be in terms of implementing good data policy and procedures.

## How Was the Primer Developed?

The Primer was developed by the Data Governance Standards Working Group of the ACL Data Council. The content is based on current data governance frameworks, models, and guidance (resources) that apply to ACL and its programs. These resources inform what ACL is required to do and what ACL staff must consider when making decisions about data policies, practices, and standards. For a list of the resources and details about how they apply to ACL, see this **annotated bibliography**.

The content of the Primer should not be considered static. Rather, data governance is an ongoing iterative process that must be responsive to current federal guidance as well as ACL's data needs and related infrastructure needs.

# Background

## What is Data Governance?

According to the Office of Management and Budget (OMB, 2020):

**Data governance** is the process of setting and enforcing priorities for managing and using data as a strategic asset. A Data Governance Body with authority and oversight over the management of agency data assets is a key piece of data infrastructure. These bodies are commonly called by such names as Data Governance Boards, Data Councils, or Data Strategy Teams. The Data Governance Body establishes policy, procedures, and roles for developing, overseeing, and coordinating data management policy and helps prioritize data resource allocations to answer agency key questions and meet stakeholder needs.

**Data governance is the strategy that enables stewardship.**

## What Is the Difference Between Data Governance, Data Management, and Data Stewardship?

Whereas **data governance is the strategy of data**—meaning the policies, procedures, and rules that govern the data, **data management is the logistics**

**of data**—meaning the processes used to plan, specify, enable, create, acquire, maintain, use, archive, retrieve, control, and purge data. Data management encompasses the entire **lifecycle** of a **data asset**, from the very initial creation of the data to the final retirement of the data. **Data stewardship** is the actual implementation of the policies, procedures, and rules that govern the data.

Data stewardship can be thought of as a collection of data management methods covering acquisition, storage, aggregation, and deidentification, and procedures for data release and use. The concept of a data steward is intended to convey a fiduciary (or trust) relationship with data that turns on a data manager whose loyalty is to the interests of individuals and entities whose data are stored in and managed by the system (Rosenbaum, 2010).

Data governance is the strategy that enables stewardship.

## Why Does Data Governance Matter?

With the generation of data comes the risk of poor data quality, duplicate data, and data breaches. A coordinated strategy for standardizing, collecting, storing, analyzing, sharing, and using data helps mitigate these risks.

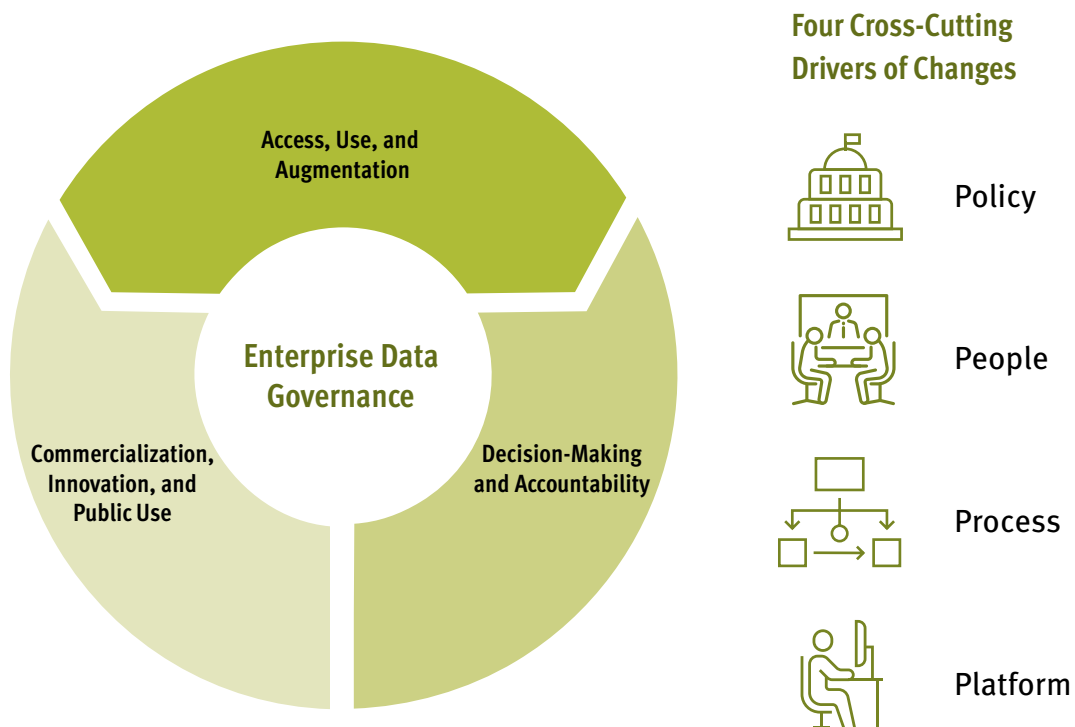
Thus, data governance plays a fundamental role in an enterprise data management approach that realizes the importance of managing information as an asset and turning data into valuable information. According to OMB **Memorandum M-19-18: Federal Data Strategy – A Framework for Consistency**, a data governance structure helps agencies use data to answer important questions while meeting legal and

ethical requirements essential to maintaining public trust, including protecting privacy and ensuring confidentiality (OMB, 2019b).

Instituting a data governance framework provides entities coordinated and systemic oversight for identifying and controlling the collection, storage, and disposition of information resources. Thus, the Federal Data Strategy (OMB, 2019a) encourages agencies to take action focused on data governance:

**Prioritize Data Governance:** Ensure there are sufficient authorities, roles, organizational structures, policies, and resources in place to transparently support the management, maintenance, and use of strategic data assets.

### A Federal Data Strategy for Today's Work Environment





# Data Governance Framework

## Data Governance Participants and Roles

Effective data governance requires the participation of different types of data stakeholders. In addition, the specific concerns and rules that will be addressed will influence the mix of data stakeholders and the amount of effort required of stakeholders to make sound data-related decisions and take appropriate action.

### **DATA GOVERNANCE LEADERS**

#### *What is the role of ACL's data governance leaders?*

ACL's data governance leaders champion data solutions to ensure accountability and increase the value of **data assets** (OMB, 2020). Their role is to promote data governance within ACL, set and enforce priorities for managing and using data as a strategic asset, communicate data-related changes to ACL staff, and monitor progress on data governance tasks and projects. They may engage relevant technical advisors as needed. At ACL the primary responsibility for data governance falls with the Office of Performance and Evaluation (OPE), which focuses on data quality and usability issues, and the Office of Information Resource Management (OIRM), which ensures that ACL's data are accessible and secure. Below are other important roles for ACL staff in maintaining high quality data.

## DATA STEWARDS

### *What is the role of data stewards?*

ACL **data stewards** are the subject matter experts and points of contact for the ACL programs they oversee. Data stewards are responsible for managing each program's data elements, including content and **metadata** (National Aeronautics and Space Administration, 2015). This responsibility includes informing data policies as determined by ACL's data governance rules and data quality standards, as well as defining, managing, controlling, and preserving the integrity of the program data. Thus, the role of a data steward requires both technical and business-oriented skills.

### Key Roles of Data Stewards



## EXAMPLE

---

Program staff in the Office of the Long-Term Care Ombudsman Program serve as the steward for that program's data. This role has involved close collaboration with the team conducting the process and outcomes evaluations, including helping to draft the program logic model, reviewing and commenting on data collection tools and approaches, getting buy-in from program stakeholders to increase their participation in the evaluation, ensuring high response rates, reviewing final reports for clarity, and promoting data dissemination. The role has also involved serving as the task lead for the development of a new web-based data tool for the collection of annual performance data, which includes working with system developers and technical assistance providers, drafting guidance for grantees that collect and report the performance data, and reviewing the data for final acceptance by ACL.

## TECHNICAL ADVISORS

### *What is the role of technical advisors?*

Technical advisors assist data governance leaders in developing all formal documentation and providing guidance for data-related decision-making. Technical advisors should be selected based on their knowledge of the ACL program as well as data governance principles and best practices.

#### EXAMPLE

---

The technical advisor for ACL's Administration on Disabilities (AoD) data management leads several efforts, including those focused on the following:

- how to organize and “wrangle” large data sets in order to get actionable insights from them, which may include finding innovative ways to combine fields of data that do not naturally mesh together and ensuring good data quality management techniques
- how to use algorithms and programming to efficiently go through large data sets and apply treatments, filters, and conditions to the data
- how to conduct exploratory data analyses in order to navigate a data set and come to broad conclusions based on some initial appraisal
- how to create meaningful data compilations and analytics that communicate and relate findings to AoD and ACL programs' mission—more specifically, possibly focusing on

continued

---

- using information about developmental disability populations and demographics so outlier data can be identified
  - using historical data on a program to identify outlier data
  - maintaining trends of key data points
  - helping to complete our understanding of underlying data points needed for performance measures
  - working closely with AoD project officers and managers to maintain data quality guidelines
  - providing contents for data calls and information requests
  - providing subject matter expertise to other ACL projects
- working with ACL's OIRM to address the following:
    - making sure OIRM and system developers understand AoD's reporting tools and the ways data sections, subsections, and data tables are related or interrelated
    - making sure data structure and integrity are maintained and reflected in the data storage portal for an easy and seamless data export and for maintaining the quality of data
    - identifying ways to collect and use data from different sources

## OTHER DATA STAKEHOLDERS

### *What are other types of data stakeholders?*

In addition to data governance leaders, data stewards, and technical advisors, there are other people who are essential to good data governance. Other data stakeholders are individuals or groups who could affect or be affected by ACL's data. Data stakeholders may be grouped into broad categories according to their data role(s), but any given stakeholder may also represent one or more categories.

These categories include:

- **producers** — stakeholders or technologies that create or collect primary data
- **managers** (sometimes referred to as data owners) — stakeholders who receive the primary data, ensure its quality, and compile it into a **data repository**
- **curators** (sometimes referred to as data custodians) — stakeholders who create, manage, and reconcile data repositories and information technology systems
- **analysts** — stakeholders who use data to conduct analyses and to develop data summaries and findings (e.g., reports, journal articles, briefs); may be ACL staff or external stakeholders
- **consumers** — stakeholders who receive the data summaries and findings generated from the data (e.g., program leaders, state agencies, legislators)

## Data Maturity Assessment

### *What is a data maturity assessment?*

A **data maturity assessment** is a high-level assessment of an organization's current data capabilities and supporting processes. It is a tool to help an organization analyze all aspects of its policies, procedures, and operations related to data and data infrastructure, including data governance, data management, data culture, data systems and tools, data analytics, staff skills and capacity, resource capacity, and compliance with law and policy (OMB, 2020).

### *Why is a data maturity assessment important?*

A data maturity assessment allows an organization to evaluate itself against documented best practices, determine gaps, and identify areas to prioritize for improvement (OMB, 2019a).

In addition, it helps with strategic communication between agency data professionals and agency leadership, and it provides a common language and framework to help promulgate solutions and best practices across federal agencies to advance data-driven decision-making. Thus, the Federal Data Strategy (OMB, 2019a) encourages agencies to take action focused on assessing data maturity:

**Assess Maturity:** Evaluate the maturity of all aspects of agency data capabilities to inform priorities for strategic resource investment.

### **A data maturity assessment can help ACL and its Centers**

- assess existing data capabilities and processes,
- prioritize time and resources to improve data and related infrastructure, and
- assess progress over time in improving data infrastructure and operations.



### ***What is an example of a data maturity assessment?***

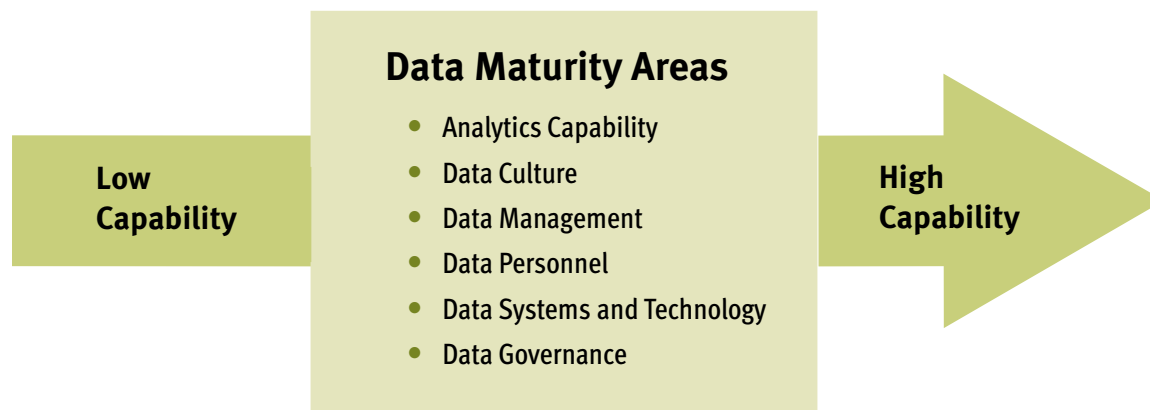
Different data maturity assessment models are available.<sup>1</sup> When selecting a model, data stakeholders should choose one that is roughly aligned with ACL's or their Center's current capacity and that can measure current successes while identifying areas for improvement. It is important that the model balances the information needed and the resources available for the assessment with the types of actions ACL or a Center can realistically take based on the results.

---

<sup>1</sup> The government-wide resources repository required by the Foundations for Evidence-Based Policymaking Act of 2018 (Evidence Act), [resources.data.gov](https://resources.data.gov), includes a collection of references and examples of commonly used data maturity assessment models, including both open-source and proprietary models.



One example of a data maturity model is the **Federal Government Data Maturity Model (see Appendix B)**. This model would help ACL assess the current state of its data capabilities by determining where ACL’s capabilities lie on a continuum from low (=1) to high (=5) in six key data maturity areas (or “lanes”). The Federal Government Data Maturity Model would also help ACL identify future milestones to signify improvement in its data capabilities. This model could be applied at strategic time points to assess progress over time. **Appendix C** provides an example of how the model might be applied to assess the current state of ACL’s data capabilities.



## ACL Governance Guidance

### DATA PRINCIPLES

#### *What are key principles of data governance?*

To ensure successful data governance, there are five core data governance principles to consider (EOL IT Services, 2018; Seiner, 2017; University of New South Wales, n.d.):

- **Valued and strategic enterprise asset:** In any organization, data are a critical asset. They have real, tangible, and measurable value and are the foundation for organizational decision-making. Thus, data must be recognized as a valued enterprise asset across different layers of an organization and across programs. In addition, **data assets** need to be defined, controlled, and accessed in a careful and process-driven way.
- **Data ownership and accountability:** In a successful data governance process, ownership and accountability of data have to be clearly defined. If no one takes ownership of an organization's data governance, it will be directionless, blunt, and ineffective. Ownership and accountability have to be applied across the organization, rather than being centered in just one department. To ensure cross-organization accountability, organizations may establish a data governance council that is composed of representatives from all departments or programs, including leadership. The council should define data procedures and policies to be implemented across the organization and should have the authority to ensure they are strictly followed.

- **Standardized rules and regulations:** A data governance process should follow standardized rules and regulations to protect the data and to ensure they are used in accordance with all relevant external regulations. Rules and guidelines should focus on topics such as data access, data definition, privacy policies, and security standards. The rules and guidelines should be established at the data governance council level and enforced by the data steward(s).
- **Data quality standards:** The quality of data has to be managed consistently from the very beginning. Stakeholders can only make effective decisions if the data are high quality and trustworthy. Thus, enterprise data must be tested on a periodic basis against the defined quality standards.
- **Transparency:** All data governance processes throughout an organization should be implemented with the utmost transparency. The organization should retain permanent records of all relevant data governance actions and processes, including what the data have been used for and why they have been used in that way. Transparency will protect an organization in the event of a data breach and will also allow stakeholders to continually learn from the way data are used throughout the organization.

### ***Why are data principles important?***

Principles provide overarching rules or norms to guide policies and actions. For data governance, principles can help govern the design of new data systems and inform changes to existing data sets and collection efforts.

## DATA QUALITY

### *What do we mean by data quality?*

**Data quality** refers to data being useful, accurate, clear, complete, and unbiased.

“Quality” is defined by OMB as a broad term that encompasses the notions of “utility,” “objectivity,” and “integrity.”

- “Utility” refers to the usefulness of the information to the intended users.
- “Objectivity” refers to whether the disseminated information is being presented in an accurate, clear, complete, and unbiased manner and, as a matter of substance, is accurate, reliable, and unbiased.
- “Integrity” refers to security—the protection of information from unauthorized access or revision, to ensure that the information is not compromised through corruption or falsification (OMB, 2002).

The Transparent Reporting Project<sup>2</sup> adopted eight dimensions of data quality (relevance, accuracy, reliability, timeliness, punctuality, consistency, comparability, and access) which the project considers compatible with the four OMB core concepts (Prell et al., 2019).

---

<sup>2</sup> The Federal Committee on Statistical Methodology established its Working Group on Transparent Quality Reporting in the Integration of Multiple Data Sources in 2017. One of the key projects of the working group was to examine current practices for transparent reporting and identify improvements. For a report on the project, see Prell et al. (2019).

### ***Why does quality matter?***

Quality data can be used to draw accurate conclusions and make good decisions. The better the data quality, the more confidence you can have in your outputs, outcomes, and decisions. In addition, quality data can enhance productivity, resulting in less time spent on correcting errors.

Federal agencies, like ACL, are also governed by regulations and rules focused on ensuring data quality. Per OMB guidelines,<sup>3</sup> federal agencies are directed to develop information resources management procedures for reviewing and substantiating the quality (including the objectivity, utility, and integrity) of information before it is disseminated (OMB, 2002).

**The better the data quality, the more confidence you can have in your outcomes and decisions.**

---

<sup>3</sup> Section 515 of the Treasury and General Government Appropriations Act for FY 2001 (Public Law 106–554) directed the Office of Management and Budget to issue government-wide guidelines that provide policy and procedural guidance to federal agencies for ensuring and maximizing the quality, objectivity, utility, and integrity of information, including statistical information, disseminated by federal agencies.

### What are some key data quality issues?

There are many issues that can lead to poor data quality. Below are a few examples.

| Data Quality Issue    | Description  |
|-----------------------|--|
| <b>Bias</b>           | A statistic is biased if it is calculated in such a way that it is systematically different from the population parameter being estimated. For example, data collection processes could result in lopsided, misleading results. Selection bias is present when researchers are more likely to select one kind of individual within the study population for study than others, and bias can happen when respondents have different characteristics from nonrespondents.                                  |
| <b>Inaccuracy</b>     | Inaccuracy refers to whether the data values are the <i>correct value</i> , whether they correspond to reality. In order for data to be accurate, the data value must be the right value and must be represented in a consistent and unambiguous form; for example, when transferring data between data systems, the transferred data points must be identical.  |
| <b>Incompleteness</b> | Data are complete when all available data elements or data points are included in the data set and there are <i>no missing data</i> . Missing data can occur because of nonresponse or because participants drop out of the study. If there is a high level of missing data, there may not be enough information to make sound decisions. This is especially true if the missing data are systematic, meaning that some groups or types of people are more likely than others to not have provided data. |
| <b>Inconsistency</b>  | Data are consistent when there are <i>no contradictions</i> in data across variables and/or time. Inconsistency in data may include duplicate data or discrepancies in data definitions between data sources during analysis which produce unintended results.   |
| <b>Outliers</b>       | Outliers are data points with characteristics that are <i>considerably different from most of the other data points</i> in the data set. Outliers could indicate “interesting” cases or errors in the data.  |

### *What are ACL examples of strategies to enhance quality of data?*

Some ACL programs have made dedicated efforts to enhance the quality of their data. Below are two examples from ACL.

#### EXAMPLE

---

##### **Burn Model Systems (BMS)**

As part of the BMS, staff at the National Data & Statistical Center developed standard operating procedures (SOPs) to standardize procedures for internal and external use of data from the BMS database. SOPs are available for a range of topics, including data quality and resolving coding issues. To access the SOPs, use this link: <https://burndata.washington.edu/standard-operating-procedures>

##### **Traumatic Brain Injury (TBI) Model Systems**

As part of the TBI Model Systems, staff at the National Data & Statistical Center developed a data dictionary for its TBI Model Systems National Database. It provides information for each variable in the database, including variable group names, group IDs, codes/keys, definitions, notes, examples, history, and Q&A for each variable. To access the data dictionary, use this link: <https://www.tbindsc.org/Syllabus.aspx>

## DATA LIFECYCLE

### *What is a data lifecycle?*

A **data lifecycle** refers to the process or stages data move through, from creation to distribution and use, to archive, to deletion or destruction. It consists of multiple stages, each of which has its own characteristics and builds on the previous stage.

### *Why is a data lifecycle important?*

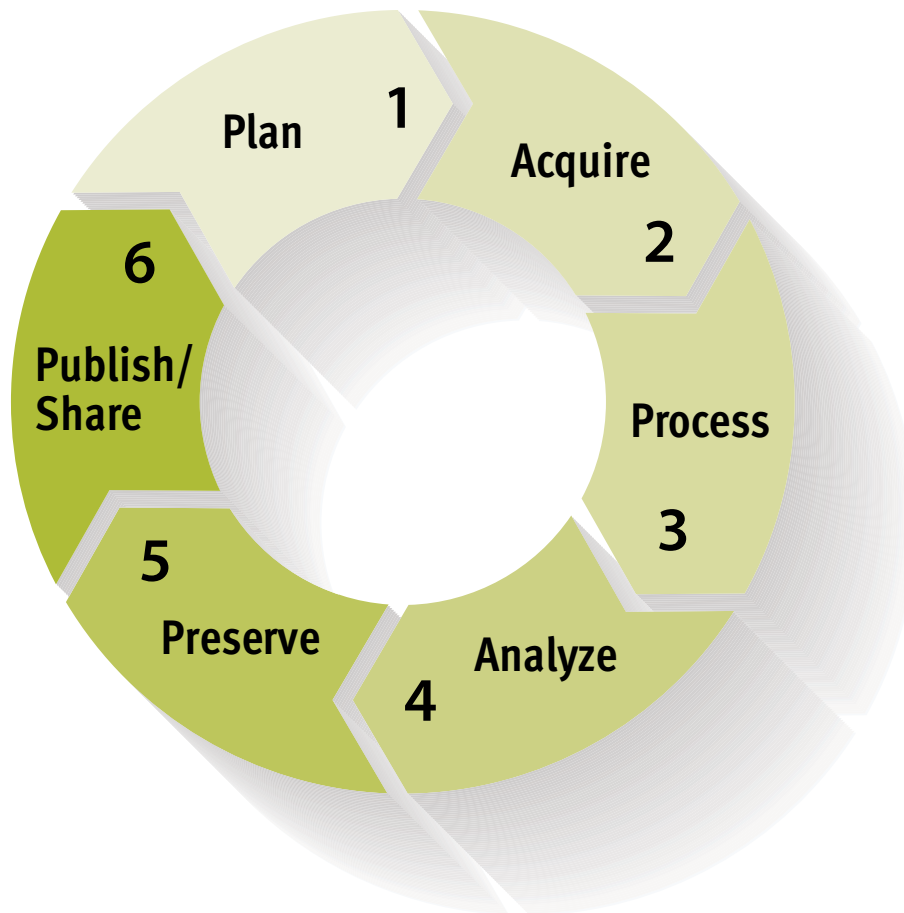
Knowing the stages of a data lifecycle will help you plan and conduct your program activities while, at the same time, supporting efforts to establish evidence for your programs.

A data lifecycle model offers a high-level overview of the individual actions, operations, or processes that must be undertaken at different stages. As a visual tool, it can assist [ACL staff] in anticipating and planning for specific actions that need to be taken at each stage to manage the data, and thus help to ensure timely, comprehensive, and secure approaches to data curation (United States Geological Survey, 2013).

### *What are the stages of a data lifecycle?*

Different data lifecycle models exist, and users may enter at any stage of the lifecycle, depending on their current area of need (United States Geological Survey, 2010). The diagram below presents a model adapted from the United States Geological Survey Science Data Lifecycle Model (United States Geological Survey, 2013).





- 1. Plan:** This should be the first stage when working with new data assets, which helps ensure that all data-related activities are considered. During this stage, staff should determine approaches, resources, and intended outputs for each stage of the data lifecycle. Ideally, a **data management plan** should be developed as an output of the planning process. For existing data assets, planning should focus on reviewing and evaluating outputs and outcomes and identifying actions for improving or updating activities for the subsequent stages. If needed, a data maturity assessment may be conducted at this stage. ACL staff from both OPE and OIRM can help with planning for new or revised data collection activities.

## EXAMPLE

---

Planning for the implementation and rollout of the Title VI module of the Older Americans Act Performance System was a partnership between the Office for American Indian, Alaskan Native, and Native Hawaiian Programs and OPE. The project would not have been successful without thoughtful and expert feedback and mutual support. Together offices worked on creating the data variables (e.g., service categories, service units) that measured what was important to grantees as well as what was required by statute. After securing the platform for collecting the data, they worked with developers to create a user-friendly system, including developing and revising system business and validation rules and regularly conducting user acceptance testing as iterations of the system were completed. The team met weekly, and during the height of planning, multiple times per week, to ensure proper coordination of the project. Later, OPE was even able to lead ACL to design data visualization tools that will help its grantees better understand and advocate for their programs.

- 2. Acquire:** In this stage new or existing data are collected, generated, or considered and evaluated for use. The output for this stage is the data.

#### EXAMPLE

---

Sometimes the data that we need to answer important programmatic questions have already been collected. It is important to know what information may be available and how to access it, as well as how to use it responsibly. For example, as part of the outcomes evaluation of the Older Americans Act Nutrition Services Program, to describe participants' geographic access to food, the research team used participant address information for each respondent in the outcomes survey, data from the Census Bureau, and address data for food retailers from the U.S. Department of Agriculture. The research team geocoded the address information and calculated measures of geographic access to food based on distance from each participant to the nearest store in the area and the number of retailers, by type, within three distances from each participant's residential address. Using existing data reduced evaluation costs and greatly enhanced the resulting findings and recommendations.

- 3. Process:** This stage consists of various activities associated with the preparation of new or previously collected data. Processing of data may entail definition of data elements, integration of disparate data sets, extraction and validation of data to prepare the data for analysis, and opportunities to improve data collection if data are part of ongoing evaluation activities. The outputs from this stage are data sets that are ready for integration and analysis.

#### EXAMPLE

---

Data transformation relating to the AGing, Independence, and Disability (AGID) database involves validation, summary calculations, and quality control of multiple data elements. Source data, generally from surveys, is validated to ensure the format remains consistent with past imports and to identify new or changed variables. The data are then converted and aggregated based on logical groupings (e.g., age category, race or ethnicity, income level) and linked to the appropriate internal database identifiers. Summary statistics, such as percentages and standard errors, are then calculated as needed and made available within the system. After the data have been prepared for the latest year, values are compared against previous years to identify any potential errors in the data or process. Upon review by data specialists, the newly updated data is made available for the public through the AGID website (<https://agid.acl.gov/>).

- 4. Analyze:** Activities in this stage are related to the exploration and interpretation of processed data, where hypotheses are tested, discoveries are made, and conclusions are drawn. Analytical activities include summarization, graphing, statistical analysis, and modeling, and they are used to produce data-driven results and information. At this stage, data sets may be modified (e.g., with construction of new variables) and new data generated (e.g., a data set created by linking two or more separate data sets); versions are tracked; and processes are documented. Data management during analysis improves the efficiency of data analysis activities, preserves documentation that is critical for data integrity, and creates a foundation for evaluation efforts. The outputs from this stage are new or updated data sets or interpretations/findings, which often are published in written reports.
- 5. Preserve:** This stage includes activities associated with storing data for long-term use and accessibility, including **data archiving**, **data disposition**, and the creation of **data repositories**. It is important to plan for long-term preservation of data, metadata, ancillary products, application-neutral storage formats, and any additional documentation, to ensure availability and reuse of data.
- 6. Publish/share:** Activities in this stage are associated with making the data available for use by others through peer-reviewed publications, social media, and ACL's website and data catalogs.

#### EXAMPLE

---

ACL has a published **evaluation policy** that requires ACL to release evaluation results regardless of the findings. Evaluation reports describe the methods used, including strengths and weaknesses, and discuss the generalizability of the findings. Evaluation reports also present comprehensive results, including favorable, unfavorable, and null findings. ACL also makes the majority of its performance data publicly available through its **AGID** data portal.

In addition to these discrete stages, other activities need to be considered and performed across all stages of the lifecycle. These activities include the following:



**Describe (metadata, documentation): Document key information at each stage**

**Manage quality: Plan and implement data quality control processes**

**Backup and secure: Ensure the safeguarding of data**

# METADATA

## *What are metadata?*

**Metadata** are often thought of as data about data, or facts about data. Specifically, they provide structured information that describes and defines data, including their content, quality, and location (National Oceanic and Atmospheric Administration, 2010).

Metadata can range from basic to advanced, from allowing one to discover the mere fact that a certain **data asset** exists and is about a general subject all the way to providing detailed information documenting the structure, [variables], definitions, processing history, quality, relationships, and other properties of a **data set** (Resources.data.gov, 2014).

## *Why are metadata important?*

Metadata provide the data consumer with the information needed to retrieve, process, understand, use, analyze, and manage the described data. Good metadata answer the Who, What, Why, Where, When, and How about data (National Oceanic and Atmospheric Administration, 2010), including:

|   |   |
|---|---|
| <b>Who</b> collected the data?<br><b>Who</b> owns the data?                 | <b>Where</b> were the data collected?<br><b>Where</b> are the data located?       |
| <b>What</b> are the data about?<br><b>What</b> format is used for the data? | <b>When</b> were the data collected?<br><b>When</b> were the data processed?      |
| <b>Why</b> were the data collected?   | <b>How</b> were the data collected?<br><b>How</b> were the data quality assessed? |

### *What should I consider when writing metadata?*

Before writing metadata, consider these fundamental standards of quality metadata:

- 1. Plan early:** Plan metadata development during the initial planning stages of a project. Doing this can save time, effort, and money in the long run.
- 2. Document roles:** Determine and document contact information and roles of individuals or positions, if you do not want to name specific individuals, responsible for different aspects of the data. These individuals may include those who create data, those who process data, and those who provide access to data.
- 3. Establish a protocol:** Establish a format or protocol for data documentation, and then stick to it. Document as much as possible during the data development. It is much easier to capture this information in real time than to document it retroactively.
- 4. Build on existing procedures:** Examine the current procedures established within ACL or your Center for requesting, submitting, certifying, and storing data to use as a foundation for writing quality metadata.
- 5. Follow a style guide:** Develop and apply standard grammar, style, and format conventions, and use them consistently throughout the document. If needed, create a generic style sheet for reference. ACL currently does not have a style guide but will develop one in the future.



- 6. Use consistent terminology:** Use consistent terminology and definitions throughout the document. Define synonyms and abbreviations at the beginning of a paragraph or file if you plan to use them. Provide a rubric for sections that use a large amount of technical terminology.
  
- 7. Review your work:** Review your metadata file for accuracy and completeness. Add any details or context needed to ensure the metadata can be used effectively. Have someone else conduct a second review. Proof the file for typos, grammar, and formatting errors before publishing.

## DATA ACCESS

### *What do we mean by data access?*

Data access refers to a stakeholder's ability to read or retrieve data stored within a database or other **data repository**. Data access can help distinguish the role of data stakeholders. For example, data managers and curators may have the ability to remove, edit, and add data, while general data consumers may only have “read” rights.

### *Why is data access important?*

Data are critical for making informed decisions.

However, data only have value if they can be accessed and used. Emphasizing this need, the Foundations for Evidence-Based Policymaking Act of 2018 (Evidence Act) requires agencies to readily provide access to data for evidence. Specifically, the act states that “the head of an agency shall, to the extent practicable, make any data asset maintained by the agency available, upon request, to any statistical agency or unit for purposes of developing evidence” (Foundations for Evidence-Based Policymaking Act of 2018, 2019).

**Data only have value if they can be accessed and used.**

In addition, the Open, Public, Electronic, and Necessary (OPEN) Government Data Act, which is Title II of the Evidence Act, requires that government data be made available in open, machine-readable formats, while continuing to ensure privacy and security.

By ensuring access to its data, ACL has an opportunity to further the agency's mission by supporting programs that effectively maximize the independence, well-being, and health of older adults, people with disabilities across the lifespan, and their families and caregivers.

### *What are considerations for making data accessible?*

When thinking about data access, consider these fundamental standards

(Burwell et al., 2013):

- Make data available in convenient, modifiable, and open formats that can be retrieved, downloaded, indexed, and searched.
- Ensure formats are machine readable (i.e., data are reasonably structured to allow automated processing).
- Make data available to the widest range of users for the widest range of purposes, often by providing the data in multiple formats for consumption. To the extent permitted by law, these formats should be nonproprietary and publicly available, with minimal restrictions placed on their use.
- Describe data using common core **metadata**, including information about origin, linked data, geographic location, time series continuations, data quality, and other relevant indices that reveal relationships between **data sets** and allow the user to determine the fitness of the data source.
- Ensure systems are scalable and flexible and facilitate extraction of data in multiple formats and for a range of uses as internal and external needs change, including potential uses not considered in the original system design.

- Create a process to engage with stakeholders to solicit help in prioritizing the release of data sets and determining the most usable and appropriate formats for releases.
- Ensure that roles and responsibilities are clearly designated for the promotion of efficient and effective data release practices across the agency, and that proper authorities have been granted to execute related responsibilities.
- Incorporate privacy analyses into each stage of the **data lifecycle**. Specifically, review the information collected or created for valid restrictions to release in order to determine whether it can be made publicly available.

## Next Steps

This document should be considered a work in progress. Later versions will include additional topics, such as ensuring data security, protecting respondent privacy, risks related to linking data sets, and tips for responsible data usage.

# References

- Burwell, S. M., VanRoekel, S., Park, T., & Mancini, D. J. (2013, May 9). *Open data policy: Managing information as an asset (Memorandum M-13-13)*. Office of Management and Budget. <https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2013/m-13-13.pdf>
- EOL IT Services. (2018, January 2). The 5 principles of data governance. <https://www.eolitservices.co.uk/2018/01/02/5-principles-data-governance/>
- Executive Office of the President. (n.d.). *The federal government data maturity model*. <https://my.usgs.gov/confluence/download/attachments/624464994/Federal%20Government%20Data%20Maturity%20Model.pdf?api=v2>
- Foundations for Evidence-Based Policymaking Act of 2018, 5 U.S.C. § 101 et seq. (2019). <https://www.congress.gov/115/plaws/publ435/PLAW-115publ435.pdf>
- National Aeronautics and Space Administration. (2015). *Data strategy white paper*. <https://files.meetup.com/19117935/NASA%2520-%2520Data-Strategy-2015-03-17-White-Paper-Small.pdf>
- National Oceanic and Atmospheric Administration. (2010). *Metadata*. <https://www.ncei.noaa.gov/resources/metadata>
- Office of Management and Budget. (2002, February 22). Guidelines for ensuring and maximizing the quality, objectivity, utility, and integrity of information disseminated by federal agencies. *Federal Register* 67(36), 8452–8460. <https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/assets/OMB/fedreg/reproducible2.pdf>
- Office of Management and Budget. (2019a). *Federal data strategy 2020 action plan*. <https://strategy.data.gov/assets/docs/2020-federal-data-strategy-action-plan.pdf>
- Office of Management and Budget. (2019b). *Federal data strategy: A framework for consistency* Memorandum for the Heads of Executive Departments and Agencies (M-19-18). <https://www.whitehouse.gov/wp-content/uploads/2019/06/M-19-18.pdf>
- Office of Management and Budget. (2020). *Federal data strategy: Data governance playbook*. <https://resources.data.gov/assets/documents/fds-data-governance-playbook.pdf>
- Prell, M., Chapman, C., Adeshiyani, S., Fixler, D., Garin, T., Mirel, L., & Phipps, P. (2019, September). Transparent reporting for integrated data quality: Practices of seven federal statistical agencies (FCSM 19-01). Federal Committee on Statistical Methodology. [https://nces.ed.gov/FCSM/pdf/Transparent\\_Reporting\\_FCSM\\_19.01.pdf](https://nces.ed.gov/FCSM/pdf/Transparent_Reporting_FCSM_19.01.pdf)
- Resources.data.gov. (2014). DCAT-US Schema v1.1 (Project Open Data Metadata Schema). <https://resources.data.gov/resources/dcat-us/>

- Rosenbaum, S. (2010). Data governance and stewardship: Designing entities and advancing data access. *Health Services Research*, 45(5 Pt 2), 1442–1455. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2965885/>
- Seiner, R. S. (2017, March 1). Defining data governance core principles. *The Data Administration Newsletter*. <https://tdan.com/defining-data-governance-core-principles/17087>
- United States Geological Survey. (2010). Data lifecycle models and concepts. [https://my.usgs.gov/confluence/download/attachments/82935852/Data%20Lifecycle%20Models%20and%20Concepts\\_sm.pdf?version=1&modificationDate=1362680814819&api=v2](https://my.usgs.gov/confluence/download/attachments/82935852/Data%20Lifecycle%20Models%20and%20Concepts_sm.pdf?version=1&modificationDate=1362680814819&api=v2)
- United States Geological Survey. (2013). The United States Geological Survey science data lifecycle model. <https://pubs.usgs.gov/of/2013/1265/pdf/of2013-1265.pdf>
- University of New South Wales. (n.d.). Data governance principles. <https://www.datagovernance.unsw.edu.au/data-governance-principles>

# Appendix A. Definitions

|                         |  |
|-------------------------|--|
| <b>Data archiving</b>   | A process that supports long-term storage of scientific data and methods used to read or interpret them.   |
| <b>Data asset</b>       | A collection of data elements or data sets that make sense to group together. Each community of interest identifies the data assets specific to supporting the needs of their respective mission or business functions. Notably, a data asset is a deliberately abstract concept. A given data asset may represent an entire database consisting of multiple distinct entity classes, or it may represent a single entity class. |
| <b>Data disposition</b> | The final stage in the records lifecycle, resulting in destruction of temporary records or the legal and physical transfer of permanent records to the National Archives and Records Administration.   |
| <b>Data governance</b>  | The process of setting and enforcing priorities for managing and using data as a strategic asset. Refers to the strategy of data, meaning the policies, procedures, and rules that govern the data.  |

|                                 |   |
|---------------------------------|---|
| <b>Data lifecycle</b>           | Refers to the process or stages data move through, from creation to distribution and use, until deletion or destruction. It consists of multiples stages, each of which has its own characteristics and builds on the previous stage. |
| <b>Data management</b>          | Refers to the logistics of data, meaning the processes used to plan, specify, enable, create, acquire, maintain, use, archive, retrieve, control, and purge data.   |
| <b>Data management plan</b>     | Describes what data twill be acquired or produced; how the data will be managed, described, and stored; what standards will be used; and how data will be handled and protected.  |
| <b>Data maturity assessment</b> | A high-level assessment of an organization's current data capabilities and supporting processes.  |
| <b>Data quality</b>             | Data that are useful, accurate, clear, complete, and unbiased.  |
| <b>Data repository</b>          | A centralized place to store and maintain data. A repository can consist of one or more databases or files which can be distributed over a network.   |



|                         |   |
|-------------------------|---|
| <b>Data set</b>         | <p>A data set is an organized collection of data. The most basic representation of a data set is data elements presented in tabular form. Each column represents a particular variable, and each row corresponds to a given value of that column's variable. A data set may also present information in a variety of nontabular formats, such as an extensible mark-up language (XML) file, a geospatial data file, or an image file.</p> |
| <b>Data steward</b>     | <p>Subject matter experts and points of contact for the program they oversee. They are responsible for the management of the program data elements, including content and metadata.</p>   |
| <b>Data stewardship</b> | <p>The implementation of the policies, procedures, and rules that govern the data. Data stewardship can be thought of as a collection of data management methods covering acquisition, storage, aggregation, and deidentification, and procedures for data release and use.</p>   |
| <b>Metadata</b>         | <p>Metadata are often thought of as data about data, or facts about data. Specifically, they provide structured information that describe and define data, including their content, quality, and location.</p>  |

## Appendix B. Federal Government Data Maturity Model

Appendix B shows a data maturity model that lists the criteria for data maturity from low capability to high capability. Data maturity is assessed based on the following areas: analytics capability, data culture, data management, data personnel, systems/technology, and data governance. Data with low maturity—or low capability—are characterized by disparate systems and groups, reactive data management at the individual system level, poor data quality, and low decision-making capability. Data with high maturity—or high capability—are characterized by transparency and mission return on investment through executive collaboration and accountability for data quality, government-wide standards, and automation and decision support.

## FEDERAL GOVERNMENT DATA MATURITY MODEL

|   |  |   |  |  |  |
|---|--|---|--|--|--|
| <b>Analytics Capability</b>   | Summary reports  | Descriptive analytics   | Diagnostic analytics   | Predictive analytics   | Cross-functional prescriptive analytics  |
| <b>Data Culture</b>   | Data use is uncoordinated and ad-hoc. Quality issues limit usefulness          | Data use is by request. Quality programs are nascent                              | Some data and analytics are routine and have quality programs supporting key assets  | High demand for data across agency. Drives decision making   | Inter-agency data communities of interest share analyses, best practices                                   |
| <b>Data Management</b>  | Data Managed in silos; documentation sparse; standards not regularly applied   | Data managed in silos; some documentation exists; standards not regularly applied | Data managed across the agency; documentation is uniform; some standards are applied   | Data are managed with cross-functional applications in mind; documentation is uniform; standards regularly applied | Data are managed considering agency-wide needs; documentation is uniform; standards are uniformly applied  |
| <b>Data Personnel</b>   | No dedicated personnel performing data duties                                  | Some siloed data teams; no clear career path for data personnel                   | Professional development path established for data personnel   | Data Professionals integrated with subject matter experts  | Multidisciplinary teams solving agency mission and operational challenges                                  |
| <b>Systems/ Technology</b>  | Data is stored in siloed systems; data are frequently copied to facilitate use | Data are stored in siloed systems; some data can be programmatically accessed     | Some common data systems' some data can be programmatically accessed   | Some common data systems; key data can be programmatically accessed; some common tools exist                       | Core common data systems; key data can be programmatically accessed; common tools are in use across agency |
| <b>Data Governance</b>  | Loose affiliations of technical staff  | Bureau-level collaboration, data ownership and stewardship                        | Agency-level collaboration, data ownership and stewardship   | Agency-level organization accountable for data governance  | Multi-agency advancement of data ownership and stewardship   |
| <b>LOW CAPABILITY</b>   |  |   | <b>HIGH CAPABILITY</b>   |  |  |
| Disparate Systems and Groups, Reactive Data Management at the Individual System Level, Poor Data Quality, Little Decision Making Capability |  |   | Transparency and Mission ROI Through Executive Collaboration and Accountability for Data Quality, Government-wide Standards, Automation and Decision Support |  |  |

**Source:** Executive Office of the President. (n.d.). The federal government data maturity model. <https://my.usgs.gov/confluence/download/attachments/624464994/Federal%20Government%20Data%20Maturity%20Model.pdf?api=v2>

## Appendix C. Application of the Federal Government Data Maturity Model to ACL

Following is an overview of the current state of data capabilities (**on a 1–5 continuum, with 5 being highest level of capability**) of ACL **as an agency** (capabilities may differ for specific programs), as shown for the six key areas listed in the Federal Government Data Maturity model. It also lists milestones that ACL may target to improve its data capabilities in each area.

- **Analytic Capabilities (2):** ACL is able to conduct descriptive analytics to assess “what happened.” Sufficient data exist to create metrics or compare data points from year to year. To enhance its analytic capabilities, ACL may build out its data collection and analytics to identify causes of poor performance or operational problems in real time in order to determine what is happening and why it is happening.
- **Data Culture (3):** Within ACL, some high performing programs/Centers exist that transform quality data into useful information that support decision-making. There is a defined data quality strategy within parts of ACL. Some user feedback is collected and incorporated into programs that produce data-related products and services. To enhance its data culture, ACL may

implement efforts to manage data as a strategic asset across the agency. Customer feedback loops and accountability for data and data quality should be realized across all Centers/programs. Data should be used to drive strategic and operational decision-making agencywide.

- **Data Management (2):** Currently, data are managed in siloes. Although some documentation exists, standards may not be applied regularly and across the entire agency. To enhance its data management, ACL may take steps to ensure that data are managed across the agency, documentation is uniform, and standards are applied consistently.
- **Data Personnel (2):** Dedicated data professionals exist within ACL, but data teams may not be integrated into all operations/Centers/programs. Data personnel may spend most of their time on data maintenance, analysis, and report writing, rather than helping solve operational and mission challenges using complex analytics. To encourage further contributions from its data personnel, ACL may establish professional development training for data professionals and program/regional staff.
- **Systems/Technology (3):** ACL program data can be accessed for decision-making, and ACL is actively working on improving its data systems (e.g., data restructuring project) to eliminate siloes and allow for use of data across programs. To further enhance its systems/technology, ACL may continue to build common data systems and common data tools and encourage their use across the agency.

- **Data Governance (2):** ACL program managers and system owners assume accountability for data sets based on projects. Program leads work with technical staff on incremental policy and data management improvements (e.g., Office of Elder Justice and Adult Protective Services) based on specific business needs and program goals. To enhance its data governance, ACL may focus on agency-level collaboration, such as establishing a review and evaluation process for data governance for continuous improvement.

**For additional information contact:**

Office of Performance and Evaluation  
Center for Policy and Evaluation  
Administration for Community Living  
330 C St SW  
Washington, DC 20201  
[www.acl.gov/contact](http://www.acl.gov/contact)

Administration for Community Living  
Office of Performance Evaluation  
ACL Data Council