ITRC has developed a series of fact sheets that summarizes the latest science, engineering, and technologies regarding environmental data management (EDM) best practices. This fact sheet describes:

- the phases of the environmental data lifecycle
- considerations for effectively managing environmental data in each phase of the lifecycle

Additional information related to data management planning is provided in the ITRC fact sheets on Data Management Planning; Data Governance; Data Access, Sharing, and Security; Data Storage, Documentation, and Discovery; and Data Disaster Recovery.

1 INTRODUCTION

The data lifecycle is a conceptual path describing how data are created, stored, maintained, and retained (Figure 1). The stages of the data lifecycle defined herein include:

- Plan
- Acquire
- Process and Maintain
- Publish and Share
- Retain



Figure 1. Environmental data lifecycle.

Each stage is described in further detail below. This data lifecycle was developed based on a review of several resources including but not limited to:

- The United States Geological Survey (USGS) science data lifecycle model, which includes the following stages: 1) Plan, 2) Acquire, 3) Process, 4) Analyze, 5) Preserve, and 6) Publish/Share, with Describe (metadata, documentation), Manage Quality, and Backup and Secure defined as cross-cutting elements performed at all stages of the lifecycle. This model is also referenced in the United States Environmental Protection Agency's (USEPA) Best Practices for Data Management Technical Guide published in November 2018 (USEPA 2018). Additional information regarding the USGS science data lifecycle model can be found in the 2014 USGS report titled "The United States Geological Survey Science Data Lifecycle Model" (Faundeen et al. 2013). Figure 2 was originally presented in that report and is shown below for reference as a graphical depiction of the USGS science data lifecycle model.

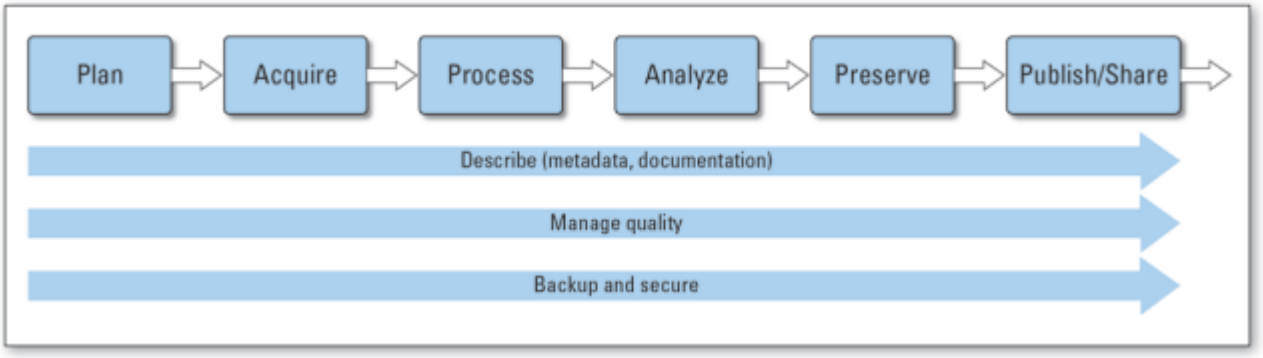


Figure 2. USGS science data lifecycle model.

- The U.S. Fish and Wildlife Service (USFWS) data management lifecycle, which includes the following stages: 1) Plan, 2) Acquire, 3) Maintain, 4) Access, 5) Evaluate, and 6) Archive, with Quality Assurance/Quality Control as an

element applicable at all stages. Additional information regarding the USFWS data management lifecycle can be found in the Data Management section of the USFWS website. Figure 3 was originally presented on the USFWS website and is shown below for reference as a graphical depiction of the USFWS data management lifecycle.



Figure 3. US Fish and Wildlife Service data management lifecycle.

The ITRC data lifecycle defined here is consistent with these two sources for the first three stages of the lifecycle (Plan, Acquire, Process/Maintain). Beyond these stages, the ITRC data lifecycle presented here differs slightly from the USGS and USFWS models as follows:

- The lifecycle presented here intentionally omits any stages related to data analysis, data evaluation, data interpretation, or any other actions related to drawing conclusions from a given data set. This is an intentional omission, as the focus of these ITRC documents is on data management, not on data analysis. The lifecycle presented in these ITRC documents represents the path of the data themselves, regardless of any potential prior, current, or future use of the data. The focus is purely on management of the data through their own lifecycle, not on the lifecycle of how the data may be used.
- The USGS and USFWS lifecycles differ in their order of the Access/Publish/Share stages and the Archive/Preserve stages, with the USGS model placing Preserve before Publish/Share while the USFWS model places Access before Archive. Because the focus of the ITRC data lifecycle is on the pathway of the data themselves, it is more appropriate to have the data retention stage (inclusive of data archival/preservation, retention, and deletion) as the last stage in the data lifecycle, as this represents the final state of the data. The Publish/Share stage of the ITRC data lifecycle (inclusive of access, publishing, and sharing) is therefore placed before the Retain stage, as this represents the time in the data's lifecycle when they are still actively being accessed by users prior to their final retention state.
- Cross-cutting elements are listed in both the USGS model (Describe, Manage Quality, Backup and Secure) and the USFWS model (Quality Assurance/Quality Control). Cross-cutting elements such as these are represented in the ITRC guidance in the Data Management Planning Fact Sheet in Figure 2. Specifically, Data Quality is represented as a row on this figure. Describe and Backup and Secure are considered to be elements of data governance, which is also represented in Figure 2 of the Data Management Planning Fact Sheet. For additional details on these and other cross-cutting elements, please see the Data Governance, Data Access, Sharing, and Security, Data Storage, Documentation, and Discovery, and Data Quality Fact Sheets.

Data management planning (including data governance and data management plans [DMP]) should consider and address all stages of the data lifecycle. Ideally, policies and plans related to data management at each stage of the data lifecycle should be in place *before* the data lifecycle process begins or should be developed during the Plan phase of the lifecycle.

Ideally, data governance policies should be in place before the data lifecycle begins or, if not, should be developed during the planning phase of the lifecycle. Organizations should address the requirements of each stage of the data lifecycle while maintaining documentation, quality, and security of the data. Also note that it is assumed that your organization has a suitable Environmental Data Management System (EDMS) in place. If not, please see the supplemental white paper

Environmental Data Management Systems for further details.

The data lifecycle is integral to many aspects of data management, including planning, quality assurance, data exchange, security, and communication. For information specific to each of these topics, see:

- Data Management Planning Fact Sheet
- Data Quality Overview Fact Sheet
- Data Exchange Overview Fact Sheet
- Data Access, Sharing, and Security Fact Sheet
- Public Communication and Stakeholder Engagement White Paper

2 PLAN

The Plan phase is used to evaluate the data project and define project- or task-level data quality objectives (DQOs), scope, or other objectives. These objectives can be determined by asking questions to define a problem. The Plan phase is also used to develop a data management plan (DMP), a document that outlines aspects of data management for a given project, data set, or task, with a focus on the objectives of the data collection and management. It should describe how data will be collected, processed, stored, and communicated. DMPs and governing policy provide support structure before a data lifecycle begins and serve as guidance during the data lifecycle. Moreover, the DMP should be considered a "living document" that will adapt to the changing needs and abilities of an organization's data management strategy. More information is provided in the Data Management Planning Fact Sheet.

3 ACQUIRE

Data acquisition includes all activities where new, existing, or legacy data are collected, generated, or considered for use in a given data set. Data may be acquired from many different and disparate sources, such as field data collection activities, affected people (surveys, interviews, testimonials, photos, videos), existing data sets, remote sensing activities, etc. A DMP should outline procedures to document data sources and data quality.

Governance policies and the DMP can provide a framework for consistent, effective, and reliable data acquisition if they address the following:

- methods for acquiring data, such as collecting data, converting/transforming legacy data, sharing and exchanging data, and purchasing data
- business or project needs, including DQOs
- data collection standards
- data update frequency
- data sharing agreements to ensure that shared data is used appropriately
- data entry protocols and data acquisition tools to control and standardize data entry
- security requirements, which may be different for different acquisition methods (for example, field collection versus loading from another data source)
- training and other requirements for staff who collect data to ensure they are qualified and knowledgeable about data collection methods

Please note that each of the above may be highly dependent upon the type of EDMS that is being used. Please see the Environmental Data Management Systems White Paper for additional information on that topic. More information on data acquisition is provided in the following fact sheets:

- Introduction to and Overview of Field Data Collection Best Practices
- Overview of Best Practices for Management of Environmental Geospatial Data
- Acquiring Traditional Ecological Knowledge Data
- Data Exchange Overview
- Data Migration Best Practices
- Electronic Data Deliverables

4 PROCESS AND MAINTAIN

Data processing and maintenance incorporates the steps and processes required to store, organize, verify, summarize, transform, integrate, organize, and derive data. This step is iterative and recurring, so plans for processing and maintenance should include protocols for updating data and managing change while ensuring that data quality is maintained and all data security requirements are met.

Data maintenance includes processing and evaluating data so they are suitable for project or organizational use, creating metadata, and making sure data are in a format that others can access in the future. Data processing may include integration of different data types and extraction, transformation, and load (ETL) operations needed to add and maintain data in the organization's EDMS. The selection of an appropriate EDMS is critical to processing and maintaining data. See the Environmental Data Management Systems White Paper for additional details.

Plans for data processing and maintenance should ensure that:

- Governing policies relating to maintenance, access, content (main data), and data protection are followed.
- DQOs and valid values (see Valid Values Fact Sheet) address the different environmental data types that will be used.
- Security requirements—as specified in DMPs or data governance documents—are considered. Note that different workflow steps may need specific attention depending on the type of security requirements specified.
- Staff who maintain the data are qualified and knowledgeable. This may require additional training in specific aspects of each individual DMP.

More information is provided in the Data Management Planning Fact Sheet, the Data Storage, Documentation, and Discovery Fact Sheet, Environmental Data Management Systems White Paper, and the Data Quality Overview Fact Sheet.

5 PUBLISH AND SHARE

If data are, or may be, published or shared between different entities (agencies/businesses etc.), data sharing methods and policies should be developed. For government agencies, data sharing protocols are especially important because data acquired by the agency become subject to open record requests or legal discovery. Data sharing agreements may be needed to define restrictions or appropriate use of shared data.

When data are published or shared, associated metadata should also be provided. Organizations should address the Americans with Disabilities Act, Section 508 of the Rehabilitation Act (29 U.S.C. § 794d), as amended by the Workforce Investment Act of 1998 (P.L. 105-220), which requires federal agencies and collaborators to develop, procure, maintain, and use information and communications technology that is accessible to people with disabilities, regardless of whether or not an individual works for the federal government. For additional information, see the Public Communication and Stakeholder Engagement White Paper.

For more information on data sharing and publication, see the Data Access, Sharing, and Security Fact Sheet and the Public Communication and Stakeholder Engagement White Paper.

6 RETAIN

Retention refers to the storing of data beyond the original intended use of the data to meet secondary use of the data and to meet regulatory and recordkeeping obligations. A data retention policy should be outlined in the planning stage if there are no overarching governance policies for data retention. Consideration should be given to any applicable state and federal agencies that may have specific data retention schedules that must be followed. The following questions should be considered:

- What data are retained?
- Who will be responsible for the data?
- What format should be used?
- How long should the data be stored?

- Should the data be archived or deleted?
- Who has the authority to dispose of data?
- When should the data be archived or deleted?
- What procedure(s) should be followed in the event of a policy violation?
- If data are purged, how can it be done in a manner that adheres to security and confidentiality protocols?
- Does your system have a data backup system, and is it performed regularly (either automatically or manually)? Is there sufficient redundancy so that data are recoverable in case of a system crash or other disaster?

Data retention options range from archiving to deleting. The decision regarding whether or not data should be archived or deleted and when these actions should occur should be made in accordance with your organization's retention policies. For additional information on data retention, see the Data Storage, Documentation, and Discovery Fact Sheet.

7 REFERENCES AND ACRONYMS

The references cited in this fact sheet, and the other ITRC EDM Best Practices fact sheets, are included in one combined list that is available on the ITRC web site. The combined acronyms list is also available on the ITRC web site.