U.S. Fish and Wildlife Service

Data Management

Data Management Handbook

May 2021

Prepared for:

U.S. Fish and Wildlife Service

Edited by Jason Duke, Jason Goldberg, and Cameron Paleologopoulos

In collaboration with US Fish and Wildlife Service Data Science Committee

Version History

Version Number	Release Date	Notes
2021.0	17MAY2021	First version

Preface

The purpose of the Data Management Handbook is to provide suggestions for creating and maintaining Fish and Wildlife Service data throughout the entire data lifecycle. This Handbook is the culmination of technical collaboration and partnerships.

Acknowledgements

Many individuals have contributed to the completion of this Handbook. The editors would like to recognize the contributions from other members of the US Fish and Wildlife Service's Data Science Committee who wrote and reviewed this handbook Daniel Adams, Andrew Allstadt, Marene Baker, Jonathan Blythe, Matthew Bobo, Joshua Bradley, Megan Burdi, Thomas Cady, Liz Cruz, Yvonne Detlaff, Aubin Douglas, Cinthia Eichhorn, Kenneth Elsner, Angela Erves, Mark Fritts, Cari-Ann Hayer, Matthew Heller, Teri Jackson-Hicks, Kurt Johnson, Bill Kirchner, Jean Kvasnicka, Toni Linenberger, Aaron Macy, Hilmar Maier, Brandon Maynard, Erica Mize, Erik Osnas, Tamantha Paterson, Rebecca Rau, Jennifer Schmidt, Keon Sheffield, Emily Silverman, Khem So, Michiko Squires, John Sweka, Karen Torres, and Maren Tuttle-Lau.

The Service people above continue to enhance the data management world with their insights.

Suggested Citation

US Fish and Wildlife Service. 2021. *Data Management Handbook*. Duke, J., Goldberg, J., and Paleologopoulos, C. *eds.* Washington DC: US Fish and Wildlife Service. Pp 81. DOI: TBD

Contents

Executive Summary

This Handbook is intended to assist with the implementation of Service Manual Chapter 274 FW 1, Data Management. It provides resources and guidance for employees new to data management practices and standards. While some of the material is detailed for specific users, all employees responsible for data should find useful references for all their data management needs.

The Handbook is organized by the phases of the Data Management Life Cycle. Chapter 1 provides the scope, responsibilities, and data management roles. Chapter 2 provides useful information on metadata (data about data). Chapter 3 reflects quality assurance and quality control that is required in each phase of the data management lifecycle. Chapter 4 discusses the importance of planning and the need for data management plans. Chapter 5 addresses data acquisition including collecting new data and purchasing data from an external entity. Chapter 6 is about maintaining data over time including conversion of data to new formats or software so it can remain readable. Chapter 7 covers access to Service data, limitations on public access management programs. Chapter 8 provides guidance to employees on how to evaluate their data management efforts. Chapter 9 covers archiving and long-term storage of data and the Appendix provides additional references such as Employee Appraisal and Performance Plan language and a glossary.

After its people, data are the Service's most valuable and enduring asset and serves as the foundation for conservation decisions. Maintaining control of the rapidly spiraling amounts of data scattered across Headquarters, Regional, and Field Offices, National Wildlife Refuges, other Service locations, and in public and private clouds requires a new and innovative management approach. Modernizing data management to keep pace with growing application and security demands isn't only important, it's essential. Managing data safely and effectively requires a strategy and reliable methods to access, integrate, organize, govern, store, and prepare data. This Handbook is a tool to help the Service achieve its data management goals. It is intended to serve as a living document that will be updated periodically, ensuring it remains accurate and useful.

Introduction

Purpose and Objectives of the Handbook

The U.S. Fish and Wildlife Service's (Service) mission is, working with others, to conserve, protect and enhance fish, wildlife, and plants and their habitats for the continuing benefit of the American people. Good data are a pillar of that mission. Without good data, we cannot make good decisions. The revised Service Manual chapter, 274 FW 1, Data Management, helps ensure the Service is a good steward of public data, values and manages data as a key strategic asset throughout its lifecycle. The updated chapter reflects recent data management laws, policies, and guidance, outlines roles and responsibilities throughout the organization, and provides policy on:

- Compliance with Federal open data requirements, including requirements for providing external and internal access to data;
- Planning data collection and acquisition activities;
- Maintaining data to keep it secure and preserve its accessibility; and
- Preserving data for long-term use.

This Handbook is intended to assist with the implementation of the new policy. It provides resources and guidance for data management practices and standards. While some of the material is detailed for specific user needs, all employees of the Service are Data Producers, and the Handbook is expected to be a useful reference for all data management needs. The Handbook is intended to serve as a living document and will be periodically updated to ensure it remains accurate and useful.

Data Management Vision

The Service's data are the foundation of conservation success. As Service employees, we will produce data that are rigorous, discoverable, documented, accessible, and secure. Our ability to provide the best scientific information along with recommendations for critical conservation decisions rests on the quality and availability of data. To that end, we will foster a data-driven culture that produces and manages data as a strategic asset. Data management will be a fundamental and mission-critical responsibility of all Service employees, requiring accountability across the organization. We will commit the necessary resources to ensure Service data can be used and retained to make effective and transparent decisions to meet the Service's data-driven mission.

Data and Data Management

Federal laws, regulations, and policies require that we manage all the data we generate through our various mission and business activities. 274 FW 1 provides a list of these authorities, including the Open, Public, Electronic, and Necessary (OPEN) Government Data Act of 2018, which is part of the Foundations for Evidence-Based Policymaking Act of 2018. Also, as detailed in the data management vision statement above, we believe that data are the foundation for the success of the Service's mission. But what does data management entail? And what do we consider to be data?

274 FW 1 defines data as "any recorded information regardless of form or the media on which the data are recorded (FWS, 2020)." It also specifies that data can be processed or unprocessed and represented as text, numbers, or multimedia. Data management, in turn, is defined as the "planning, collection,

storage, maintenance, retrieval, quality control, and analysis of data throughout its lifecycle (FWS 2020)."

It is important to note that the definition of data does not include things like specimens (e.g., herbarium sheets, pollinator collections) or other types of physical objects. While data can be generated *about* these objects, and that data may need to be managed to meet the requirements of 274 FW 1, the specimens themselves would not be considered data under 274 FW 1.

This Handbook provides guidance on how to perform the planning, collection, maintenance, and other management tasks that are necessary to ensure our data meet the requirements of 274 FW 1. We conceptualize data management as occurring throughout the "lifecycle" of the data as a way to demonstrate that good data management begins prior to the data being generated (planning) and ultimately ends once we decide they no longer need to be actively managed and are stored for long-term preservation (archiving). More information about the Service's data management lifecycle model can be found in subsequent sections. The Handbook is generally organized to coincide with that model.

What are the Benefits of Good Data Management?

As noted, several laws and policies identified in 274 FW 1 require the Service to be good stewards of our data. Beyond that, there are numerous benefits associated with practicing good data management. Some of these benefits are:

- When our data are discoverable and accessible to Service employees and members of the public, it helps facilitate its use and reuse. This can allow us to get more out of our initial investment by facilitating collaborative efforts and potential new applications. It also helps reduce duplication of effort and ensures the Service can focus its resources in the most effective way possible.
- More informed and transparent policy and regulatory decisions, as we will have greater confidence in the accuracy and reliability of the data used in decision-making. In addition, our scientific data will be more readily available to provide to the public.
- The Service can operate more strategically and will be able to capitalize on opportunities and manage risks in response to unprecedented changes and events. For example, having strong baseline information in advance of an oil spill can help with response and recovery efforts.
- A reduced Freedom of Information Act (FOIA) burden. By providing the public with appropriate access to Service data, in accordance with relevant open data laws and policies, we may preemptively answer inquiries without needing to route a request through our FOIA processes.
- Provides greater assurance that data collected today will be available and useful in the future through proper maintenance and preservation. Data can continue to be used to support the mission even as employees retire or transition to new jobs.
- Strengthened business practices and organizational efficiency.

You can find real examples of the benefits of good data management by reviewing the Success Stories on the Service's Data Management SharePoint site.

Data Management vs. Project Management

Often, data are the product of "projects." Projects can occur at many levels of the organization, from a project undertaken at a particular Field Office to a large, complex project involving multiple governmental and non-governmental partners. Project management is a distinct discipline from data management, though data management will often be a necessary component of project management.

The diagram below, created by the U.S. Geological Survey (USGS), illustrates the difference between tasks associated with data management and project management.

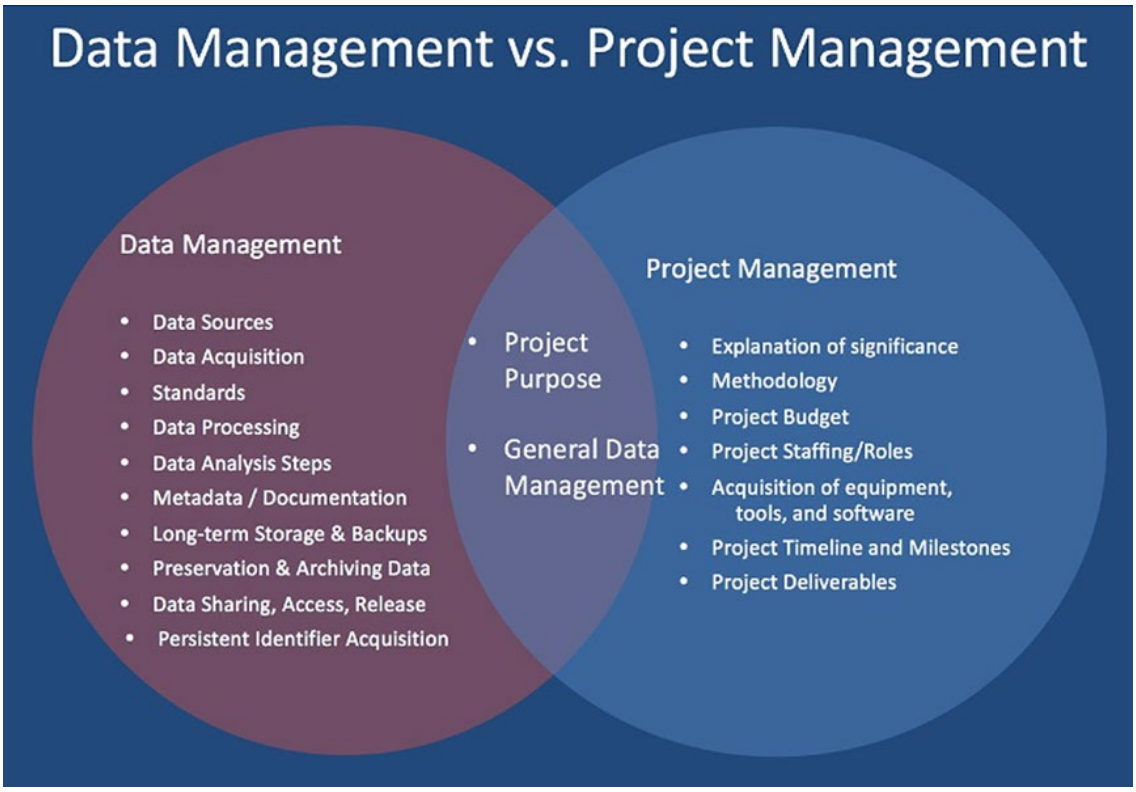

Figure 1-1-Diagram of Data Management vs. Project Management *Note*. The image is from USGS (https://www.usgs.gov/media/images/data-management-vs-project-management-venn-diagram).

Data will not necessarily have a 1:1 relationship with a project, and the relationships between data and projects could be extremely complex. For instance, geological data could be associated with a particular project at a particular site but could also be pooled with other data to support a large-scale, Service-wide project. Regardless of the data's particular relationship to a project or projects, the same data management principles will apply. This also illustrates the need to treat data as an asset that is not just needed to support current projects but could also have potential applications to future projects.

While good project management is an essential task, this Handbook is focused primarily on data management though it may occasionally touch on project management related issues in a general way. This means that there will not be any guidance on developing project methodology, budgeting, staffing/roles and responsibilities, and project deliverables such as reports, presentations, and peer-reviewed papers.

Data Management Lifecycle

The data management lifecycle is a step-by-step approach to data stewardship that covers all aspects of data management from planning to archiving. The lifecycle organizes and illustrates the elements of data management.

Figure 1-2-Diagram of the Data Management Lifecycle *Note*. from U.S. Fish and Wildlife Service (https://www.fws.gov/data/).

The data management lifecycle is not linear, but a series of interconnected and iterative processes. Each of the phases of the lifecycle must be considered while documenting your data set. As you explore each phase of the lifecycle, be mindful of the questions you will need to answer related to documentation, storage, quality assurance, and stewardship. More information on the lifecycle can be found on the Data Management SharePoint site.

The phases of the data management lifecycle are briefly described as follows:

- **Planning** involves making decisions about data management and potential products, including roles and responsibilities. It is important to document all stages of the data management lifecycle and quality control prior to beginning a new project.
- **Acquiring** data involves collecting or adding to data holdings. This could include collecting data through automated collection (e.g., of sensor-derived data) or the manual recording of empirical observations. It may also involve converting legacy data or contracting to acquire data from a partner. Requirements may differ based on how the data was acquired.
- **Maintaining** data includes the steps of processing data for analysis, creating metadata, and making sure data are in a format that is accessible over time. Maintaining data is important to the Service because unmaintained data do not have enough information for future managers to analyze or interpret the information.
- **Accessing** involves the processes and tasks needed to prepare, release, and share our data. Data repositories and peer-reviewed journals are just some of the ways in which data can be accessed, either internally or by the public.
- **Evaluating** data, metadata, and data management processes allows us to ensure that Service data are accessible, discoverable, and of high quality. By establishing evaluation criteria, we can measure progress and success of our data management program.
- **Archiving** data is a process that supports the long-term storage of data and the methods used to read or interpret archived data. We generally archive data that is historically significant in coordination with the National Archives and Records Administration (NARA).
- **Data quality management** involves the prevention of data defects or issues within a dataset that reduce our ability to apply data towards our mission or business-based activities. There are two components of data quality management that represent different concepts, though they are often lumped together. In **Quality Assurance (QA)**, processes prevent data defects from

occurring. In **Quality Control (QC)**, defects are detected and repaired once you have the data. QA and QC are practiced throughout all the other lifecycle steps.

References

US Fish and Wildlife Service. 2020. Data Management. Service Manual Chapter 274 FW 1. Washington DC, 14 pages. 274 FW 1, Data Management, Fish and Wildlife Service Manual (fws.gov)

Chapter 1-Scope/Responsibilities

Sections 1.2, 1.6, and 1.7 of 274 FW 1 establish the policy scope and data management responsibilities. This chapter provides general supplemental guidance on the data we are required to manage, who is required to manage it, and how data management responsibilities are structured.

For additional questions, you can contact the Associate Chief Data Officer (ACDO) or Data Management Governance Board (Board) either directly or through your program representative. The ACDO and Board are available to assist with determining applicability to data you are working with and the establishment of new or update of existing data management systems to comply with policy. For programs that do not have existing data management systems, we recommend beginning by identifying the individuals to carry out the responsibilities of Data Trustee, Steward, Custodian, and Producer. This team can then take the lead on actions to ensure compliance with policy requirements.

Policy Scope

Section 1.2 of 274 FW 1 defines the scope of the Service's data management policy. The policy focuses on data-level management rather than on projects, which may require further clarification.

The policy applies to Service data. *Service data* means data collected, created, acquired, or distributed by Service employees, contractors, and volunteers, including data acquired from an external entity with the primary purpose to acquire it for the Service's direct benefit or use. The policy is not limited to specific types of data, as our goal is to manage all Service data under one data management framework. The policy applies to all types of data we are responsible for managing, including financial, personnel, property, safety management, award, and other administrative records as well as research and scientific data. This means that most, if not all, Service employees have some data management responsibilities under the policy, as will many of our contractors and volunteers who we hire or authorize to collect and manage data for the Service's direct benefit or use. For more information on purchasing data, see Chapter 5, Acquire.

Regarding legacy Service data, the policy does not require us to transform and manage *all* legacy Service data (created, acquired, or distributed before October 1, 2020). It does require us to consider legacy Service data assets and manage any needed to support ongoing priorities and responsibilities per 274 FW 1, as determined necessary or appropriate. In other words, the policy would not require a program to digitize paper records of data from the 1970s, unless transformation of those would support an ongoing or anticipated future Service priority or activity. For such data assets, we should make efforts to bring them into compliance with the policy, which may include digitizing, converting to machine-readable format, and moving to an approved repository. Even if we are not currently using legacy data, we may decide to bring the data into compliance.

Programs and Regions are encouraged to review and determine applicability of 274 FW 1 to legacy Service data in their control. The employees assigned to the Data Trustee role are responsible per 274 FW 1 for making such determinations, which we assume they would do in coordination with officials in their Region or program and using a cost-benefit analysis. Criteria for prioritizing legacy data for transformation to meet the requirements of this policy include if:

- It has significant historical value and/or qualifies for long-term or permanent retention per the Service's records schedule;

- It has current or potential future value for ongoing Service initiatives, or otherwise supports the accomplishment of a Service strategic goal or other priority activity, at any organizational level; and/or
- We have the financial and staff resources to bring it into compliance.

For example, legacy data may have value in understanding climate change trends. The data may not have an immediate short-term value but could be valuable in the long-term. Service programs may want to consider the potential utility for legacy data in applications and needs we do not have insights on in the present time. If you believe that you have legacy data that could provide value to the Service and would like to move forward with performing the tasks necessary to implement the requirements of 274 FW 1, you can also contact the ACDO who can work with the Board and your program or Regional leadership to identify potential resources to assist.

Regarding the scope of Service data, it is important to note that Service data does not mean all data generated by Service grant and cooperative agreement (financial assistance) recipients in performance of their award. Implementation of 274 FW 1 does not authorize or require us to begin collecting data generated under financial assistance awards when we do not otherwise have reason or authority to do so. The primary purpose of financial assistance is to support and assist eligible entities in carrying out a program or project to accomplish a public purpose authorized by Federal statute per the Federal Grants and Cooperative Agreements Act. We must not use a financial assistance award for the primary purpose to acquire property or services for the Service's direct benefit or use, unless otherwise explicitly authorized in statute. For more information on selecting the appropriate funding instrument, see the Service's "Selection of Funding Instrument Decision Diagram." Further, the information we collect from financial assistance applicants and recipients is limited per the Paperwork Reduction Act to the minimum we need to carry out our Federal awarding agency responsibilities, primarily those in Title 2 of the Code of Federal Regulations (C.F.R.) and applicable awarding program statute. Here are some examples of data we may collect:

- We may require recipients to submit specific project performance-related data to support our assessment of recipient and program performance.
- When the Service uses or anticipates using published research or scientific findings produced under a financial assistance award to support rulemaking or other Service action with the force and effect of law, we may require the recipient to provide us a copy of the related data that we then make available to the public. Alternatively, we may require the recipient to publish those in an approved public repository. We also may impose special award terms and conditions to require the recipient to develop and follow a data management plan as part of their approved project plan and/or adhere to the Department's code of conduct when engaging in scientific activities (see 305 DM 3, Integrity of Scientific and Scholarly Activities), as the program finds appropriate. In such cases, the Service's use for the data must not be the primary purpose for the award.

Once received, we should manage the data we have the authority to collect from financial assistance recipients per 274 FW 1. For further questions on applicability of 274 FW 1 to financial assistance, please contact the Financial Assistance Policy and Oversight Division Manager.

Data Management Roles and Responsibilities

This section describes the general roles and responsibilities associated with managing the Service's data. 274 FW 1 itself does not necessarily prescribe any one way of implementing the roles within a program

or Region, as the needs, resources, and structure of the different Service organizations necessitates a flexible approach. This Handbook provides some general guidance on the roles (Data Trustee, Data Steward, Data Custodian, and Data Producer) needed to effectively manage data, the responsibilities associated with each role, and how these roles fit in within the Service's overall data management program structure.

Data Management Governance

Associate Chief Data Officer (ACDO)

The Service's ACDO plays an important role in governing data management within the Service. This includes but is not limited to promoting data management best practices, providing leadership, advice, and technical assistance to programs and Regions, collaborating with others across the Service to develop and implement policies and standards, and leading strategic planning activities related to data management.

However, one of the most important roles the ACDO plays is to serve as a liaison to the Service's scientific, business, and Information Resources and Technology Management (IRTM) communities on data management-related issues. The ACDO ensures that relevant parties within those communities are brought together, either in a formal setting such as the Board or Data Science Committee or informally, to discuss cross-cutting data management issues and identify potential tasks to put before the Board for acceptance. In addition to this internal liaison role, the ACDO works with the Department of the Interior's (Department) Chief Data Officer on issues related to Department-wide data management and notifies the Board of any issues that require Service implementation.

Data Management Governance Board (Board)

The Board, which is chaired by the ACDO, is a chartered body tasked by the Service director with providing leadership and direction for data management. Membership is made up of Regional and HQ employees from a variety of different programs. In general, the Board is tasked with:

- Setting the Service's overall data management strategy, vision, and policies,
- Providing a venue for collaboration and coordination across the Service on data-related issues,
- Providing advice to Regions and programs on proposed data management-related investments and ongoing data management initiatives,
- Monitoring and measuring data management processes to help achieve greater efficiency, and
- Communicating and promoting the importance of data management throughout the Service.

The goal of these efforts is to support and empower the Service's data management community, improve the oversight, management, and usability of Service data assets, and identify the solutions and resources needed to define and prioritize data management needs across the Service.

For more information, please review the Board's charter.

Data Science Committee

Another important function of the Board is to charter technical teams and work groups to help support Board activities and ongoing Service wide initiatives. To do this, the Board has established the Data Science Committee (Committee), which is a technical committee responsible for overseeing the work of a variety of ad-hoc technical teams and working groups. The Committee, and its associated teams and

working groups, assist the Board in a variety of data management areas including architecture, data development, data quality, metadata, file management, and training.

Employees who have a data management-related issue they would like to put before the Board can submit a proposal to the Committee. Once approved by the Board, the Committee will either assign it to an existing team or working group or the Committee will put out a call to the wider data management community to participate on a new team.

Employees who are interested in joining the Data Science Committee (or even just a working group) can contact the ACDO. Committee members serve two-year terms. Supervisor and Board approval is required. For more information about the Data Science Committee, to the Data Management intranet site.

Roles and Responsibilities

Everyone in the Service has an important role to play in data management. 274 FW 1 denotes the specific roles and responsibilities that staff have, both in their respective positions within the agency as well as roles that specifically apply to data management. For a full list of all the roles and responsibilities associated with data management, including up to the Director, review Tables 1-1 and 1-2 in 274 FW 1. This section of the handbook provides a brief look at the roles associated with managing data directly, the Data Trustee, Data Steward, Data Custodian, and Data Producer.

The data management policy requires that someone is assigned to fulfill the responsibilities of each role. Table 1-1 provides a brief description of each role, associated responsibilities, and provides a couple of examples of the types of employees that might serve in each role. It is important to note that each of these roles may be filled by a different employee or the same employee may fill multiple roles depending on the activity in question. This may depend on the amount of data involved, available resources, and other factors. To be clear, there is no requirement to have four different staff members assigned to these data management roles if one person can or should fulfill the roles outlined in Table 1-1.

Table 1-1-Data Management Roles and Responsibilities

Role	Simple Definition	Responsibilities	Examples
Data Trustee	The person responsible for ensuring that adequate resources are available to support data management for data within their scope of responsibility.	1. Ensuring Data Stewards, Custodians, and Producers follow applicable policies and practices. 2. Ensuring data are covered under a data management plan. 3. Providing adequate resources to manage the data throughout its lifecycle.	Assistant or Regional Directors (for certain high-profile data) Managers/Supervisors Project Leaders

Role	Simple Definition	Responsibilities	Examples
Data Steward	The technical point of contact for data and related attributes. Performs certain oversight functions for data and metadata.	1. Serving as the technical point of contact for employees with questions about the data. 2. Ensuring data meets quality, privacy, etc. standards. 3. Ensuring metadata quality. 4. Developing a data management plan in coordination with the Trustee if one doesn't already exist.	Lead biologist for a survey or project (for scientific data) Coordinator at the Regional Office or HQ level
Data Custodian	The person assigned to perform the technical tasks necessary to manage the storage, integrity, and security of data.	1. Maintaining data and metadata in a usable format for storage and accessibility in an appropriate repository. 2. Performing data transfers and migrations as legacy systems are decommissioned. 3. Assisting data stewards with data maintenance activities. 4. Applying and maintaining appropriate access rights to data. 5. Periodically reviewing and updating metadata fields.	Data Managers within an office or program Any Service employee with the necessary technical skills
Data Producer	The person directly involved with creating, acquiring, analyzing, or distributing data.	1. Creating, collecting, and acquiring, and maintaining data in accordance with Service policies. 2. Creating metadata. 3. Following applicable data management plan(s). 4. Complying with data use constraints and limitations.	Field biologists working on collecting data for a survey or project. Potentially any Service employee.

What May Data Management Look Like in an Office?

It is easy to identify the level where someone serves in the Service organization. It can be more challenging to identify your data management function: Are you a **Data Trustee**, **Data Producer**, **Data Steward**, or **Data Custodian**? Any **Service employee** may have more than one of these roles for a given assignment, or you may work as part of a team where everyone has individual roles. One employee, for example, can serve as the **Data Steward** for multiple datasets for colleagues in their field office while also serving as a **Data Producer** for a species survey under a separate data management plan. For some management plans, the same person may fill all four data management rolls. The decision of how to assign data management roles is flexible and subject to the needs of the Region or program.

For example, assume a Field Office conducts an annual monitoring study of several threatened and endangered species on a National Wildlife Refuge. The following offers one example of how data management can occur:

- The **Data Producers** in the office conduct the surveys. Before **Data Producers** conduct the survey, the team reviews the methods of the survey and how to collect data.
- The Field Office **Supervisor** serves as the **Data Trustee** and ensures that staff are assigned to serve as the **Data Steward** and **Custodian.** In this case, the **Data Steward** is the lead biologist for the survey. The **Data Custodian** is a full-time data manager working from the Regional Office who coordinates data management for multiple Field Offices.
- If no data management plan exists to cover this data, the **Data Trustee** will work with the **Data Steward** to develop a new Data Management Plan. Otherwise, they may be able to modify or apply an existing Data Management Plan.
- While planning the survey, the **Data Producer** prepares and records the metadata that describes the data before they conduct their surveys; the **Data Steward** reviews the metadata.
- After the **Data Producer** conducts their survey, the **Data Custodian** makes sure the data are saved in an accessible location. The **Data Steward** reviews the data for any potential problems and will discuss them with the **Data Trustee** if further action is needed.
- The **Data Producer** conducts their analysis of the data and uses the survey results to meet conservation goals.
- If a member of the public requests the data, the **Data Trustee** ensures that the data are provided or verifies that there is a valid reason not to do so.
- As time passes, and employees take new positions, the **Data Custodian** and **Data Trustee** ensure that points of contact are maintained in case the data are needed in the future. The **Data Custodian** also works to ensure that data are saved in an appropriate repository for the data's lifespan.

Chapter 2-Metadata

A vital component of good data management is the creation of effective metadata, or "data about data". As a result, we would like to introduce this topic here before discussing the data lifecycle. It is important to note that Service metadata standards are under active development, based on anticipated Department standards and a Service-wide metadata requirements analysis that will begin in FY 2021. Therefore, this chapter is also under development. The current iteration serves two goals: 1) introduce general metadata concepts; and 2) ensure all active data practitioners are on the same conceptual playing field regarding the importance and use of metadata.

What is Metadata?

As defined in 44 U.S.C. 3502, metadata is "structural or descriptive information about data such as content, format, source, rights, accuracy, provenance, frequency, periodicity, granularity, contact information, publisher or responsible party, method of collection, and other descriptions." In simpler terms, the National Information Standards Organization defines metadata as "data about data". Metadata provide the who, what, when, where, why, and how of the dataset, and are vital to making the Service's data FAIR (see Figure 2-1, Wilkinson et al., 2016).

The 274 FW 1 Data Management Policy requires that all Service data must be described using metadata following an applicable metadata standard (e.g., Federal Geographic Data Committee-compliant metadata for geospatial data). The metadata must be maintained and updated throughout the data lifecycle. Section 1.8B(2) of the DMP policy describes the requirements for metadata in detail.

How does metadata make data FAIR?

- **F**indable – posting metadata to a metadata catalog allows other users to search for datasets.
- **A**ccessible – metadata indicates where a dataset is stored, ideally with a direct repository link.
- **I**nteroperable –metadata provides the field names and standards in the dataset, necessary to integrate with other information.
- **R**eusable-metadata includes detail on how a dataset was generated and info on limitations and use restrictions.

Figure 2-1-Metadata and the FAIR Principles *Note*. The image is from Wilkinson et al. (2016).

In general, metadata can be associated with a project, a file or collection of files, or even objects in a museum. However, for the purposes of this Handbook, we will focus specifically on data-level, rather than project-level metadata. We use the term *dataset* to stand in for a variety of terms, for the purpose of this Handbook.

Different Types of Metadata

We classify metadata into three levels simplified from the National Information Standards Organization structure: discovery, business (operational), and archival. Depending on the use case, the levels are not mutually exclusive, but are inclusive, i.e., discovery may have elements of business, and archival includes all elements of discovery and business. In other words, archival level metadata may be used as discovery metadata.

Discovery-level metadata allows the resource to be uniquely identified, generally assessed for fitness and quality, and accessed. These fields are often used by metadata "catalogs" to index data resources, and associations among records are defined at this level. Discovery metadata may contain simplified elements of business and archival metadata.

Business-level metadata document specific operational needs concerned with administration and management of the resource. This level identifies roles and responsibilities related to the data. Some information in the metadata may have practical use only within the context of the organization, e.g., a URL pointing to a local network server, detailed funding allocations, or keyword themes used to classify data resources according to operational objectives.

Archival metadata allows the data to be consumed and integrated independent of its original creator. Archival metadata may have more extensive requirements like expanded abstracts, data quality information, and supplemental information needed to understand and use the data resource. Lineage and data dictionaries are important elements included at this level.

What Do Metadata Look Like?

Metadata can be represented in a variety of ways. For example, an old-fashioned card catalog is a metadata catalog. Other types of metadata may be directly embedded in an object, like a timestamp in an image file. However, the Service focuses on electronic, machine-readable metadata such as might be stored in an Extensible Markup Language (XML) file or stored on website data catalogs (e.g., data.gov, ScienceBase, ServCat).

The Service's metadata can "in theory" be created and viewed in any text editor but various desktop and web applications can display or provide editing capabilities of machine-readable metadata in user-friendly ways. Figures 2-2 and 2-3 provide examples of what metadata files look like in their most basic forms. Figures 2-4 shows the same metadata in a user-friendly metadata viewer (ArcGIS Pro Metadata editor).

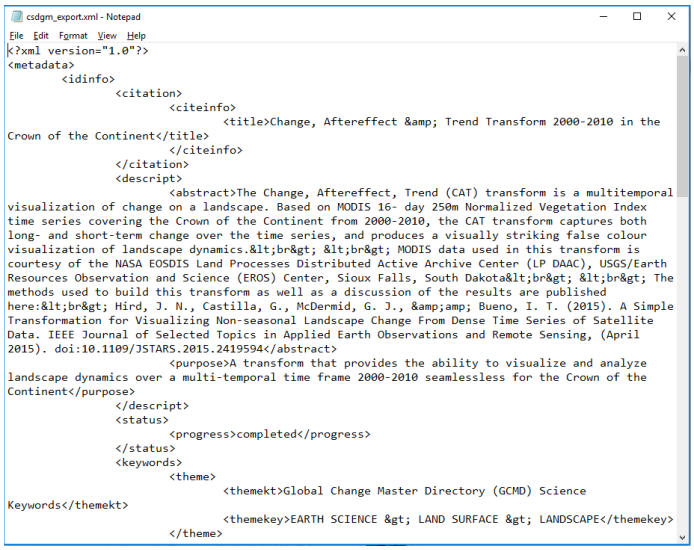

Figure 2-2-XML Metadata File *Note.* the image was created from a screen shot of metadata visualized through Microsoft Notepad from (Hird et al. 2015).

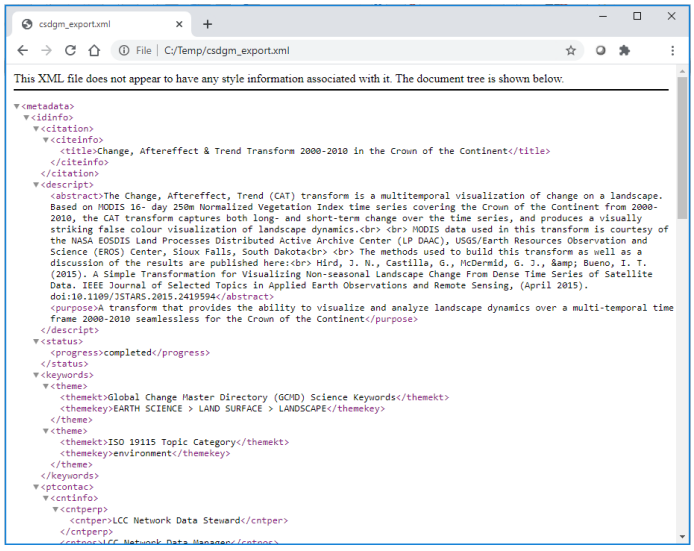

Figure 2-3-XML Metadata File *Note*. the image was created from a screen shot of metadata visualized through Chrome Web Browser (Hird et al. 2015)

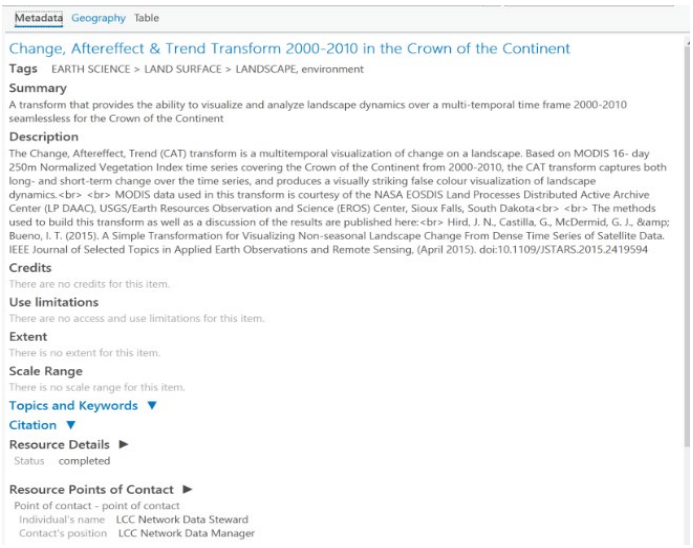

Figure 2-4-ArcGIS Pro Metadata Viewer *Note*. the image was created from a screen shot of metadata visualized through ESRI (Hird et al. 2015)

Though you might not see machine-readable metadata every day, you are using it. Searching any website for information utilizes data about data.

Metadata Standards

A metadata standard is a minimum set of fields that encompass basic attributes required for all metadata, regardless of the type of resource. Standards may specify content necessary for your metadata record to be valid. There are many existing standards and the level of detail required for each may differ. For example, an International Organization for Standardization (ISO) metadata format may only require minimal metadata entries to be valid while the Service may require many fields in a standard. Another example could be where one content standard may simply require a name for a contact entry and another content standard may require features such as the first name, last name, and email.

The Service's metadata standard is currently under development. Once complete, it will be a base requirement for all metadata. This standard will complement and can be built upon by individual Service programs and partner systems (without modifying The Service's one metadata standard), by adding additional metadata requirements. The metadata standard will likely have three parts, depending on the tools we are required to use.

- Department-level requirements – The Department adopts a vocabulary and selection of fields that they will require from all bureaus to facilitate operation of a Department-wide data catalog and data inventory.
- Service-level requirements – We will then lay our own standards on top of Department requirements to meet the specific needs of our work. Programs, and even groups within programs, will build on these requirements for their workflows.
- Storage format: The Department or the Service may require metadata be stored in a certain format, like XML or JSON, as companion files with data and inclusion in data catalogs and repositories.

Well-defined metadata standards can greatly aid validation and translation from one storage format to another, thus feeding a variety of data repositories and catalogs as needed. Well-defined metadata standards also help achieve the goal of only having to write metadata once to fulfill obligations for discovery level, business level, and archival metadata (see descriptions above). Metadata translation tools such as mdTranslator, ArcGIS Pro's "Import Metadata", and ArcGIS Pro's "Export Metadata" can help translate metadata between standards. Schemas may be used to validate structure and syntax, please see the [Review and Validation] section below for more information.

There may be case-by-case programmatic (i.e., Science Applications, Refuges Inventory & Monitoring, IRTM Branch of Geospatial Information Systems) and partner system (e.g., ScienceBase) metadata vocabulary, content, and format standard requirements that the Department and Service standards may or may not address. Similarly, different data types may have required industry standard metadata elements. For example, industry standard geospatial requirements may indicate lineage or horizontal accuracy metadata elements are required, but the general Department or Service standards may not address these elements. Currently, this topic is out of the scope of this metadata section, though you can refer to the appendix for a non-comprehensive list of known standards.

What Needs Metadata?

274 FW 1 requires that **all data** created, acquired, or distributed by or for the Service and managed by or for the Service, in any medium or form, needs metadata, as described in section 1.8B(2). Depending on the activity, this may take the form of a project with overarching metadata that will link likely link to child products including datasets. The specific metadata standards that should be applied to individual data assets should be identified by the Data Trustee/Steward during the Planning Phase in the data management plan (DMP).

Standing alone or under the umbrella of an overarching project, the data that the Service produces have metadata associated with them to enable Service staff or the public to find, understand the data, and more. Here are a few examples of what needs metadata and some of the different types of fields you might need to provide, additional examples are found in Chapter 6, Maintain:

- **Datasets**: what project the data belongs to, why the data was collected, the time frame the data was collected, who collected the data, where it was collected, and how the data were collected, the file format of the dataset, distribution information.
- **Documents/Books/Papers**: Metadata can be used to describe a digital document and include information such as authors, last edit date, keywords, information about its publication, a link to the document, number of pages in the document, file format, distribution information.
- **Images and Videos**: type of camera used to take the image, camera settings used to capture the image, the photographer, date of photo, location of photo/video, title, and description of the image, file size, file format, distribution information.
- **Models/Scripts:** programming language, version/build, technical requirements/dependencies, author, date last modified, inputs, output format.
- **Websites/Webpages:** URL, purpose, date last updated/modified/ web administrator, related projects/products.
- **Digital Maps:** static/interactive

Creating/Maintaining Metadata

Though you *could* sit down and type an XML file in a simple text editor, you would not enjoy the process. Thankfully, a variety of tools can help you create metadata, typically centered on more user-friendly form(s). For example, specialized tools like ServCat, mdEditor (Figure 2-5), ArcGIS Pro (Figure 2-6), or United States Geological Survey's (USGS) Metadata Wizard facilitate creation by organizing your entries and reusing values where possible.

When the Service adopts a comprehensive metadata standard, we will be able to recommend specific tools for metadata creation and validation. In the meantime, you should contact your manager/supervisor or Region or program's data steward (if applicable), if you have any questions in the short-term.

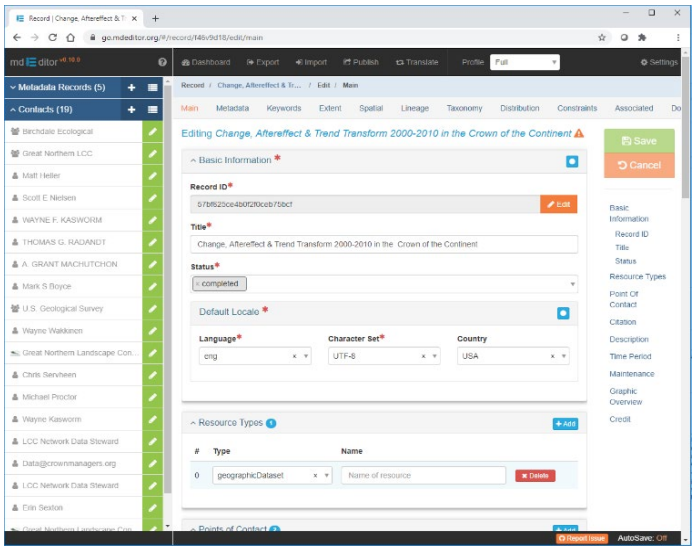

Figure 2-5-Creating Metadata using mdEditor *Note*. the image was created from a screen shot of metadata visualized through mdEditor (Hird et al. 2015)

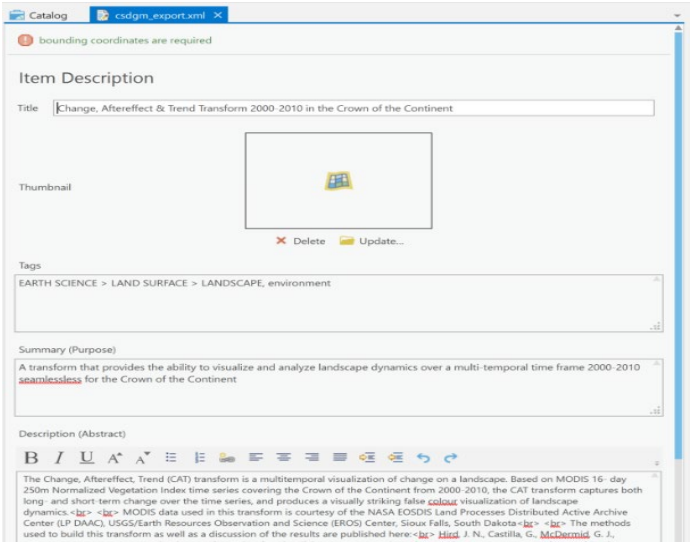

Figure 2-6-ArcGIS Pro Metadata Editor *Note*. the image was created from a screen shot of metadata visualized through ArcGIS Pro (Hird et al. 2015)

Who Creates/Updates

Metadata should be created and updated by those who have direct knowledge of the data (or project), as errors are often introduced when metadata is created or updated after the fact by someone without direct knowledge. Section 1.7 of 274 FW 1 gives Data Producers the responsibility to create metadata and Data Stewards the responsibility to verify the quality of the metadata.

It is also important to note that when we acquire data from outside of the Service using a contract, we should ensure that the data we receive is described by metadata that meets our quality requirements and complies with whatever standard is applicable. In these cases, the entity we have hired to collect data for us should create the metadata, to ensure that we don't introduce any errors by trying to create the metadata later in its lifecycle. More information on the requirements associated acquiring data from outside the Service can be found in Chapter 5, Acquiring Data.

What to Create/Update

You must follow all applicable Departmental, Service, or programmatic standards when creating or updating metadata. You should also ensure that any requirements related to a specific information system, including systems that a Service partner may maintain, are followed to ensure that discovery, business (operational), and archival level metadata is available. There will be more detail in this section in future versions of the Handbook.

When to Create/Update

Metadata should be created early in the data lifecycle, ideally during the Acquire Phase, to ensure that information is not lost or forgotten as a project proceeds and distractions multiply. Metadata should be updated whenever needed, but DMPs should identify a regular time for a thorough metadata review. For example, contact information for the Data Custodian or Data Stewards may change frequently over the life of a dataset.

Where to Store Metadata

Various Department, Service, programmatic, and partner data repositories/catalogs require metadata. Whenever possible, write complete metadata that meets all standards for the repository or catalog that the metadata will feed into. Using metadata editors that allow all required metadata to be entered is key to helping reduce duplicate metadata entry. Once metadata is written, it can most likely be propagated manually or in automated fashion to data repositories/catalogs to be stored alongside the data itself.

Workflow Examples

National Wildlife Refuge System Field Station Example – PRIMR and IMP Workflow

As new National Wildlife Refuge Inventory and Monitoring Plans (IMPs) are developed or updated, metadata about each selected survey are entered into the Planning and Review of I&M activities on Refuges (PRIMR) database. This is an example of a workflow for collecting project-level metadata.

1. Organizational points of contacts control user access to PRIMR for each refuge.
2. Survey metadata is documented in PRIMR by the Refuge Biologist, Refuge Manager, or Principal Investigator for the project, in their role as a Data Producer.
3. The form used to enter metadata ensures that minimum requirements for survey documentation are met. This includes information such as the survey name, survey type, station where survey will be conducted, timing of the survey (time of year and frequency of survey). This form also requires users to document the goals, objectives, survey costs, rationale, and links to survey protocols.
4. Organizational points of contact can use tools within PRIMR to generate reports to identify unpopulated, required metadata fields to perform Quality Assurance (QA) and Quality Control (QC) populated metadata.
5. Refuge staff are responsible for metadata record maintenance and QA/QC. Metadata for each survey record in PRIMR get updated when the project is completed, or if any of the related data for that record need updating.

Science Applications Example

1. Metadata is written by the Data Producer.
2. If mdEditor was not used to write original metadata, metadata is translated using mdTranslator to mdEditor's format (mdJSON).
3. Science Applications metadata requirements are added to the metadata (i.e., related project, point of contact) by the Data Steward using mdEditor. Batch metadata updates are handled via python scripting.
4. The Data Steward validates and QC's the metadata
5. Science Applications data steward publishes project and product metadata to ScienceBase
6. Automated processes post/update project and product metadata from ScienceBase to the Science Applications science catalog
7. Automated processes post/update flagged product metadata from ScienceBase to data.doi.gov, administered by Science Applications data stewards.
8. Automated process post/update flagged product metadata from data.doi.gov to data.gov, administered by SA HQ office.

ServCat Example

ServCat metadata can be entered directly in a web form as you create records and upload data, and records that meet minimum Departmental criteria are shared to data.doi.gov and data.gov as in steps 7 & 8 above. However, you can also import metadata following the Federal Geographic Data Committee (FGDC) standard using the Desktop XML Upload tool, offering another example of entering metadata once and reusing it.

Detailed instructions can be found on the ServCat site, but in general you create FGDC-compliant metadata in ArcGIS Pro, mdEditor, or other available tools, including some Service-specific fields like Cost Center Codes that ServCat will use to organize the records and local paths to files that will be uploaded along with the metadata. When you run the tool, the metadata is imported, the files are uploaded, and the full FGDC-compliant XML file is also stored alongside the files.

Review and Validation

Responsibilities

While final metadata QC is usually done by the Data Steward, anyone writing metadata should ensure it meets defined metadata requirements. Ensuring that requirements are met is often referred to as validation.

How to Validate

Metadata editor software and systems generally have built-in tools to validate against industry standards. Some metadata editing tools (e.g., mdEditor) also have custom validation capabilities that validate against other, pre-defined standards.

Access/Publishing/Distribution

The authoritative metadata record is stored alongside the data per 274 FW 1. However, metadata is often published (copied) to other data catalogs and systems (i.e., ServCat, ScienceBase, data.gov) to help make data more discoverable. Because of this, metadata records must have web links (URLs) that point to the location of the data. Links to the data are often referred to as distribution elements in metadata.

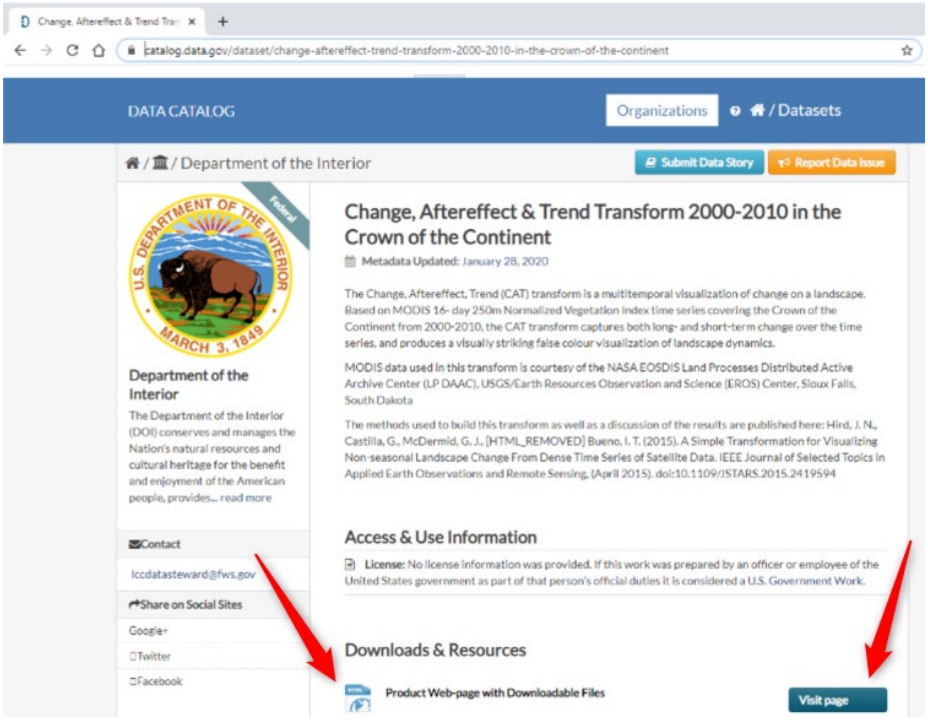

Figure 2-7-Example of Linkable Files in a Metadata Record *Note*. The image was created of a screenshot of data.gov (Hird et al. 2015).

Distributing Metadata

As mentioned at times above, we all want to only have to write metadata once and avoid double entry whenever possible. Implementing a comprehensive standard helps achieve this. There are many data repositories that we are obligated, or want, to distribute metadata to and each of these repositories will have their own guidance on how to distribute metadata. You should refer this guidance for specific instructions. Chapter 7-Access also provides general information on distributing data and metadata.

Limited Use/Restricted Use

See sections “Public Access to Data”, “Limitations on Public Access”, and “Documentation of Restrictions to Public Access” of 274 FW 1 for requirements related to distribution of data that requires limited access. In certain circumstances, even though the data itself may not be available to the public or to a general Service audience, the current guidance is that metadata must be made available and are only rarely withheld. See Chapter 7-Access for more information.

Integrating/Distributing Metadata to Other Systems

Well-defined metadata standards also support the concept of the “semantic web”, which is a “web of data”. Through standardized metadata elements (e.g., dates, titles, biological properties, and any other data one might conceive of) The Service’s machine-readable metadata can become machine-actionable, enabling existing or future data catalogs and repositories to have trusted interactions to talk with each other. These trusted interactions help get the most mileage out of metadata benefiting the employees who create metadata, who will only have to write metadata once, data catalog/repository administrators who provide solutions, and end users who will be able find/discover data/metadata on different data catalogs/repositories. The Service is actively looking at semantic web standards such as

Resource Description Framework (RDF) which provides the foundation for publishing and linking data with the semantic web. Such technologies can allow us to embed metadata in documents or expose data in relational databases or make it available as standalone metadata files.

References

Hird, J. N., Castilla, G., McDermid, G. J., & Bueno, I. T. (2015). A Simple Transformation for Visualizing Non-seasonal Landscape Change from Dense Time Series of Satellite Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, (April 2015). doi:10.1109/JSTARS.2015.2419594

Wilkinson, Mark D et al. “The FAIR Guiding Principles for scientific data management and stewardship.” *Scientific data* vol. 3 160018. 15 Mar. 2016, doi:10.1038/sdata.2016.18

Chapter 3-Quality Assurance and Quality Control

The following chapter describes data quality management and its two associated components: Quality Assurance (QA) and Quality Control (QC). QA and QC are activities conducted throughout the entirety of the data lifecycle and so it is important to understand, document and implement the various practices within each area as part of the planning stage and before beginning work with data. It is important to note that a full discussion of all the requirements related to data quality are outside the scope of this Handbook, as this chapter focuses on quality in the context of data management. You should refer to the Service's Information Quality Guidelines or 212 FW 7, Scientific Integrity and Scholarly Conduct for additional requirements.

Data Quality Management

As mentioned, managing data quality is an important component of the entire data lifecycle. Data quality can be defined in many ways, but herein we share two:

1) The International Organization for Standardization (ISO) defines data quality as: " the degree to which a set of inherent (existing) characteristics fulfills requirements" (ISO 2015a).

2) OMB defines data quality based on the following three characteristics:
   - Objectivity: (a) The information itself must be accurate, reliable, and unbiased, and (b) the manner in which the information is presented must be accurate, clear, complete and unbiased.
   - Utility: The information must be useful for the intended users.
   - Integrity: The information may not be compromised through corruption or falsification, either by accident, or by unauthorized access or revision (OMB 2013).

Data quality management has two components that should be identified during the Planning Phase by the Data Trustee and/or Data Steward as they construct a data management plan (DMP):

QA: QA begins before the data are collected and are processes and procedures used to prevent errors while collecting or entering the data.

QC: QC is the discovery and correction of errors in the data and generally occurs during or after data collection (e.g., detection of outliers, typographical error, a character datum where a numeric value is expected, using an incorrect species code). Quality control should occur as soon as possible after collecting the data and before submitting, archiving, or sharing data.

You can see a comparison of QA and QC characteristics in Table 3-1.

Why is QA/QC Important?

As stated in EPA (2019) "collectively, an organization's policies, processes, and procedures for implementing quality management activities are known as a Quality Management System (QMS). As described by the American Society for Quality, a QMS is the 'blueprint' or framework by which an organization applies sufficient QA/QC practices to produce results that meet or exceed the organization's objectives and expectations (ASQ 2014). The QMS encompasses "both management and technical activities pertaining to planning, implementing and assessing environmental programs within the organization's mission and scope" see Appendix below for more information).

Table 3-1-Quality Assurance and Quality Control Feature Comparison.

Quality Assurance	Quality Control
• Process oriented • Focuses on preventing deviations from project objectives and requirements (proactive) • Involves identifying measurable quality requirements needed to support project objectives and specifying the measurements used to verify that these requirements have been met • Involves developing plans, processes, and procedures before implementing data collection and management activities • Involves assessing processes to detect deviations and areas for improvement, and identifying corrective actions to address findings	• Product oriented • Focuses on identifying deviations from project objectives and requirements (reactive) • Involves performing QC checks during monitoring and data management activities, and assessing results to determine if plans, processes, and results meet requirements • Involves collecting data using modified or corrected processes to verify that changes are yielding the desired results

Note: From "U.S. Environmental Protection Agency. 2019. *Application of Quality Assurance and Quality Control Principles to Ecological Restoration Project Monitoring*. Publication No. EPA/905/K-19/001. Chicago, IL: Great Lakes National Program Office."

The QA/QC practices, procedures, information, and concepts associated with a good QMS can provide the following benefits to practitioners and stakeholders:

- Save time and resources by enhancing the consistency of documentation and procedures in current and future projects.
- Improve data quality, aid in evaluating project success, and incorporate long-term effectiveness monitoring as feedback to adaptive management.
- Encourage a common approach to QA/QC across multiple entities to improve data comparability over time.
- Serve as a consolidated collection of the best QA/QC practices for projects across agencies (EPA 2019).

QA/QC and the Data Lifecycle

Plan

Planning for a data-related project involves making decisions about data management, potential products, as well as the various roles and responsibilities. It is important to document all stages of the data management lifecycle and QA/QC activities prior to beginning a data collection. You should document all the following points listed below within a DMP or elsewhere as appropriate:

- **Project and/or sampling objectives**

  The level of rigor for your objectives, statistical needs and investment in QA/QC strategies should match the data you need for your project. For example, if you are doing a quick assessment vs. a high-profile, nationally funded data collection, your level of QA/QC and statistics would be different. However, it is required that there is some basic level of QA/QC to ensure data quality (validation of data collection activities; EPA 2019).

- **Data collection methods**

  Have all of your data collection methods been identified prior to starting your project (e.g., paper data sheets, digital data collection)? You may need to train Data Producers on data collection methods in order to collect proper observations or measurements. For each of these observations or measurements, identify the **Data Quality Indicators (DQIs)** (i.e., precision, bias, accuracy, representativeness, comparability, completeness, and detectability) that will be used to evaluate results. For each planned observation or measurement, state the **Acceptance Criteria** or associated information with each DQI and make sure it is incorporated in the Metadata content. These acceptance criteria should be stringent enough to control measurement error while also being achievable by properly trained staff using well-defined procedures (EPA 2019).

  It is also important to have procedures listed for calibration of any equipment that will be used to collect data (e.g., D.O./pH/temperature probes) and what will be done if data collected from the instrument does not meet the identified standards. Also, identify how soon after data are collected will the data be transcribed from paper data sheets into the database or uploaded from digital data collection to the database (EPA 2019).

- **Standard Operating Procedures (SOPs)**

  In practice, SOPs may be identified as manuals, methods, protocols, work instructions or other names. These documents play a critical role in maximizing the quality of data by: promoting efficiency; minimizing miscommunication; helping ensure procedures are performed consistently by multiple individuals and over long periods of time (both within and across projects); providing a basis for training staff; serving as references if confusion arises in the field or during data transfer, reduction, review, extraction or analysis; and serving as references to data users. Refer to Table 3-2 for recommended SOP content (EPA 2019).

- **Review Procedures**

  It is important to refer to the acceptance criteria in the metadata content, identify when data review activities will be performed and who is responsible for reviewing the data during the Planning Phase. Include procedures on how data review findings will be conveyed to data users. Describe the procedures to verify, validate and certify your data. How will this be accomplished, who will be responsible for this step? Is there a way to automate parts of the review process? Generally, review and validation of data are performed by the Data Steward, but the actual structure may vary (EPA 2019).

- **Describe how data quality will be evaluated**

  Provide plans for any preliminary reports that will be made with the data. Provide plans to identify and examine the data for any outliers. Describe steps on how this will be done. Provide any initial data assessments, either through reporting or programming language-based approach (e.g., Python, R, Java; EPA 2019).

- **Reporting/documentation**

  Describe any reporting and how often the reporting will be done. Identify where you will store the reports. Indicate any documentation that will be provided and how often reporting is done. Also provide any information regarding how reporting will be disseminated and if the results are intended to be released to the public (EPA 2019).

Table 3-2- Recommended SOP Contents – *If Program Specific Guidance is Unavailable*

UPFRONT INFORMATION	CONTENT	APPENDICIES AND ATTACHMENTS
• Title • Identification number (e.g., SOP FS-34) • Version number, revision date & history Name of organization • Table of contents • Acronyms, abbreviations, and explanations or definitions of uncommon terms/ phrases	• Schedule and timing of data collection • Health and safety warnings • Cautions and interferences • Required equipment and supplies • Personnel qualifications and responsibilities/Detailed procedures • QA/QC, including: Type and frequency/timing of QC checks; Acceptance criteria; Corrective actions • Data and records management (including data entry; reporting units; data uploading, handling, and storage) • Pertinent references • Charts and maps • Graphic representations (e.g., diagrams, illustrations)	• Data sheets, forms, and checklists • Reference tables/materials

Note: From "U.S. Environmental Protection Agency. 2019. *Application of Quality Assurance and Quality Control Principles to Ecological Restoration Project Monitoring*. Publication No. EPA/905/K-19/001. Chicago, IL: Great Lakes National Program Office."

Acquire

Data can be acquired by collecting new data, processing legacy data, collaborating with partners, and contracting others to collect data. Data must be reviewed and updated on a regular schedule to maintain a high standard of quality.

Collecting/Updating New Data

Before teams begin the actual collection or updating of data, they should consider the following QA strategies, including but not limited to:

- Verifying that Data Producers have received training and demonstrated competency in the activities they will perform before they can gather or update data without supervision.
- Making arrangements for analysis of collected data by laboratories that have demonstrated competency in performing the required determinations.
- Preparation activities include but are not limited to conducting site reconnaissance and obtaining any required permits; organizing and preparing equipment and information needed; and supplementing SOPs with site-specific information that will guide crews in their daily efforts at each project site.
- See Table 3 for good field practices for data collection to assure data quality.

Good Field Practices during routine data collection include:

- Adhere to all safety requirements.
- Understand the purpose of data collection.
- Ensure that site packets and equipment are readily available and in good condition.
- Establish effective communication methods between field crews and the project office or designated project leader to ensure that any questions/concerns or support requests can be readily addressed.
- Use teams of at least two field crew members to collect data (if practical).
- Observe and document any conditions that could impact the data collected.
- Calibrate equipment onsite when practical and ensure proper functioning prior to use.
- Ensure site is marked correctly and markings are stable.
- Have a copy of the SOP in the field and follow SOPs as written. Note any deviations to those procedures that were necessary based on site conditions.
- Prioritize collection of measurements or observations on target variables considered least stable (i.e., transitory) during the presence of the crew, or least resilient to crew disturbance.
- Use verbal repetition to confirm data during data transcription.
- Ensure data collection completeness and legibility prior to leaving the site by completing the following activities: (1) double check all collected data, notes and comments, and (2) ensure all data forms, photographs, recordings, charts, samples and/or specimens are labeled, organized and stored correctly, and (3) ensure samples requiring laboratory analysis are handled in accordance with specified shipping and chain of custody procedures (EPA 2019).

Evaluating Existing (Legacy) Data

Evaluate each data source relative to the quality needs (e.g., acceptance criteria, DQI's) of your data project. The rigor should be determined by the overall role and importance of the existing data and your intended use. A flow chart was created by EPA (2002) to help evaluate existing data needs:

- Determine your data needs.
- Identify existing data sources that might meet your project needs.
- Evaluate existing data relative to your project's data quality specifications.
- Document quality issues on planning documents or in the final report.

Collaborating with Partners or Contracting to Collect Data

Ensure some documented agreement is in place about how partners or contractors will follow established data collection procedures and/or guidance documents. This agreement could be in the form of a Statement of Work or a Service Level Agreement that details how data will be collected and expected deliverables prior to project initiation. It might also be useful to have an established QA Project Plan (QAPP) to outline data quality metrics to ensure data quality is being met.

Quality Control Practices (Acquire)

- QC steps should be outlined in the Planning Phase of the data lifecycle. Once data are collected, evaluate the data for any notes or issues that are found right away. It is good practice to perform QC checks on data right after it is collected or acquired, as any anomalies or issues are fresh in your mind. Check any notes or identified issues when transcribing the data into data repositories (such as databases or spreadsheets). These steps can also help verify data integrity issues. When transcribing, detail the steps or procedures to check the data and ensure there are no transcription errors. If errors are found, document the reasoning for leaving the data point in or removing it from analysis. If joining digital data collection, describe the procedures in place to QC data when uploading.
- See Table A1-4 in Appendix material for a comprehensive list on best practices to QC data.

Maintain

Data maintenance includes processing data for analysis, creating metadata, and making sure data are in a format that can be accessed by others.

- ***Data Review*** is a process, undertaken by Data Stewards with the assistance of other employees as necessary, to confirm data quality, identify any associated limitations, and help managers understand how confident they should be in these data if they are used to support decision-making activities. Components of a data review include:
  - Verification: *Confirmation, through provision of objective evidence that specified requirements have been fulfilled* (ISO 2015b, ASQ 2014). This can include verification that specified procedures were followed, results comply with data quality requirements (e.g., established acceptance criteria for data quality indicators were achieved), data entry and calculations were performed correctly, and data integrity has been protected.
  - Validation: *Confirmation, through provision of objective evidence that the requirements for a specific intended use or application have been fulfilled* (ISO 2015b, ASQ 2014). This includes

ensuring that the data are valid, and that they support broader project and sampling objectives.

- Certification: Ensuring a secure validated database has been completed, documented, and certified (if applicable) and that the data within the database are suitable for final usability assessments in preparation for analysis, reporting, distribution, and archiving (EPA 2019).

Stages of Data Review: There are four potential stages of data review, but these are not absolutes: (1) before results leave the field, lab or office (if collected by the Service); (2) upon receipt of reported data; (3) before upload to databases or repositories; and (4) after merging with other data.

- Active data review methods would involve a person going through the established QC checks, verifying, and validating data. Other methods are more passive, with automated procedures or dynamic reporting providing QC checks on data.
- Data Stewards (or other reviewers) will need to examine requirements governing the collection or generation of the data they will be reviewing, results of field and laboratory QC checks, and other information (e.g., post-season debriefing results) that may shed light on the data being reviewed. Having access to the requirements and all field and laboratory records during data review is vital.
- Procedures in place for version control or backup/retention plans to ensure original data are maintained (EPA 2019).

Data Review Checklist

- See below data review checklist (Table 3) provided in Appendix material (EPA 2019).

Creating Metadata

- Requirements for creating metadata are detailed in Chapter 2 of this Handbook.

Access

The ability to prepare, release, and share quality data to the public, other agencies, and internally is an important part of the lifecycle process. Some QA strategies are:

- When applicable, provide procedures or identify in the planning step on how you plan to share your data; ensure that access restrictions are clearly defined. This may be accomplished in your DMP, by including some scheduled review of whether the access/use or security restrictions are being implemented. A periodic review of how the access restrictions are defined might also be useful.
- Identify and format the data properly depending on how you anticipate the end user will use it.
- Provide all of the necessary documentation that assists in understanding the data and accompanying metadata so the intended audience can interpret it properly.

Evaluate

Evaluating data and data management processes allows us to ensure that Service data are accessible, discoverable, and of high quality. By establishing criteria to evaluate data, metadata, and the process of data management, we can measure progress and success of our data management program.

Data Assessment **"**is a continuation of the data review process that involves addressing data quality flags, assessing data quality indicators (DQIs) across measured, observed, and calculated variables, and reconciling the data quality with project assumptions and stated sampling objectives" (EPA 2019).

When planning data assessments and/or reporting activities, project planning teams should determine and document:

- who will be responsible for performing the assessment and/or reporting;
- when the activity will take place;
- how the assessment will be performed;
- how the quality of the data assessment and/or reporting activities will be controlled; and,
- how the project data will be used in decision-making (EPA 2019).

Data Trustees, working in conjunction with Data Stewards, should establish QA/QC strategies for evaluation and data assessment during the Planning Phase. In most cases, this involves reviews and spot checks.

- The data assessment step is itself a QC process to determine data usability, but steps in this process should be reviewed internally – and possibly externally – to ensure the correct procedures and assumptions were used.
- QC steps should include spot checks to ensure the correct procedures were used to analyze the data and that results were accurately calculated.
- Good data documentation is essential throughout the data lifecycle. The following includes some best metadata practices:
    - Recording what the data are, how the data are collected, and the why the data are collected
    - Document how the data are structured / formatted
    - Ensure that the data are completed, documents instances in which data are incomplete
    - Track and document data version histories.

Data Stewards/Trustees should also need to evaluate the potential impact of flagged data and determine any limitations on the use of the data. Specific objectives of this process include the following examples (EPA 2019):

- Identify different types of data issues and/or instances of flagged data.
- Determine the possible effects of these data issues for the evaluation of project objectives.
- Decide how to address data issues such that the impact on data analysis is minimized.
- Communicate corrective actions taken and data usability limitations to data users.

Table 3-3 provides examples of potential limitations on data use as identified during data review efforts that affect all DQIs. Table 3-4 displays the potential effects of those limitations.

Table 3-3 Examples of Potential Limitations on Data Use as Identified During Data Review Efforts that Affect All DQIs"

Type of Flagged Data	Possible Causes
Missing data values	• Weather conditions or other factors prevented collection • Data values overlooked or not recorded
Data collected outside of data quality acceptance criteria	• Imprecision or bias for continuous variables • Species identification error
Violations of SOPs	• Incomplete training and crew certification • Procedures for collecting data not followed
Other data flagged as suspect from verification or validation checks	• Data collection errors • Data entry errors • Spelling errors

Note: From "U.S. Environmental Protection Agency. 2019. *Application of Quality Assurance and Quality Control Principles to Ecological Restoration Project Monitoring*. Publication No. EPA/905/K-19/001. Chicago, IL: Great Lakes National Program Office."

Table 3-4 Potential Effects on Data Usability

• Reduction in sample size affecting Type 1 and Type 2 error rates • Potential reduction in ability to compare key variables over time or space • Potential reduction in usable data for decision-making • False positives, false negatives, incorrect classifications, or other data errors affecting data interpretation • Incorrect calibration of equipment resulting in reduction of usable data • Poor data quality adds unforeseen expenses to projects when data collection methodologies need to be repeated to correct mistakes

Note: From "U.S. Environmental Protection Agency. 2019. *Application of Quality Assurance and Quality Control Principles to Ecological Restoration Project Monitoring*. Publication No. EPA/905/K-19/001. Chicago, IL: Great Lakes National Program Office."

Archive

Data archiving supports the long-term storage of data and the methods used to read or interpret them. Some QA strategies for archiving data are: (1) Identifying policies and documenting those policies on how long you will maintain the data for; (2) Who is responsible for the data and where the data will be stored once archived; (3) How will all the accompanying information (electronic data, paper datasheets) be stored; and (4) How will it be documented that the physical connection between the two can be maintained.

Summary

QA/QC touches all aspects of the data lifecycle and is useful in establishing and maintaining data integrity, quality, and utility. It is recommended to provide some basic QA/QC practices to your project and/or data collection efforts. Below is a checklist of best practices for QA/QC.

PLANNING
- ✓ Data management roles and responsibilities
- ✓ Standard Operating Procedures (SOPs)

ACQUIRE
- ✓ Observer training and testing
- ✓ Data collection forms
- ✓ Quality control practices during field activities

MAINTAIN
- ✓ Procedures for data review
- ✓ Review notes for unusual occurrences or events that may help explain data anomalies

ACCESS
- ✓ Schedule data-quality reviews at important points in your workflow
- ✓ Maintain data-quality metadata and documentation

EVALUATE
- ✓ Track data changes and implement a versioning scheme for your data
- ✓ Data assessment, analysis, and reporting
- ✓ Periodically run test data through all processing scripts to verify expected functionality
- ✓ Compare new data to historical values
- ✓ Plot spatial data on a map to verify locations
- ✓ Use data quality indicators, or at least comment fields, to qualify data anomalies
- ✓ Calculate summary statistics for data or display data using common graphs such as box plots to evaluate for possible anomalies

ARCHIVE
- ✓ Policies in place for maintaining the data and all project documentation

Additional checklists adapted from the EPA are included in the Appendix.

References

American Society for Quality. 2014. *ASQ/ANSI E4:2014: Quality Management Systems for Environmental Information and Technology Programs*. Milwaukee, WI: ASQ Quality Press.

International Organization for Standardization. 2015a. *ISO 9000:2015, Quality Management Systems – Fundamentals and Vocabulary*. Geneva, Switzerland: International Organization for Standardization.

International Organization for Standardization. 2015b. *ISO 9001:2015, Quality Management Systems – Requirements*. Geneva, Switzerland: International Organization for Standardization.

Office of Management and Budget (OMB). 2013. M-13-13. Open Data Policy-Managing Information as an Asset. In *Memorandum for the Heads of Executive Departments and Agencies, From Silvia M. Burwell, Steven VanRoekel, Todd Park, and Dominic J. Mancini. Executive Office of the President, Office of Management and Budget*. Accessed January 15,2018. https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2013/m-13-13.pdf.

U.S. Environmental Protection Agency. 2019. *Application of Quality Assurance and Quality Control Principles to Ecological Restoration Project Monitoring*. Publication No. EPA/905/K-19/001. Chicago, IL: Great Lakes National Program Office.

U.S. Environmental Protection Agency. 2002. *Guidance for Quality Assurance Project Plans (EPA QA/G-5)*. Publication No. EPA/240/R-02/009. Washington, DC: Great Lakes National Program Office.

Chapter 4-Planning

Why is Planning Important?

Planning underlies every step of the data management lifecycle. By planning before investing time, effort, finances, and other resources (i.e., data acquisition, maintenance, access, evaluation, archival, and quality assurance/quality control), you optimize each of those investments by accounting for their relationship to each of the other stages. The final product – discoverable, accessible, and usable data – is a result of effective data management across each stage.

Data Management Plans

Per the Service's data management policy (274 FW 1), a Data Management Plan (DMP) is defined as documentation that:

> *Describes data that will 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 during and after the completion of the activity where the data are involved.*

In other words, a DMP is a formal document that outlines how you will handle your data both during and after the project is completed. The goal of a data management plan is to consider the many aspects of data management, metadata generation, data preservation, analysis, and the resources required to develop and maintain all these functions before the project begins. These functions are key components of well-managed data system across its entire lifecycle.

Section 1.9 of 274 FW 1 provides the minimum required elements for a DMP. In addition to those elements, a comprehensive or well-planned DMP should include the following:

- **Describe.** Describe data inputs acquired from existing sources (e.g., provenance, documentation, and use restrictions); anticipate the full array of data products and types generated including primary (e.g., field-collected) and secondary (e.g., derived from analysis or modeling) data;
- **Acquire**. Describe how new data will be collected or existing data will be leveraged or reused, including analytical tools and software;
- **Maintain**. Describe how the data will be updated. If there is no intent to update data, that should also be noted.
- **Quality Assurance/Quality Control**. Describe any Quality Assurance (QA) and Quality Control (QC) procedures used;
- **Access.** Specify how and when the data will be shared and/or accessed; and
- **Archive.** Describe archiving, data delivery, and long-term maintenance measures.

To ensure your DMP meets the requirements of 274 FW 1, use the DMP template that is available on the Data Management SharePoint site. Data Stewards are responsible for creating data management plans using the template and updating and maintaining that plan throughout the lifecycle of the data. We encourage you to add details suitable to your project to maximize the utility of data produced.

Who is Responsible for Data Management Planning?

As mentioned above, and per 274 FW 1, the Data Trustee and Data Steward coordinate to develop a DMP and update it as needed. Data Producers and Custodians are responsible for following DMPs.

When Do You Need a Data Management Plan?

Service employees should always determine whether relevant data resources exist prior to generating any new data or DMPs. Whenever existing data resources do not meet your needs, you should begin planning how to generate the needed data.

The completion of a DMP is a critical first step in this process. The DMP is a 'living document' that will change as projects evolve and mature over time. Therefore, it is unnecessary to completely predict how you will manage data *ad infinitum* while developing initial data management plans. DMPs should, instead, be examined and updated regularly to reflect the changing objectives and products of the project.

Although this guidance leaves room for interpretation, the following activities may trigger the development of a DMP:

- Prior to data collection of any field-based or lab-based survey or administrative program;
- When acquiring biological or physical data from a partner or contractor, where practical given constraints in the mechanisms by which such data are acquired; or
- Prior to any subsequent sampling of ongoing projects lacking a DMP.

Scale of Data Management Plans

DMPs should be developed on a scale or level of organization that is most appropriate and efficient for meeting project objectives and complying with policy. The processes and procedures described within a DMP should contain a level of detail commensurate with the project scale. The roles and responsibilities of the Data Trustee, Stewards, Custodians, and Producers should be adapted to the scale of each DMP. Smaller-scale projects may require more localized oversight than larger-scale programs that may require oversight from staff at Regional office(s) or HQ. In some cases, project and data management planning may need to be documented at multiple spatial and/or organizational scales to accommodate hierarchies and dependencies between scales. Below are some case study examples that illustrate the scale-dependency of data management planning. These examples are not inclusive of all potential scenarios for data management in the Service.

Data collection or generation project where the data lifecycle is primarily implemented at a single organizational level or scale

A data collection or generation project occurring at a single organizational level such as an individual National Wildlife Refuge, National Fish Hatchery, Ecological Services field station, or Regional or national program where the entire data lifecycle is primarily conducted at that single organizational level would require a single DMP. One example of this scenario would be an inventory of wetland invertebrates at Hanalei National Wildlife Refuge. This project and associated data assets are managed at the local scale even while information is published and archived within national repositories. Another example would be the Corporate Master Table, which contains Service business data, and is administered at the national level.

Data collection or generation project where responsibility for the data lifecycle is distributed across multiple organizational levels or scales

Regional, multi-Regional, or national-level data collection or generation projects typically involve acquiring data at local levels and then aggregating those data into larger-scale, centrally managed information systems. Under this scenario, one or more DMPs *may* be required. A factor in determining whether multiple DMPs may be required is whether a single DMP could contain adequate descriptions for processes and procedures that would apply across organizational levels or scales. In other words, could a single DMP contain enough detail to ensure consistent implementation of data management practices across the entire lifecycle and between Data Producers and Custodians at different locations? If so, then a single Regional, multi-Regional, or national scale DMP could suffice. If not, then a hierarchy of DMPs could be considered where a single broad-scale plan provides guidance for (i.e., steps down to) multiple finer-scale plans. The finer-scale (i.e., more local-level) plans would inherit attributes from upstream plans; this would promote more consistency and accountability while also allowing for local-level deviations.

An example for the former scenario would be the environmental DNA monitoring for bighead and silver carps in the Midwest Region including the Great Lakes and Mississippi River Basins. Data for this program is collected by at least 10 Fish and Wildlife Conservation Offices throughout the Region. Standard operating procedures are defined by the eDNA Quality Assurance Project Plan (QAPP) and ensures standardization among each field office. All phases of the data lifecycle are implemented by the Service's eDNA program at the Regional level. In this situation, a single DMP produced to cover the full scale of the survey would suffice.

An example for the latter scenario would be the monitoring program for common murre and Brandt's cormorant breeding colonies in the California Current System. This monitoring program includes multiple different types of data acquisition performed across multiple National Wildlife Refuges and other lands within three states and two Interior Regions. Data are maintained, analyzed, and reported at multiple scales, including the local and regional (California Current System) scales. While aggregated data are managed regionally, state-level data are managed locally, with variability between states. Thus, in this situation, a broad-scale plan could provide guidance for Regional data management while multiple finer-scale plans could provide additional detail for local data management.

Chapter 5-Acquire

There are three general means of acquiring data: converting/transforming legacy data, collecting new data, or purchasing/exchanging from an external entity.

Finding Existing Data

Before using Service resources to acquire new data, employees should determine whether relevant data that would meet their purpose or need already exists and access these data if possible (Table 5-1).

Table 5-1-Summary of Data Acquisition Considerations

Consider…	Explanation
Program, Agency, and Departmental Policies	*What policies beyond Service Data Management and Standards (274 FW 1, 274 FW 2) may be applicable?* The Joint Administrative Office (JAO) Contracting site has guidance for specific project needs.
Business Needs (i.e., Data Management Plans)	*Why are we acquiring the data and what administrative, science, or management question(s) are being addressed? Are the data to be collected relevant to the questions needing to be answered?*
Data Standards	*Are there standard processes identified for how you should be acquiring the needed data? If developing your own, what attributes are needed to answer your fundamental questions? Is a new standard needed?* Consult the FWS Data Standards site for a list of approved Service-wide standards.
Accuracy Requirements	*How accurate and precise do the data need to be to be useful for the questions being asked? What analyses will you perform, and will those analyses adequately answer your question?* Consulting a statistician before acquiring data can help you identify exactly how much and how precise data you need to acquire (i.e., power analyses), saving resources and ensuring impact. Be sure to work in a bit of redundancy as a safeguard, where applicable. Service employees, contractors, and volunteers all assume responsibility for maintaining the Service's Scientific Integrity policy (212 FW 7), and written agreements/contracts should affirm this.
Cost	*Is it cheaper to buy than collect the data in-house?* Consider the full cost of the acquisition, such as federal labor time.

Consider…	Explanation
Currency/Maintenance (Maintain Section)	*How current do the data need to be? Real-time? How often should the data be updated? Who will maintain the data? Think through the whole data lifecycle before acquiring the data, not just the one-time collection event. Is it critical the Service retains access (e.g., if data access is on a subscription basis)?*
Timeliness	*When should the data be collected? For instance, vegetation data used to train satellite imagery should be collected around the same time of year as when the imagery will be collected. Will the data's usefulness expire and will updated datasets be needed?*
Formats (Access Section)	*What are the appropriate data formats to ensure data are open and machine-readable? Where will those data be stored? Do file names adhere to appropriate conventions (Maintain section)? Are files organized into appropriate directory structure (Maintain section)?*

We are developing this Handbook as other Service data management infrastructure is also being established, including a list of Board-approved repositories. As Service data becomes more discoverable and accessible, finding existing data will become more efficient. In the short-term, consult the Service Systems Inventory for a complete list of systems within the Service and explore systems that may reasonably contain your data. Points of contact are available to consult for each system.

Converting/Transforming Legacy Data

Data created, acquired, analyzed, or distributed prior to October 1, 2020 are considered *legacy data*. When converting and transforming legacy data, consider:

- *Are legacy data of sufficient quality*? Legacy data may require extra scrutiny to ensure an error is not propagated;
- *At what cost can data be converted into a usable format?* Converting or transforming data produces new datasets that would not themselves constitute legacy data.

If the data are considered relevant to support the Service's mission, a Data Trustee must bring legacy data into compliance with the data management policy at 274 FW 1 to the extent possible. A few scenarios where legacy data would need to become compliant include:

- Where the legacy data supports an ongoing project or activity;
- Legacy data supports a pending or anticipated regulatory action, rulemaking, or policy decision; or
- Legacy data are provided to the public.

Collecting New Data

Data collection should follow protocols outlined in the project's data management plan (DMP), including documentation of all applicable metadata. Ensure data collection standards are identified and followed. Both automated and manual data recording require quality control and quality assurance, but automated recording is encouraged to limit errors. Physical data sheets should be transcribed, and files

produced from transcription or those which were automatically generated should be made available for updates through the data lifecycle.

Human error is an inherent random error factor when collecting data manually. Follow Quality Assurance (QA) and Quality Control (QC) protocols to minimize this error (identified in DMPs; e.g., ensure equipment is calibrated and functioning properly), sample the data that is being collected early in the project to ensure it is as expected Document deviations from protocols where applicable.

Automated collection is another way to collect data. Several tools provide efficient workflows (e.g., Collector app, Survey123, drones, loggers), reducing random human error but also may introduce potential systemic errors. Follow QA/QC protocols to minimize these errors. As with any measurement device, documenting tools also may incur error in transcribing measurements. Follow appropriate QA/QC protocols and document which protocols were followed. In transcribing and converting data, transformation materials (e.g., coding scripts, translation software) used are important data themselves and should also be stored with the created datasets.

Additional Policies for Data Collection

Compliance with the Paperwork Reduction Act requires approval from OMB for any data collection involving 10 or more people outside of the Federal Government including the public and any collaborators. Check the Department Information Collection Clearance Program for updates as your first point of reference.

Protocols for data collection outlined in the DMP should follow standards compliant with the OPEN Government Data Act, such that information resources are accessible, discoverable, and usable by the public. Raw data may then be noncompliant due to factors like proprietary file formats associated with specialized equipment or be housed in local storage devices, but where possible and appropriate, and consistent with guidance in 274 FW 1 that data must be stored within a Board-approved repository throughout its useful lifecycle. Data should be converted to a usable format and uploaded to a repository where such data become discoverable and accessible.

Purchasing Data

There are several means of purchasing data:

- **Inter-agency Agreements (IAA**) are written agreements between entities of different Federal agencies to procure property and/or services from another Federal agency. Work is performed with or for another Federal agency under the Economy Act, 31 U.S.C. Section 1535.
- **Procurement Contract** is used to establish a relationship between the Federal government and a non-Federal entity (generally a commercial entity but may be any type of entity) when the principal purpose is to acquire property or services for the Service's direct benefit or use.

Special Conditions

Often with agreements, our partner may have very different goals related to the data than the Service has. If data need to be released to the Service immediately, specify that in the agreement documentation. It is common, for example, for a university or science agency to want to publish their findings and data prior to sharing with Service. It is best to clearly state how and when data are to be shared prior to executing any agreement.

Standards

Service Policy 274 FW 2 establishes Service data collection efforts shall follow data standards for documenting and formatting collected data, when they exist, approved by the Data Governance Board. When contracting for data that have an existing standard, this information should be included in the Statement of Work and require the selected vendor to provide data in that standard. See the Service Data Standards website for a complete list of currently approved Service-wide standards. The Federal Geographic Data Committee also maintains a list of national standards that should be followed when the Service is contracting for new data collection of those types

Licensing

Whenever possible, we should seek to procure non-licensed data that can be shared with the Public. This isn't always possible, however. Vendors usually apply a [restrictive] license that may limit distribution as opposed to an "open license". We should then acquire the most open license option available that meets project budgets.

Metadata

To protect Service investments, we should require metadata as a condition of all data acquisition awards. We need to clearly state data requirements and specifications such as metadata standards and file formats of deliverables in procurement announcements and awards Statement of Work. For example, when purchasing geospatial datasets, metadata shall be provided in a FGDC compliant-XML format.

Evaluation

Clearly state the evaluation factors that will be used in rating vendor proposals. Below is some sample language that could be used in a Statement of Work:

> *All Geospatial data developed by the contractor shall meet the federal and industry accuracy standards for geographic data.*
>
> *Any data developed by the contractor is the property of the government and shall be provided electronically to the Service with delivery dates specified in each individual task order.*
>
> *Any Contractor-supplied geospatial deliverables shall comply with applicable national and federal geographic information standards, particularly those determined by the Federal Geographic Data Committee (FGDC) as supporting the National Spatial Data Infrastructure (NSDI). Any geospatial data provided will be complete and will include a spatial component, attribute information and metadata. The data will be used for the MP-EIS and will be in compliance with applicable FGDC content standards.*
>
> *Accuracy statements reported by the contractor shall be completely and thoroughly substantiated by Metadata. The National Standard for Spatial Data Accuracy provides guidelines for Accuracy Reporting for reporting positional accuracy in Metadata.*
>
> *The contractor shall ensure that the metadata delivered to the Service is compliant with the Federal Geographic Data Committee Standard "Content Standard for Digital* Geospatial

Metadata", FGDC-STD-001-1998, which is available at: http://www.fgdc.gov/metadata/geospatial-metadata-standards.

Chapter 6-Maintain

Data maintenance includes the steps of processing data for analysis, creating metadata, and making sure data are in a format that is accessible over time. Maintaining data is important to the Service because unmaintained data do not have enough information for future managers to analyze or interpret. In addition, the format of the data must be kept current because the Service cannot often convert the information from an outdated version to a useable format, thus causing data to be lost. It is the responsibility of the Data Custodian to make sure the data are compatible with current software.

Data maintained by Service employees are often used to make significant decisions about wildlife or habitat management, so it is imperative that these data are of high quality, can be understood and used by others, and are stored in a format that is secure and safe.

File Directories, Folder, Dataset Titles, and File Names

Consistent and organized file naming conventions and directory structures can assist with finding data and understanding their content. Best practices for file naming include being consistent and descriptive in your file and folder names (OHSU 2020), however, there is no single correct method for naming files or organizing directories.

File Directories

Start with setting up a clear directory structure. Folders can be organized by project, date, researcher, experimental run, or whatever makes sense for you and your research or project. Keep in mind that file structure should be meaningful to the next generation of staff. It is important that you stay consistent once you set up a directory structure. Visual interpretation tools such as mind maps are helpful for designing a file structure (OHSU 2020).

Include a readme.txt file in the directory that explains your naming convention, format, and other features involved with naming your data along with any abbreviations, acronyms, or codes you have used.

Within Appendix B of Alaska's Interim Region Data Resource Management User Guide you can find an example of a tree structure for file organization on categories one might use as a baseline for content when setting up a directory.

File Folders, Dataset Titles and File Names

When naming the project folder (short title), files within, and even variable names (column headers in tables) keep the following best practices in mind:

- Be consistent throughout the life of your project.
- Keep names short, but meaningful that reflect the content of the folder/file. File or folder names may contain information such as project acronym or name, study title, location, investigator, year(s) of study, and type of data (Martin and Ballard 2010).
- Long file names don't always import well into various types of software. More descriptive information can be incorporated into the metadata.

- Use the International Organization for Standardization (ISO) date format 'back to front': YYYY, YYYYMM, or YYYYMMDD. It ensures that files with the same name and different dates can be sorted in order.
- Use leading zeros for clarity and to make sure files follow sequential order. For example, (01_FileName, 02_FileName...) instead of (1_FileName, 2_FileName...).
- When using personal names, give the family name first, followed by the first name or initials (e.g., SmithMary or SmithMC).
- Use only letters, numbers, dashes, "-", and underscores, and "_". Alternatively, use no separation between words or capitalize each word in the file (e.g., FileName.xxx). Do not use spaces or any other special characters because some software will not recognize them.
- Avoid using "draft," "version," or "final" in file names. Use date (in ISO format) to distinguish versions.
- Don't use acronyms, initials, abbreviations, and codes that are not commonly understood or make sure to document or define them within easily accessible readme.txt files.

Types of Data and Organization Best Practices

Service projects generate a few common types of data. These are tabular data, databases, geospatial data, collections of similar types of data (e.g., digital images, data logger outputs), and source data, which is data not generated by the project but is used during the course of the project (e.g., data used to inform project design). Best practices for organizing these types of data for long-term storage and access are described below; more specific guidance and trainings will be developed for each datatype in the future.

When organizing data, a Data Steward should remember to keep data simple and tidy. All too often, those who collect and enter data want to fuse data, analysis, and presentation. These are three separate things and should not be mixed together in a single platform.

For example, imagine a single sheet in a Microsoft Excel workbook that stores data in individual cells. Different cells have different formats to convey some meaning and formulas reference particular cells to preform calculations and return the results of the analysis, which ultimately generate a graph based on those results. On the surface, this appears to be efficient data management because everything is contained within a spreadsheet. However, this fusion of data, analysis, and presentation can become unwieldy. If new data are added, deleted, or changed, updating of formulas and associated graphics may require much work depending on their complexity, and constant updating can easily result in errors in results. Thus, it is best to keep data in a "raw" format that appears very simple when viewed and call those data into a separate platform when conducting an analysis of those data. Examples would be data stored in a flat .csv (comma separated values) file that is read into a statistical software program like R for analysis; another sheet within the workbook that uses a pivot table to summarize data; or tables within a relational database and queries in the same or different database that are used to perform a summary analysis.

Tidy data refers to a standardized way to structure a dataset (its physical layout). An excellent document on "Tidy Data" is Wickham 2014. A dataset is a collection of values which can be quantitative (numbers) or qualitative (text or strings). Values are organized by the variables and observations to which they belong. Variables refer to measurements of the same attribute (e.g., length, weight, temperature) and observations contain all the values measured on the same unit for each of the variables. Tidy data are most often organized in a rectangular fashion with rows and columns where the rows represent

individual observations and the columns represent the variables measured on an observation. The collection of observational units forms a table. Wickham (2014) identified five common data management problems that lead to messy data and should be avoided:

1. Column headers as values instead of variable names,
2. Multiple variables stored in a single column,
3. Variables stored in both rows and columns,
4. Multiple types of observational units are stored in the same table, and
5. A single observational unit is stored in multiple tables.

When data are "tidy," sharing of data among users and transfer of data between various analytical software applications is seamless without the need for manual manipulation of data.

Data collected by the Service is often stored in spreadsheets, relational databases, geospatial databases, and other platforms. Below are some best practices for each storage platform as recommended by the Alaska Region Interim Data Management User Guide.

Spreadsheets

Multiple spreadsheets within the same file can contain data and derivatives of the data (i.e., tables, summaries, pivot tables, formulas, figures). There are some best practices to use when dealing with these types of data.

- One sheet in the file should contain only a clean version of the data. **Nothing else** should be on this sheet. This is the tidy data which is simple in appearance.
  - The first row contains variable names; each column represents one variable. Variable names should use only letters, numbers, dashes, "-", and underscores, "_". **Do not** use spaces or any other characters.
  - The variable name should include the unit where this is relevant (e.g., length_cm and weight_gm).
  - Each row after the first row should represent one observation.
  - Avoid formatting information in this sheet (e.g., comma in the thousands place, font settings, border lines, colors). If the formatting is there to convey some information, consider adding a new variable to record that information instead.
  - For the purposes of tidy data, blank cells indicate that the data point is missing. "0" in a cell means that the data point was collected, and it was "0." (For example, a given quadrat set during a survey showed no species of interest in that particular quadrat.) Deviations from this convention should be recorded in the data dictionary (see below).
  - The tidy data sheet, in addition to being part of the workbook, should also be saved in an open format (e.g., TXT or CSV) using the same name as the Excel file (e.g., fish_data.xlsx and fish_data.txt) in the same archive folder as the Excel file.
- One sheet in the file should provide a brief description of each variable in the tidy data. Each row of this sheet represents one variable. This is termed the data dictionary and will be part of the metadata record for the tidy data.
- Other sheets in the file can contain summaries of the data (pivot tables), graphical representations of the data (figures), or derived quantities from the data (e.g., formulae, macros). Keeping the data in one sheet and summaries and graphics in another sheet prevents the fusion of data, analysis, and presentation.

- One sheet in the file should provide a brief description of each sheet in the file (what does it contain, any relevant information about its use). Each row of this sheet represents one sheet of the file. The first column is the sheet name and the second column is the sheet description.
- Save the original workbook in the most recent format supported by the application. For example, save Excel files in XLSX format rather than XLS format.

Relational Databases

Databases are essentially a collection of tidy datasets with relationships between the tables specified. Generally, but not exclusively, databases developed by field stations within the Service use MS Access. If data spans a Regional or national project, consider using SQL, Structured Query Language, a code-based way to work with a relational database through PostgreSQL or other tool.

When using relational databases, consider the following best practices:

- Variable names in each table should be described (i.e., a data dictionary is available for each table and defined in the table design when using Microsoft Access). When possible, the database should be documented within the database application (e.g., from within MS Access, add title and abstract to database properties and add description for all tables and fields).
- Constraints should be enforced on variables to promote Quality Assurance. For example, in a variable named "Sex," inputs could be constrained to the values of "Female," "Male," and "Unknown." Or, in a variable named "Length_mm," only integers between 10 and 1000 could be made allowable values.
- If possible, consider converting MS Access databases to SQLite, an open format that will preserve the database functionality. Utilities are available to assist with this but may require additional technical assistance.
- If conversion is not possible, MS Access tables and their data dictionaries should be exported to a preferred open format (e.g., text or .csv) and the database structure (relationships diagram in MS Access) should be saved to a preferred open format (e.g., JPEG or PNG) throughout the duration of the project and at the completion of the project. These files can be saved in the same archive folder as the database.

Geospatial Data

Geospatial data are often stored in a complex proprietary format (e.g., ESRI geodatabase) that is extremely difficult to archive for long-term preservation and access. A single geodatabase can contain many individual data sets. Each individual data set within a geodatabase typically consists of multiple individual files (in proprietary formats) that record the spatial information, attribute information, and other essential database properties required to use the data. The complexity of the geodatabase (e.g., a few individual feature classes or many feature classes with related tables and attribute domains) will determine the best methods to use when creating open data formats for archival purposes. It is recommended that you consult with your program's Data Manager, GIS Analyst, and/or Records Management (FWHQ_Records_Management@fws.gov) prior to archiving geospatial data. Broad steps for managing geospatial data include:

- Fully document each individual feature class or shapefile;
- Store individual geodatabases and shapefiles in the archive folder directory structure as with other data types; and

- Convert the geodatabase or shapefile to an open format (e.g., GeoPackage) and store in the archive folder.
- Use QGIS, Program R, or other open source software to work with geospatial data.

Collections of Similar Types of Data

Large batches of data are often collected in a project via cameras (e.g., from a remote camera taking time-lapsed or motion-sensitive images or from an aerial survey), data loggers, or other equipment. These batches, called a "collection" in metadata, should be saved in a single archive data folder with an accompanying text file describing relevant information for each file (e.g., location, equipment, resolution). In the case of photographic images, information for each file may be extracted from the metadata embedded in the image itself. The collection can be documented with a single metadata record.

Source Data

Source data refers to those data resources that were used by the project but were not created by the project. Source data are commonly used as input in the creation of new data products. Examples include base layers used in GIS processing or sensor input used to generate analytical output. If the source data are discoverable and permanently accessible through another means (e.g., United States Geological Survey (USGS) Streamflow data), this data does not need to be maintained in a project archive folder. However, if the source data are not readily available in the form used by the project, it would be appropriate to save the data and metadata in the archive folder by following the best practices for each data type as described in prior sections of this guide. In either case, a metadata record for the source data must be in the source data folder of the archive. The relationship between source data and products are described in the Lineage section of the product metadata record.

Relational Databases Versus Flat Files

A Data Steward is faced with the question of what platform to store data, whether a relational database or flat files. The choice depends on the structure, amount, and purpose of the data. Spreadsheets such as MS Excel are probably the most common method of data entry and storage. However, there are some limitations to MS Excel spreadsheets that should be considered (Alexander 2007). A MS Excel spreadsheet is limited to 1,048,576 rows which becomes problematic if the dataset contains more observations. In addition, the entire MS Excel file is loaded into RAM and every edit requires the whole spreadsheet to be loaded which can slow processing if there are a large number of cells filled with data and calculations are being ran. Conversion of a MS Excel spreadsheet to a flat file in an open file format such as TXT or CSV can eliminate some of these issues, but it also makes editing data more cumbersome. A relational database does not have these constraints.

The choice of relational database or a flat file depends on the situation. Table 6-1 gives examples of when one may be preferable over the other.

Table 6-1-Considerations for relational database versus flat files.

Data Situation	Preferred storage platform	Example File Types
Large dataset	Relational database	MS Access, SQL

Small dataset	Flat file	MS Excel, text, csv
Data that is continually updated and added to	Relational database	MS Access, SQL
Data that is collected one time only	Flat file	MS Excel, text, csv
Data with hierarchical relationships	Relational database	MS Access, SQL
Data with no relationships	Flat file	MS Excel, text, csv
Complex data manipulation and analysis	Relational database	MS Access, SQL
Multiple uses of the data	Relational database	MS Access, SQL

In general, data that has hierarchical relationships, repetitively collected through time, or has multiple users should use a relational database to ensure data standardization and quality are maintained. Flat files can accommodate such data if they are tidy, but more care is needed to ensure the structure of the data are not corrupted with continued manipulation of the data and the addition of new data. Flat files are completely adequate for data that is collected one time and one time only, such as the data collected during the course of a simple laboratory study that will only be conducted once.

Data Standardization

Data standards describe objects, features, or items that are collected, automated, or affected by activities or the functions of organizations. In this respect, data need to be carefully managed and organized according to defined rules and protocols. Data standards are particularly important in any co-management, co-maintenance, or partnership where data and information need to be shared or aggregated (National Land & Water Resources Audit 2008).

Benefits of data standards include (National Land & Water Resources Audit 2008):

- More efficient data management (including updates and security),
- Increased data sharing,
- Higher quality data,
- Improved data consistency,
- Increased data integration,
- Better understanding of data, and
- Improved documentation of information resources.

Additionally, the majority of data analysis is often spent cleaning and organizing data to get it into a format that can be analyzed and interpreted (Dasu and Johnson 2003). Just because data are entered into an electronic format does not necessarily mean that those data are accessible, analyzable, or interpretable. Standardizing our data can help us get there.

Standardized data are:

- Entered correctly and have been checked for accuracy. See Chapter 3, Quality Assurance and Quality Control for more guidance on Quality Assurance (QA) and Quality Control (QC).

- Entered in the same format: For example, phone numbers with area codes are all entered as either “123-456-7890” or “(123) 456-7890”.
- Consistent in units of measure: Do not mix units. For example, if the length of something is measured and recorded do not enter some data in feet, some in inches, and some in meters. Convert these measurements to a single unit of measure that is consistent throughout the dataset.
- Free of mixed data types: Text and numbers are not mixed. For example, if something is measured in feet, one should not enter “3’ “ or “3 ft.” Enter the number “3” and the metadata should indicate the units.
- Free of special characters (e.g., @, #, $, %, &, !) unless they have a specific meaning (e.g., @ in an email address). Special characters can have specific meanings in other computer programs and make analysis of data cumbersome, provide incorrect results, or result in programming errors.

It is the responsibility of the Data Steward to ensure that these practices to standardize data are instituted and kept up to date.

The Service provides a list of adopted data standards and provides guidance on establishing them.

Storage Options

This section of the Handbook focuses on recommended storage options for working data. Working data are actively being managed, preliminary, or that is updated frequently through a QA/QC process. Final data or products have gone through a review process, QA/QC completed, will not likely change again, and have an end goal of being made accessible to others. Examples include population estimates released annually or project completion.

All of the options mentioned below are advantageous for a user over saving files to their local hard drive (usually the C: drive) because there typically are no automatic backup security procedures already in place for local drives (see Backup Security below for more information). Storing data on recommended Service storage platforms can prevent data loss when hard drives crash or if laptops get lost, stolen, or damaged.

Working Data

Currently, the Service has a contract with Microsoft that provides users options such as OneDrive, SharePoint, and Teams.

Cloud Storage Options

OneDrive: One Drive is quickly becoming the Service’s preferred storage option because of ease of use and accessibility. OneDrive is a cloud storage service that allows you to store all your important files securely in one place and then access them virtually anywhere. Each user has a storage limit of 5TB in OneDrive.

Microsoft SharePoint: Microsoft SharePoint is a web-based platform used for sharing files and information. SharePoint allows easy access and updating by personnel from across multiple offices within our agency. It is designed for teams and provides collaboration features such as project

management, messaging, and shared document storage. SharePoint has a 1TB limit within the Service for each SharePoint website.

Microsoft Teams: Microsoft Teams is cloud-based team collaboration software that is part of the Office 365 suite of applications. The core capabilities in Microsoft Teams include business messaging, calling, video meetings and file sharing. The storage limit for Teams is currently 1Tb + 0.5Gb per enterprise licensed user. To get the total storage limit for your Team, multiply your users by 0.5Gb and add that to 1Tb.

Local Server or Network Attached Storage Device (NAS)

Service-preferred cloud storage options such as OneDrive, SharePoint, and Teams might not always be the best or fastest option. Service locations that have limited bandwidth or network speed should try and use a local server or a NAS device as their primary storage option. An NAS is used for file storage and enables multiple users at the same time to retrieve and manipulate data from a centralized space. The Service has several small offices (five users or less) that are using a NAS in place of a server for local data storage. A local server or NAS device can provide the user a redundant storage option meaning their data are not lost with a single hard drive failure (It takes multiple hard drives to fail before the potential for loss of data).

Final Data

When a user reaches the end of their working data and they feel as though it's reached a final product there are a few storage options. The Service will have a list of approved repositories which will be available later.

Backup Security

Backing up data or the storage platform it resides on protects its integrity against loss or corruption. OneDrive, SharePoint, and Teams provide cloud backup solutions and enable enterprises or individuals to store their data and computer files on the Internet using a storage service provider, rather than storing the data locally on a hard drive, compact or floppy disc, or backup media.

It is the responsibility of the Data Custodian to work with IRTM to ensure data are being backed up at Service locations with a local server. At the time of the Handbook's release, each Region within the Service is using their own preferred backup software with a goal of having one enterprise solution (i.e., one standard backup software) across all Interior Regions soon.

Staff Turnover: Best Practices

There is usually relatively high turnover within a resource agency as personnel frequently change positions or leave the agency altogether. This results in data that are often lost or abandoned after a position is vacated. It also results in confusion and inefficiencies when vacated positions are filled by a new person. It is incumbent upon managers and supervisors to create a seamless transition between employees occupying a position that has a role managing data. Below are some actions a supervisor can take to avoid confusion and loss of data:

- Have the exiting employee upload all data that has been checked for quality control to an accessible, centralized storage location.

- Follow best naming conventions for files.
- Delete preliminary versions of data and multiple copies of data. Only final, or the most up-to-date data should be retained to avoid confusion on data versioning.
- Have the exiting employee transfer appropriate metadata, data files and code to the supervisor and/or their replacement.

Implementing these few simple steps when an employee leaves a position will greatly increase the efficiency at which the next employee can be brought up to speed and become fully productive in their new position.

References

Alexander, M. 2007. Microsoft Access 2007 Data Analysis.

Dasu, T. and T. Johnson. 2003. Exploratory Data Mining and Data Cleaning. Wiley-IEEE. https://onlinelibrary.wiley.com/doi/book/10.1002/0471448354. Accessed 1 October 2020.

Martín, E., and G. Ballard. 2010. Data Management Best Practices and Standards for Biodiversity Data Applicable to Bird Monitoring Data. U.S. North American Bird Conservation Initiative Monitoring Subcommittee. Online at http://www.nabci-us.org/

National Land & Water Resources Audit. 2008. *Natural Resources Information Management Toolkit: building capacity to implement natural resources information management solutions*, NLWRA, Canberra. Online: Accessed May 2009. http://lwa.gov.au/national-land-and-water-resources-audit/natural-resources-information-management-toolkit

Oregon Health & Science University Library (OHSU). "Best Practices for File Naming." https://www.ohsu.edu/library/best-practices-file-naming. Accessed 24 July 2020.

Wickham, Hadley. 2014. "Tidy Data." Journal of Statistical Software 59 (1): 1–23.

Chapter 7-Access

Open Data and Accessibility Requirements

The OPEN Government Data Act, in 44 U.S.C. § 3506(b)(6), codified longstanding Federal policy regarding "open data" by requiring all Federal data to be available in an open format by default. Further, it requires Federal agencies to make all their public data available as open data under an open license.

To implement these requirements in Service policy, section 1.8 of 274 FW 1 requires all Service data to be machine-readable and available in an open format, to the extent possible. Further, section 1.11 requires all data provided to the public (see the following section) to also be available under an open license so that there are no restrictions on using the data. Taken together, meeting these requirements will help ensure that our data are accessible to members of the public and/or internal parties, as appropriate, in accordance with the OPEN Government Data Act.

We expect further guidance in the future regarding open data, public data, and data accessibility requirements from the Department of the Interior (Department), Office of Management and Budget (OMB), and potentially other sources. We will update this handbook as that guidance becomes available.

General Accessibility Guidance

When sharing data with the public or scientific community, data should be released in multiple formats where practical. At least one machine readable and open format type should be included.

Machine readable format types include those such as CSV, XML, JSON, and ASCII. Open format examples include PNG, CSS, and CSV. The most comprehensive list of acceptable file formats can be found on the National Archives and Records Administration (NARA) website.

Other helpful file format guidance is available from other sources. The U.S. Geological Survey offers Data Management File Format Guidelines. The Library of Congress also has a recommended Formats Statement.

Questions to consider when converting data to open and machine-readable formats include what format is appropriate for your data and what tool will be used to convert your data (e.g., FME Desktop)? Answering these questions will help you make the most cost-effective decision for your needs.

After you convert your data, ensure the new file does not contain errors such as omissions, changes to column headings and rows, and truncated fields to ensure your data has not been altered during the conversion process. Verify that the metadata is updated and accurate. Finally, check to see that any markups, such as highlights or bolded text, are either removed, or are moved to the metadata so that important ancillary information is not lost in the conversion.

Public Access to Data

The OPEN Government Data Act and 274 FW 1 make it clear that data that is releasable under the Freedom of Information Act (FOIA) (5 U.S.C. § 552) is potentially subject to release to the public. As a result, the only data that would not be candidates for public release are those that fall under one of the FOIA exceptions at 5 U.S.C. § 552(b)(1-9), also taking into account privacy, security, legal, and other concerns described in the "Limitations on Public Access" section.

As mentioned previously and described in 274 FW 1 section 2.11(A)(2), all data provided to the public must meet the general requirements that apply to all data (e.g., described by metadata, available in a repository, etc.) but also must be:

- Machine-readable;
- Available in an open format; and
- Available under an open license

Review the following resources for more information about FOIA and the Service's responsibilities relating to releasing data under that law:

- FOIA.gov FAQs gives a great summary of FOIA including short videos and text.
- The Service's IRTM webpage has the most recent updates and notices regarding FOIA. It also includes Service FOIA Contact Information by Region in case you want to reach out to your Regional FOIA representative.
- The FWS FOIA Resources SharePoint serves as a reference point for Service employees. You can find resources to assist with FOIA Review Package Submissions, a training document on how to use FOIA's electronic tracking system (EFTS), a review checklist, a control form, sample resources to FOIA requests, exceptions, and information on FOIA searches using Department's eMail Enterprise Records and Document Management System.
- 203 FW 1, Freedom of Information Act (FOIA) Policy, Roles, and Responsibilities

Limitations on Public Access

274 FW 1 outlines considerations we must take when determining whether data are appropriate for public access and release. Not all data necessarily has to be released to the public in accordance with existing laws and regulations. Data release posing a privacy/confidentiality/security risk, or subject to specialized requirements may not have to be released. If data should not be released, the Data Trustee must ensure any restrictions to public access are documented and justified in the metadata for that data.

Following laws, regulations and polices outlined within 274 FW 1, Section 1.4, metadata must still be posted within Data.gov, regardless of whether or not the accompanying data are released to the public.

If you have concerns that your data may not be suitable for public release, you should contact the Service FOIA Officer, Associate Privacy Officer, CUI Contact, or Records Officer for assistance

FOIA Exemptions

As mentioned previously, data that would exempt from disclosure under the FOIA are not subject to release to the public under the OPEN Government Data Act. The FOIA exemptions are found at 5 U.S.C. § 552(b)(1-9) and many of them are spelled out in 274 FW 1 section 1.11(B). Such data could include confidential business data, data subject to intellectual property rights, national security information, and more. More information on FOIA exemptions can be found on the FOIA FAQ page.

Employees with questions about data that may be subject to FOIA exemptions should ask the Service's FOIA Officer or the Department Solicitor's Office before releasing it to the public.

Privacy Act Information

The Privacy Act of 1974 (5 U.S.C. 552a) provides individuals' rights of access, amendment, and accounting of disclosures to records the government maintains about them. More information about the Privacy Act and the Department's privacy program can be found on their website.

This protection includes personally identifiable information (PII) such social security numbers, home addresses, general privacy related information, source selection, contract use documents, archaeological resources, information on historic properties, and facility physical security information. For records about individuals, the best resource for what can and cannot be shared is the governing System of Records Notice (SORN). Government-wide, Department-wide, and Service-specific SORNs can be found on the Department's page. Please reach out to Service Associate Privacy Officer (APO) for assistance determining which SORN may be applicable.

Other Considerations

Other data that may have limitations on access include:

- Data that has access restrictions as part of a contract or other binding, written agreement. For example, publicly sharing commercial satellite imagery or publishing rights for access given to universities or the United States Geological Survey (USGS).
- Data that when released will result in the Service violating any existing law or policy, including
  - Unauthorized release of PII,
  - Violation of a System of Records Notice (SORN),
  - Maintaining PII in a system not authorized to host such data (e.g., hosting PII or CUI on ArcGIS Online),
  - Violation of litigation hold restrictions, and
  - Releasing crime scene data associated with a current investigation.
- Controlled Unclassified Information (CUI). Note that a CUI designation by itself does not necessarily mean that the data are not subject to release to the public under the FOIA and thus the OPEN Government Data Act. Contact the Service FOIA Officer and the Service's CUI Contact for more guidance.

Documentation of Restrictions to Public Access

If access to data is restricted, the Data Trustee must ensure any restrictions to public access are documented and justified in the metadata and data management plan (DMP). Documentation must include the following:

- Who does or doesn't have access? Are the data accessible only within the Service, available to other Department employees, or are they available outside the Department?
- What is the justification why the data are being withheld?
- A description of when access restrictions can be removed or might change. For example, after a litigation hold is lifted, the sensitivity of the data might change.
- The expiry of a publication hold deadline established in award letters/agreements.

Internal Access to Data

Section 1.11(D) of 274 FW 1 says that data must be available for Service employees to access internally for lawful Government purposes such as making regulatory decisions. In ensuring data are available to Service employees, data must be available to employees in an approved repository. While still being worked on, these files need to be available to everyone working on the project/data.

Data Producers must store data in a Board-approved repository that will allow employees to access in a timely manner. They must document repository location as part of the metadata along with instructions on how to access it and include any data liability and disclaimer language, if applicable. It is also recommended that Data Producers document protocols to be used before releasing data (e.g., inter-agency vs public, publication or peer-review processes, or what might be subject to FOIA) or performing requests for data not available within an accessible repository.

Sharing Data with Partners

If partners outside the Service need access to data to further the Service mission, the Department has approved communication and file sharing alternatives for sharing data and basic instructions on how to do so. See the Data Management Resources SharePoint site for the latest version. This method of file sharing with external partners offers an easier, faster, and cheaper method than trying to allow external partners to access Service systems like a server that stores your data.

If your partner or contractor must have access to your system then additional resources, time, and human resources need to be considered. Partners and contractors using Service systems need to have background checks and become adjudicated (PIV card and Service email address). It is never acceptable to allow a partner or contractor without a PIV card on a Service computer system such as allowing a partner to use a Service computer in the field or Regional office.

Section 508 Compliance

When sharing data or data products, it is important that information is accessible to everyone including those with disabilities. Data and digital products that the Service shares via a DOI platform or pays another to host their data and digital products must be, to the extent practical, compliant with 270 FW 4 Implementing Section 508 o the Rehabilitation Act.

Additional resources to assist employees with 508 Compliance like creating and testing documents, training, and Regional or program contact information are also available:

- Section508.gov | GSA Government-wide IT Accessibility Program
- The Service's IRTM Section 508 Compliance SharePoint Home Page
- The Department of Interior also maintains a reference website with information regarding 508 compliance Section 508 | U.S. Department of the Interior (doi.gov)

Chapter 8-Evaluating Data and Data Management

For the purposes of this handbook, we use the word evaluation to refer both to evaluating data itself, as well as evaluating the data management processes used to gather or collect the data at the project or program level. Evaluating data, as well as the processes used to generate that data, helps us measure progress and the success of our efforts.

All Service personnel with data management roles have a responsibility to evaluate their data as part of data management, though Data Stewards and Trustees have the specific responsibility to ensure that the applicable employees are following the data management plan (DMP) for the data as the data moves through each lifecycle step and that the DMP is effective. Each individual needs to evaluate their role and ensure that their assigned responsibilities are being conducted according to policy (see Table 1-2 from 274 FW 1).

The Service's Associate Chief Data Officer (ACDO) and the Data Governance Board (DGB) are responsible for establishing criteria to evaluate the Service's overall data management program. Further information on that topic will be provided in future versions of the handbook.

Evaluating Data

To effectively evaluate data it is imperative to examine the processes used to organize, interpret, and assess the quality of t data to ensure that it will be useful for making decisions and can be used to improve your project and/or program. As part of evaluating your data, you should take the following actions:

- *Organize*: You need to organize and "clean" your data to make sure that data are ready for analysis.
- *Assess quality*: Data need to be evaluated for quality before analyzing or sharing. See Chapter 3, Quality Assurance (QA) and Quality Control (QC) for details and methodology.
- *Interpret & Draw Conclusions:* What is the analysis telling you about your program or project? Sometimes it's useful to compare your data with other available data to better understand results. For example, it can be useful to compare the change you see in species numbers with existing data about changes across a larger population to understand how similar or different they may be.

Evaluating Data Management

Periodically evaluating the data management process is essential at the project/program level. This requires clearly documenting the process and any data analyses you perform, to increase transparency and reproducibility. Another suggestion would be to go back through each step of the lifecycle to ensure the implemented data management processes still meet the needs of the project or determine if revisions are necessary. Issues to consider include:

- Have needs or requirements changed?
- Are the data still meeting the project's needs?
- Do the data meet all standards set forth in the DMP?
- Has the use of the data changed?
- Do the applications and programs consuming the data need updating or changing?
- Is the metadata current and up to date?

- Are all security responsibilities being met?
- Are there any new regulatory requirements that must be met?

Table 8-1 provides a checklist of things to evaluate at the project/program level.

Table 8-1-Checklist of Evaluation Needs for the Process of Data Management

Evaluation Needs	What Effectively Meets This Need?	If the Need Isn't Met, What Needs to Change?
Assess Data Management Plan (DMP)	DMP is complete, current, and regularly updated or retired	Create DMP, revisit plan regularly, and update necessary fields
Assess need for Digital Object Identifier (DOI)	Determine if a is DOI needed for accompanying data assets. If needed, DOIs are obtained and documented in the metadata	Obtain a DOI. Contact Science Applications for more information on using the DOI tool.
Assess acquisition methods	Data have been acquired using standard protocol or following contracting guidelines	Adopt SOP for data collection methods and/or reexamine contracting guidelines
Assess data appropriateness and usefulness	Data are collected in a way that is still meeting project's needs	Reevaluate data collection methods and adjust to ensure data fulfill project scope and purpose.
Assess documentation	Data are well documented following established protocols and approved data standards	Document data according to established protocol and/or approved data standards
Assess metadata	Metadata fields are filled out following established data standards	Create metadata and fill out all fields following established data standards
Assess effectiveness of QA/QC protocols	Continue using documented QA/QC protocol	Go through QA/QC protocol or consider new QA/QC protocol
Assess access to data	Data are accessible according to policy and the DMP	Make data accessible according to policy and established guidelines, taking into account any recent changes in status.
Assess systems	All applications and programs consuming the data are regularly updated	Review periodically with IRTM

Assess security	All security responsibilities are met and regularly assessed	Review periodically with IRTM
Assess regulatory compliance	All regulatory requirements are met and regularly assessed	Review with Project Officer to ensure guidelines are met (e.g., NEPA, ESA, Cultural Resources)
Assess archive	Data have a plan for archiving that fulfills policy and guidelines and once complete plan is executed	Review periodically with Records Management

Chapter 9-Archive

By the time that data reaches the archive section of the data management lifecycle, it will no longer be regularly maintained and updated by Service staff. However, because of its potential long-term applications, the data may need to be preserved for future access and use. In addition, some Service data may be considered a Federal record and therefore certain record management requirements may apply.

This chapter will describe the differences between storing data and archiving the data. It will also describe the record management requirements that apply to data that must be transferred to the National Archives and Records Administration (NARA), which is the agency responsible for managing all Federal records.

What is Archiving?

As it applies to Federal records, archiving means collating a final, historically significant product with associated metadata for the purpose of preservation and permanent storage into a repository separate from where the product may be published for public or business uses. The archived product must meet the associated records schedule criteria. A records schedule defines the policy for information retention or disposal, and the criteria within dictate how records should be managed based upon their content or the context of their creation.

All data produced by Service employees could be considered records. The Service's Records and Information Management (RIM) Program will make the final decision about what constitutes a record as it pertains to archiving, but generally records are the output of the work of federal employees. Non-records are the other things produced in the course of this work. Examples of records include reports, professional correspondence, or grey literature. Examples of non-records include duplicate copies, phone messages, or sticky notes.

Historically significant records, whether digital (preserved in specific file formats) or analog (preserved on paper, microform, or similar media), are those that capture and preserve the legacy of the Service. For example, historical (50 years or older) studies and records related to policy decision-making.

If you are unsure if your record meets this criteria please reach out to the Service's RIM program for assistance at FWHQ_Records_Management@fws.gov.

Why is Archiving Important?

Archiving is important for several reasons. First, it is important to the legacy of Service data. Archiving preserves your work and your legacy, providing insight to future researchers and expanding the value of your work beyond its current business use. Second, there are legal requirements to manage records created by the Federal Government, including the archiving of data. Finally, archiving is important to keeping data safe. Digital files stored on local drives without access conditions can be altered. External hard drives deteriorate over time. Analog files stored in non-archival conditions like basements or attics risk damage due to humidity, water, or temperature fluctuations. File formats or the software relied upon for their use may become inoperable as technology changes. Properly archiving data can mitigate the risk of loss. Contact the Records and Information Management Program for assistance with archiving data.

What is the Difference Between Storing Data and Archiving Data?

Archiving data should not be confused with storing data. Storage refers to maintaining temporary, in progress, or in use products in a certain location. Storage also refers to keeping products in a centralized repository for the purposes of sharing with others. It also refers to maintaining historically insignificant products or those that do not meet records schedule criteria in local storage or a centralized repository. The data management plan (DMP) should describe local storage, metadata requirements, and the timeframe for the dataset's metadata to be published to a data catalog and the dataset shared to a public repository. The DMP should also indicate when the data may be considered final and potentially ready for permanent archive with NARA. Moving items to NARA will potentially still allow for the metadata to be published to data catalogs, and the data to be stored in an approved public repository as well as with NARA.

For more information about storage, please see Chapter 6, Maintain.

How Long Should I Maintain my Data?

Data created through the work of Service employees is subject to the Records Disposition Schedule. Each Federal agency has a Records Disposition Schedule which defines for how long and when records created by the agency are maintained, destroyed, or archived. The Service is currently covered by both the Department of the Interior (DOI) Departmental Records Schedule (DRS) and the FWS Records Schedule while the DRS undergoes complete review and approval by NARA. The DRS divides records into overarching categories:

- **Administrative**: Covers general administrative, human resources, finance and acquisition, and information technology records. This category is currently complete and you can review the final approved DRS-Administrative schedule and crosswalk for more information.
- **Policy**: Covers controls and oversight, judicial and legislative affairs, public affairs, regulatory development, and policy-related special media. This category is currently complete and you can review the final approved DRS-Policy schedule and crosswalk for more information.
- **Mission:** Once complete, it will cover records relating to DOI's mission, including natural and cultural resource management and protection, natural resource planning and development, science and technology, and more. Biological projects and related datasets will be found under this category. The Service's RIM is updating this category.
- **Legal, Regulatory Compliance, and Enforcement:** The Service's RIM is updating this category.

Detailed guidance on retention and disposal may be found in the schedules.

Steps for Archiving Records

Your DMP should indicate when your data will be archived with NARA, in accordance with the applicable records schedule, and the Data Custodian who will be responsible for the transfer process. Once the time has arrived, contact the Service's Records Management Branch at FWHQ_Records_Management@fws.gov. Copy, Paste, and answer the bullet points below, which contain important information the Records Management Branch needs to know.

- **Type of record:** Is the record analog (paper, physical media) or electronic (digital files)?
- **Topic of record:** This will correspond to a series title in the Records Schedule. Examples include management plans, capture photos, and wildlife policy development.

- **Extent:** How much space do your records take up if they are analog? If they are electronic, how big are the files?

NARA Requirements

NARA requires 15 elements for transfer of electronic records. The elements are repeatable and have one or more associated terms which define more specific meaning. More than the minimum required metadata may be included for records transfer. The required elements and their associated terms are defined in NARA Bulletin 2015-04, Appendix A.

The Service's RIM Program will work directly with you to prepare your metadata to meet NARA standards and to transfer electronic files to NARA. The transfer occurs through a secure FTP site after RIM and NARA meet. NARA verifies the transfer, and then transfer is complete.

Who is Responsible for the Data Once Transferred to NARA?

NARA will process and accession the records once transferred, taking physical and legal custody of the records. This could be physical custody, where the records are now held by NARA, physically or electronically. If there is a FOIA request for the records, the request should be directed to NARA because they have legal custody after the records transfer.

If the records are deemed to be interpretable as individual records and valuable to public researchers, NARA may include the records in its catalog.

Records can be retrieved after they have been archived. Anyone can access and download electronic records added to the NARA catalog, which includes both data and unstructured files. All requests to retrieve transferred records from NARA should be routed through RIM to ensure data storage and access policies are followed. For further instructions, see the NARA FAQs.

Digital Object Identifiers

Digital Object Identifiers (DOIs) are a system of identification for digital resources that allow for sharing and re-use of datasets and data citation. They can help with data archiving by providing a stable location for people to access the data in the future. If you are interested in using DOI, contact Science Applications about using the DOI tool.

Appendix

Standard Employee Performance Appraisal Plan (EPAP) Language

The Service's Data Management Policy (274 FW 1) requires Directorate Members, Managers, and Supervisors to ensure employees with data management responsibilities have data management duties included in their performance plans**.** There are many ways this can be accomplished. The expectation is the Programs will be in the best position to develop performance standards as appropriate. The following offer some suggestions that can be modified as needed based on the specific assignments in the employee's EPAP.

Language reflecting data management responsibilities can be inserted under the "employee excellence" or "science excellence" critical elements. Please consider the language below (*italics*) as a suggested goal statement for staff EPAPs. You may want to simplify this language to address basic responsibilities or provide detailed goals to best address the unique duties of specific groups based on their roles and responsibilities related to data management, for example, supervisory versus staff. EPAP language for personnel who have a primary responsibility for collecting or managing data may be fairly detailed; for instance, a Fully Successful Performance Standard may read:

> *Collects and/or uses scientific data and information in a manner that fully adheres to the Service's*
>
> *Data Management policies and other Service policies involving data. These data management responsibilities may include the development of Data Management Plans that include appropriate metadata and stipulate how data will be stored, accessed, and shared within and outside of the Service, as well as how they will be addressed in any contractual agreements**.** Adheres to the best practices, guidance, and procedures developed by the Data Management Governance Board and Associate Chief Data Officer.*

EPAP language can be simplified for employees without project specific data management planning

duties. For example:

> *Collects, manages, and/or uses scientific data and information in a manner consistent with current Service policies related to data and data management.*

EPAP language can also be included under other critical elements. For example, a Refuge biologist may have a Critical Element that says in part, "Conserve/manage/restore wildlife and their habitats to fulfill refuge purposes; provide resources to support the mission. Is implementing Strategic Habitat Conservation (SHC) framework through biological priorities on the refuge. Ensures that activities conform to all applicable regulations and policies governing the NWRS." Suggested EPAP language for Fully Successful might read:

> *Develops data management plans sufficient to meet Service data management policies. Inputs biological data into databases (such as biological priorities in PRIMR) consistent with data management plans.*

For a Refuge supervisor, additional language might include for this Performance Element:

Refuge managers must align ALL of their staff to work towards their identified biological priorities. Significant progress should be made to complete associated priority establishment, protocols, delivery, monitoring and data management and recording to address those biological priorities during the fiscal year.

Manager/Supervisor critical elements should be directed where practical toward specific and measurable attributes that allow it to be determined if they are ensuring that employees are following policy.

Additional EPAP elements can be written based on the responsibilities for Managers and Supervisors and Service employees found in Table 1-1 of 274 FW 1 and the data management roles described in the Roles and Responsibilities section of this Handbook. For example, if an employee serves as the Data Steward for a Program, their EPAP should include as Fully Successful the roles and responsibilities including serving as the technical point of contact for data and developing or complying with existing standards for the data, among others.

The Department of the Interior's Performance Management Handbook (370 DM 430 HB) and policy on performance management (370 DM 430) are useful references for follow-up.

General Resources

Glossary of Terms

A. Controlled Unclassified Information (CUI). Information that the Service creates or possesses, or that an entity creates or possesses for or on behalf of the Service, that a law, regulation, or Governmentwide policy requires or permits us to handle using safeguarding or dissemination controls as specified in 32 CFR 2002 and the CUI Registry.

B. Data. Recorded information, regardless of form or the media on which they are recorded (44 U.S.C. 3502(16)). This information can be unprocessed or processed and represented as text, numbers, or multimedia.

C. Data Custodian. The person assigned to perform the technical tasks necessary to manage the storage, integrity, and security of data.

D. Data Lifecycle. The stages of data's useful life describing its creation, maintenance, distribution, reuse, and storage, until it becomes obsolete.

E. Data Management. Refers to the planning, collection, storage, maintenance, retrieval, and quality control and analysis of data throughout its lifecycle.

F. Data Management Plan. Describes data that will 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 during and after the completion of the activity where the data are involved.

G. Data Producer. The person who is directly involved with creating, acquiring, analyzing, or distributing data.

H. Data Repository. A centralized location where logically organized data may be stored and shared. A data repository may be general or focus on a particular topic and may be public or internal to the Service.

I. Data Steward. A technical point of contact for data and related attributes who performs the oversight and maintenance functions described in Table 1-2 (3) of 274 FW 1 for data under their stewardship.

J. Data Trustee. The person responsible for ensuring that adequate resources are available to support data management for data within their scope of responsibility.

K. Inter-Agency Agreement. A written agreement entered into between two Federal agencies, or major organizational units within an agency, which specifies the goods to be furnished or tasks to be accomplished by one agency (the servicing agency) in support of the other (the requesting agency).

L. Legacy Data. Data created, acquired, or distributed prior to October 1, 2020.

M. Machine-Readable. A data format that a computer can easily process, without human intervention, and with no meaning being lost. See 44 U.S.C. 3502.

N. Metadata. Structural or descriptive information about data such as content, format, source, rights, accuracy, provenance, frequency, periodicity, granularity, contact information, publisher or responsible party, method of collection, and other descriptions. See 44 U.S.C. 3502.

O. Open Format. A format for data that is unencumbered by any copyrights, patents, trademarks, or other restrictions, so that anyone may use it at no monetary cost for any desired purpose. Open formats are non-proprietary, platform-independent, and machine-readable.

P. Open License. A legal guarantee that a data asset is made available at no cost to the public and with no restrictions on copying, publishing, distributing, transmitting, citing, or adapting such an asset. See 44 U.S.C. 3502.

Q. Personally Identifiable Information (PII). Any information that permits the identity of an individual to be directly or indirectly inferred, including any information which is linked or linkable to an individual. Some PII is not sensitive and does not require special handling, such as information on a business card or in an email signature block. However, some PII is considered sensitive and requires stricter handling requirements. This is due to the fact it could result in substantial harm, embarrassment, inconvenience, or unfairness to an individual if lost, compromised, or inappropriately disclosed. Examples of sensitive PII on their own include Social Security numbers, Tribal enrollment numbers, financial account numbers, date of birth, and biometric identifiers (e.g., fingerprint, facial image). Examples of PII that may become sensitive in conjunction with an individual's identity are citizenship or immigration status, medical or health information, or performance rating.

R. Quality Assurance. The process and procedures used to prevent errors while collecting or entering data.

S. Quality Control. The process of identifying and correcting errors and quality issues with data.

T. Records. All recorded information, regardless of form or characteristics, made or received by a Federal agency under Federal law or in connection with the transaction of public business and preserved or is appropriate for preservation by that agency or its legitimate successor as evidence of the organization, functions, policies, decisions, procedures, operations, or other activities of the U.S. Government or because of the informational value of data in them. See 44 U.S.C. 3301.

Standards List

Note: Many of the standards listed below have not been adopted as official standards. However, some organizations have and adopted listed standards. This list is a living list and needs continual revision. Listed standards are annotated with what type of standard they are appliable to (i.e., vocabulary, content, procedure, data, metadata, format, classification, etc.) See Chapter 2, Metadata for more content on the standard topic.

Departmental

DCAT2 (metadata): DCAT is an RDF (Resource Description Framework) vocabulary designed to facilitate interoperability between data catalogs published on the Web. https://www.w3.org/TR/vocab-dcat-2/

Service Data Standards

The Service keeps a list of currently adopted Service-wide standards on the Data Standards internet site.

Programmatic/Regional Specific Standards

Science Applications Project and Product Metadata Content Standard: The Science Applications Project and Product Metadata Content Standard is currently available in draft format**.**

IRTM Branch of Geospatial Data Services AGOL Requirements for Shared public items (metadata)

Service or Partner System Specific Standards

ScienceBase Item Core Model – a.k.a. sbJSON (metadata)

ISO 19115-2 (metadata) is a metadata standard developed and adopted by the International Organization for Standardization (ISO) that defines how to document and describe information.

FGDC CSDGM (metadata) Content Standard for Digital Geospatial Metadata (CSDGM), Ver. 2 (FGDC-STD-001-1998) is the current version of this FGDC authored and endorsed standard.

Data Repositories

At this time, the Service Data Governance Board is working on an official list of approved repositories for Service use. This document will be updated when further information is received.

Data and/or Project Catalogs

A data catalog is an organized inventory of data assets in an organization. It uses metadata to help organizations manage their data. It also helps data professionals collect, organize, access, and enrich metadata to support data discovery and governance.

Listed below are some examples of Data and/or Project Catalogs. This list is a living list and will be continually revised. (Please note that all the links are to federal resources.)

1. ServCat (data & project)
2. AGOL (data & project)
3. ScienceBase (data & project)
4. Landscape Conservation Cooperatives Science Catalog (data & project)
5. GeoPlatform (data)
6. Data.gov (data)

Training and Help Resources

Please see the Service's Data Management SharePoint page for a list of training and help resources. The Service is looking into procedures and technology to allow the continual curation of this page to ensure it being up to date.

"One off" trainings are encouraged to be communicated using the Microsoft "Service Data Management" Team's "Training Opportunities" channel. The "Tools and Resources" channel is also a valuable resource.

Supplemental QA/QC activities adapted from EPA (2019)

Table A1 - PREPARING FOR DATA COLLECTION – CHECKLIST

- **Identify, develop or modify standard procedures for data collection and handling.**
  - Verify the procedures to be used are applicable to the specific needs of your project and are consistent with the goals and objectives.
  - Verify the procedures are applicable to project site locations, the ecological community and species of concern, and the anticipated ranges of environmental conditions including associated spatial and temporal variability.
  - Ensure the procedures are documented clearly and appropriately (including revision tracking) and can be easily understood and followed consistently throughout the project.
  - Verify the procedures are thorough, and include all instructions and details required to collect and/or handle data needs. Test the procedures before use to verify they are clear enough to be applied consistently by individuals with varying levels of experience.
  - Ensure procedures are available to all staff prior to data collection and handling.
- **Verify that personnel who will be conducting field activities have demonstrated competency.**
  - Ensure that all individuals who will be collecting and/or handling data have the required expertise and experience and have been sufficiently trained to safely and accurately perform their assigned tasks.
  - Ensure training provides an understanding of the purpose of the project, affirms the importance of each crew member's role, provides project-specific instruction and practice in both the field and classroom, emphasizes safety, and includes an evaluation of crew performance.
  - Plan and implement procedures to certify, document and monitor crew member competency, including plans for re-training or reassigning personnel who fail to demonstrate competency.
  - Plan and implement procedures for conducting periodic assessments to ensure crew competency throughout the project.
- **Address site access requirements.**
  - Obtain required state and/or federal permits and site-access permission from local landowners.
  - Notify local authorities of planned field activities on public lands; provide them with appropriate contact information and other information needed to support effective public outreach or emergency response.
  - Verify suitability and accessibility of selected sampling sites and confirm that SOPs are applicable and appropriate for the site conditions.

- **Address field logistics before initiating data collection activities.**
  - Develop and distribute schedules.
  - Assemble field equipment; ensure equipment is clean, calibrated, and in good working order.
  - Ensure crew members have all necessary equipment and supplies, including documented procedures, site logistics, maps, data forms, communication devices, equipment instructions, and first aid kits.
- **Make arrangements for sample analysis.**
  - Define and understand your analytical needs, including the number and type of samples, target analytical parameters, procedures to be used and associated QC requirements.
  - Identify qualified laboratory(ies).
  - Ensure laboratories are able to analyze your samples in accordance with sample holding time restrictions and project schedules.
  - Establish clearly defined contract(s)/agreement(s) with the laboratory(ies), including reporting requirements and schedules

Note: From "U.S. Environmental Protection Agency. 2019. *Application of Quality Assurance and Quality Control Principles to Ecological Restoration Project Monitoring*. Publication No. EPA/905/K-19/001. Chicago, IL: Great Lakes National Program Office."

Table A2 - QUALITY CONTROL DURING FIELD ACTIVITIES – CHECKLIST

- **Establish and maintain good field practices.**
  - Ensure field crews have all the necessary information, documentation, equipment and materials, training, and data reporting forms needed to perform their assigned field activities.
  - Establish communication methods between field crews and designated project leader(s).
  - Ensure field crews have the necessary expertise and materials to record collected data and associated information clearly, legibly and with the level of detail required to address all target variables, including ancillary information.
  - Establish and implement procedures for field crews to review data during and immediately following collection (e.g., use of multiple crew members, verbal repetition, reviewing transcribed results prior to leaving sampling location).
  - Establish and implement procedures to ensure all specified data collection activities are completed before leaving sampling sites (e.g., double check all collected data, notes and comments; ensure all data forms, photographs, recordings, charts, samples and/or specimens are correctly labeled, organized, and stored; and ensure samples requiring laboratory analysis are handled in accordance with specified handling, transport & chain-of-custody procedures).

- **Establish qualified QA crews.**
    - Identify and obtain access to individuals with the appropriate levels of experience, proficiency, and expertise to perform as QA crew members.
    - Clearly document QA crew activities and responsibilities.
- **Determine field QC strategies.**
    - Determine the type of QC field checks that are appropriate for your project.
    - Determine the location, timing and number of QC field checks that are appropriate for your project. Chose a strategy that allows a reasonable assessment of sample and data collection procedures, field crews, and the resulting data. Consider prioritizing checks for less experienced crew members and more challenging procedures. Ensure the strategy provides QC check data that are sufficiently representative of site conditions and target variables.
    - Use QA crews to perform QC field checks where feasible; rely on routine field crews to conduct precision checks in other circumstances or as an additional QA tool.
    - Conduct all QC field checks within a time frame that sufficiently mimics site conditions at the time of routine field crew data collection. Address transitory variables by conducting QC checks in real time (concurrent with routine data collection).
- **Use and reporting of QC field check results.**
    - Establish and implement procedures to:
        - Ensure QC field check data are complete and accurately associated with corresponding field crew data.
        - Compare QC field check data to routine field data and evaluate the results against data quality acceptance criteria.
    - Report, review and consider results of QC checks within a time frame that allow for sufficient evaluation and corrective action.
    - Establish and implement communication between field crews, QA crews, and project decision makers to evaluate and act upon QC field check results and data collection activities.

Note: From "U.S. Environmental Protection Agency. 2019. *Application of Quality Assurance and Quality Control Principles to Ecological Restoration Project Monitoring*. Publication No. EPA/905/K-19/001. Chicago, IL: Great Lakes National Program Office."

Table A3 - DATA REVIEW – CHECKLIST

- **Plan data review activities in advance.**
    - Determine who will review project data (identify data reviewers with the appropriate knowledge, skills and experience, and objectivity).
    - Identify which information will be reviewed by which data reviewers.
    - Identify when data review activities will be performed.
    - Determine how data will be reviewed and how the results will be reported.
    - Identify the materials and information data reviewers will need when conducting their reviews.
    - Determine how the data will be certified for assessment and use.
    - Document these who, what, when, and how data review decisions in written procedures.
- **Verify and validate project data.**
    - Review project data to ensure specified requirements have been met.
        - Are all required data, including QC data, present and legible?
        - Have all revisions or corrections been signed and dated?
        - Have all required samples and voucher specimens been collected and transferred to the appropriate facility within the required holding time?
        - Were the data documented and transcribed accurately?
        - Are units, species names, and chemical names reported correctly and consistently for each variable?
        - Are data completely and correctly linked to supporting or related information?
        - Are QC data correctly associated with corresponding field data?
        - Were approved field and laboratory procedures followed?
        - Have acceptance criteria been met?
        - Has feedback from routine field and QA crews been documented?
    - Review project data to ensure they are scientifically logical.
        - Have results that appear to be impossible, illogical, or outside the anticipated range been evaluated?
        - Have values reported for one or more related variables been confirmed for consistency with scientific expectation?

- Have results been confirmed for temporal and spatial consistency with scientific understanding of the ecosystem, its condition at the time of data collection, and its inherent ecological processes?
- Are data consistent with collected specimens?

- **Handle data discrepancies and errors.**
  - Identify questionable data.
  - Determine how each questionable value is to be handled (i.e., corrected, accepted, and/or flagged).
  - Document the questionable data, along with corresponding decisions, corrective actions, and supporting rationale.
  - Implement corrective actions.
- **Certify project data.**
  - Document the data review process and its results, including any necessary data modifications.
  - Provide a description of any significant data problems, along with associated follow-up actions.
  - Provide an audit trail of all changes made to the data, including the reasons for the change, time stamps, and identities of data editors.
  - Provide a description of the data relative to the data quality acceptance criteria.
  - Provide a data manager's statement to certify that the data and data review documentation are true, accurate, and complete.

Note: From "U.S. Environmental Protection Agency. 2019. *Application of Quality Assurance and Quality Control Principles to Ecological Restoration Project Monitoring*. Publication No. EPA/905/K-19/001. Chicago, IL: Great Lakes National Program Office."

Table A4 - DATA ASSESSMENT, ANALYSIS, AND REPORTING – CHECKLIST

- **Plan data assessment, analysis, and reporting activities.**
  - Consider the purpose and intended use of the project data.
  - Determine who will be responsible for data assessment, analysis, and reporting (identify individuals with the necessary qualifications, background, and authority), including those who will be responsible for QC spot checks.
  - Determine when assessment, analysis and reporting activities will take place.
  - Determine which data will be assessed, analyzed, and reported.

- Determine how the assessment and/or analysis will be performed and how the results will be documented.
- Determine how the data will be used in decision making.
- Document these who, what, when and how decisions in written procedures.

- **Assess the impacts of data quality on the intended use of the data.** (Determine how the validated dataset(s) might impact the intended use of the data.)
  - Identify types of data issues and instances of data flags.
  - Determine the extent of the issues and data flags.
  - Determine the impacts of these issues and data flags on data quality indicators (DQIs) and data use (e.g., reduction in available data, reduction in ability to evaluate key variables).
  - Determine the impacts of these issues and data flags on the evaluation of whether project objectives have been met (e.g., potential risk of false positives or false negatives on decision making).
  - Decide how the data affected by issues or data flags should be handled.
- **Analyze and interpret the data.** (Determine whether project objectives have been met.)
  - Calculate summary statistics (e.g., to determine mean, median, percentiles, ranges standard deviation, coefficients of variation, and uncertainties) and prepare graphical representations to aid in understanding the dataset.
  - Identify, evaluate, and determine how to handle censored values (e.g., values reported as less than detectable) and outliers.
  - Consider and determine how to handle potential biases in the data.
  - Select and implement data analysis approach (e.g., comparisons of means or medians, regression/trend analysis, categorical outcome).
  - Verify assumptions for the selected statistical tests are met.
  - Document results of data analyses, including calculations, data used, and corresponding decisions and their rationale.
  - Complete a QC assessment of data analysis activities and results (e.g., were data analysis procedures followed, are calculations and statistical results accurate, are statistical assumptions correct, are results accurately and thoroughly documented).

Determine whether project goals and objectives have been met.

Note: From "U.S. Environmental Protection Agency. 2019. *Application of Quality Assurance and Quality Control Principles to Ecological Restoration Project Monitoring*. Publication No. EPA/905/K-19/001. Chicago, IL: Great Lakes National Program Office."