ITRC has developed a series of fact sheets that summarizes the latest science, engineering, and technologies regarding environmental data management (EDM) best practices. ITRC also developed a Data Management Planning Tool to help guide the development of a data management plan.

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

1 INTRODUCTION

Data management planning is the process of developing data management policies and strategies to effectively acquire and manage data of known quality fit for purpose. It includes the development of data governance policies and data management plans (Figure 1).
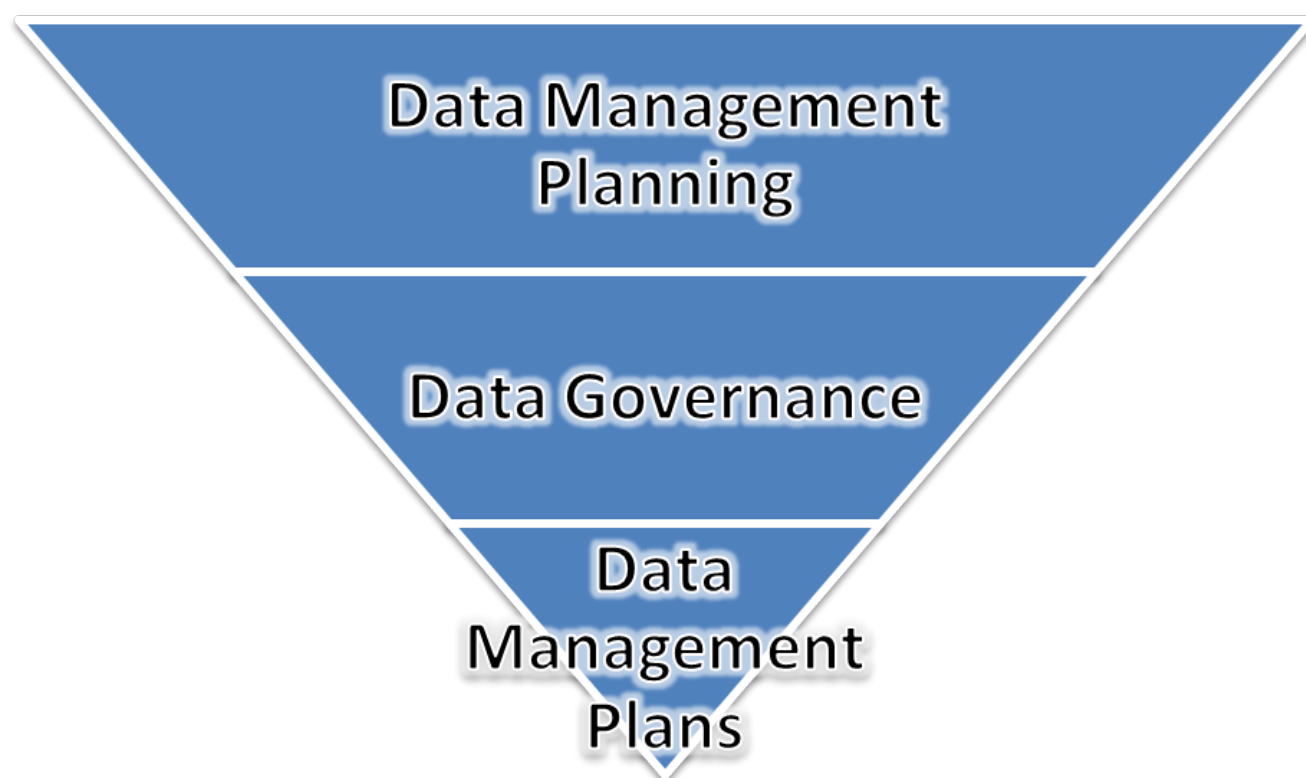
Figure 1. Relationship of data management planning components.

1.1 Data Governance

Data governance encompasses data management strategy at an organizational or agency level. Core elements of data governance include data access, usability, interfacing, storage, and retention. Data governance is overarching and extends beyond the life of any individual project. While data management plans are unique to a given project/task, data governance policies should apply to all data management activities, and as such, data management plans should always follow overarching data governance policies. Additional information on data governance can be found in the ITRC fact sheets on Data Governance; Data Access, Sharing, and Security; and Data Storage, Documentation, and Discovery.

1.2 Data Management Plans

A data management plan (DMP) is a project-level or task-specific document outlining all aspects of data management for a given project or task, with the focus being on current objectives of the project. It should describe how data will be managed throughout the data lifecycle, and it should refer to and align with existing data governance policies in place at the organizational level. If no data governance framework exists, the DMP should also address core elements of data governance, such as specifying levels of access rights and defining data retention schedules. DMPs should also refer to and align with other applicable documents (for example, quality assurance project plan [QAPP], and sampling and analysis plan [SAP]). The DMP should describe the relationship between these documents and their relevance to data management but

should not necessarily replicate content already contained within these other documents.

In 2017, the International Conference on Environmental Data Management (ICEDM) Best Management Practices (BMP) Group authored a white paper on DMPs and created an associated template specific to environmental data management considerations. This white paper explains why a DMP is important and also guides the reader in writing their own DMP. These resources can be accessed from the ICEDM website (http://www.icedm.net/icedm-bmp-group).

Several federal agencies also have published guidance on DMPs, including but not limited to the National Oceanic and Atmospheric Administration (NOAA), the U.S. Geological Survey (USGS), the U.S. Environmental Protection Agency (USEPA), the U.S. Department of Energy (USDOE), and the U.S. Department of Agriculture (USDA).

- The NOAA Data Management Planning Procedural Directive was approved by NOAA's Environmental Data Management Committee (EDMC) on February 11, 2015, as version 2.0.1. While the directive is intended for NOAA programs that produce environmental data or that commission the production of environmental data through contracts, the concepts described are generally applicable to broader environmental data management activities. An example DMP template is included in Appendix A of the directive (https://nosc.noaa.gov/EDMC/PD.DMP.php).
- The USGS maintains a webpage within the Data Management portion of their website dedicated to DMPs. The webpage includes a checklist describing items to consider when developing a DMP for new USGS projects, DMP best practices, several templates and examples of DMPs, and tools to assist with DMP development. Resources are directed toward those completing USGS or other publicly funded science center projects, but the concepts are applicable to broader environmental data management activities (https://www.usgs.gov/products/data-and-tools/data-management/data-management-plans).
- As part of USEPA's Environmental Sampling and Analytical Methods (ESAM) Program, USEPA has published brief guidance summarizing the key elements of a data management framework. USEPA defines a data management framework as a plan developed to support the data management process, including the individual tools, technologies, and processes used to plan, collect, store, retrieve, visualize, and distribute data. As such, the guidance is applicable to development of DMPs as well (https://www.epa.gov/esam/data-management).
- The USDOE Office of Science has published a "Statement on Digital Data Management" focused on data management throughout the data lifecycle, with a particular focus on sharing and preservation. The statement incorporates input from the research communities and the public, including all six Office of Science Federal Advisory Committees on the dissemination of research results. The statement includes requirements for DMPs and various additional guidance on digital data management. Of particular relevance, the statement also includes a section on "Suggested Elements for a Data Management Plan," which may serve as a useful guide when developing a DMP for broader environmental data management activities (https://science.osti.gov/Funding-Opportunities/Digital-Data-Management).
- The USDA National Agricultural Library provides guidance on data management as outlined in P&P 630: Data Management & Public Access Requirements for Agricultural Research Service. This guidance incorporates U.S. federal public access and open data directives and complies with a range of funding agency requirements for DMPs. The guidance published by USDA includes a description of the core DMP components and an example annotated DMP (https://www.nal.usda.gov/main/data/data-management-plan-guidance).

2 DATA LIFECYCLE

The data lifecycle is a conceptual path describing how data are created, stored, maintained, and retained. Ideally, policies and plans related to data management should be in place before the data lifecycle process begins or should be developed during the plan phase of the lifecycle. Data management planning (including data governance and DMPs) should consider and address all stages of the data lifecycle. The stages of the data lifecycle include: 1) plan, 2) acquire, 3) process and maintain, 4) publish and share, and 5) retain. Organizations should address the requirements of each stage of the data lifecycle while maintaining documentation, quality, and security. For additional information on the data lifecycle, see the Data Lifecycle Fact Sheet.

3 PROJECT/DATA QUALITY OBJECTIVES

Data quality objectives (DQO) establish the type, quantity, and quality of data needed to reach defensible decisions or make credible estimates (USEPA 2006). Data management activities should support DQOs. As such, data management planning must take relevant DQOs into consideration. For some projects, DQOs are formally developed in various planning documents, like QAPPs, where data quality indicators of precision, accuracy/bias, representativeness, comparability, completeness, and sensitivity (PARCCS) are well defined. Less formally, DQOs might be established in various data management standard operating procedures (SOPs) or documented best practices. For additional information on DQOs, see the Data Quality Overview Fact Sheet.

In February 2006, EPA published EPA QA/G-4 Guidance on Systematic Planning Using the Data Quality Objectives Process, which provides a standard working tool for project managers and planners to develop DQOs (USEPA 2006). The seven steps outlined in this guidance for developing DQOs are:

1) state the problem

2) identify the goals of the study

3) identify the information inputs

4) define the boundaries of the study

5) develop the analytic approach

6) specify performance or acceptance criteria

7) develop the plan for obtaining data

For additional information on DQOs, see USEPA's guidance and the ITRC Environmental Data Management (EDM) Data Quality Overview Fact Sheet.
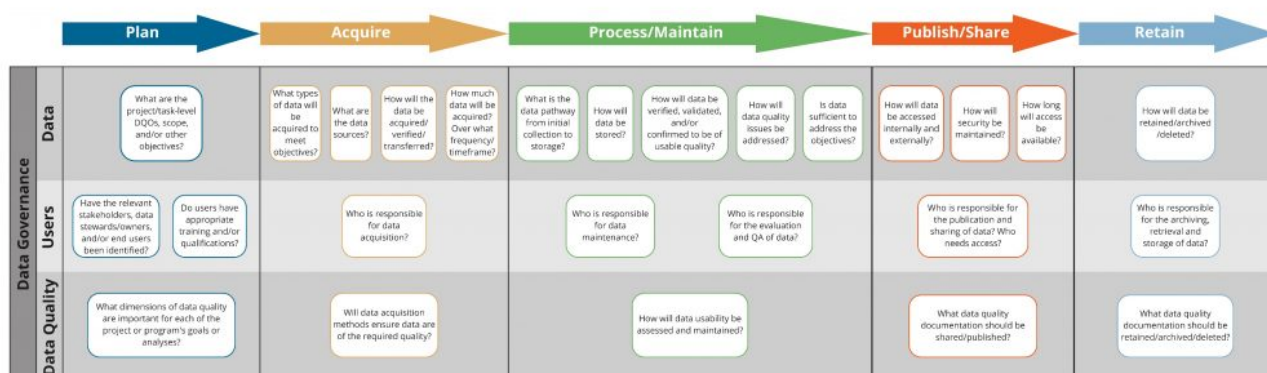
4 USERS AND STAKEHOLDERS

Data users include all people interacting with the data at any point in the data lifecycle, including those who need to access or share the data. Stakeholders are any entities who have an interest in a data set, such as the organization maintaining the data set and internal/external users. Data users and stakeholders should be identified by conducting a stakeholder analysis during the planning phase of the data lifecycle and should be considered when developing data management planning documents and policies. When identifying data users, required training and qualifications of users should also be considered. For additional information on data users and stakeholders, see the Data Access, Sharing, and Security Fact Sheet.

5 DATA MANAGEMENT PLANNING TOOL

The Data Management Planning Tool shown in Figure 2 is intended for use at the beginning of a data acquisition project to develop a DMP. The tool does not cover all aspects of data management planning but is meant to be a guide to major concepts and considerations, which should be ultimately incorporated into a DMP.

The tool's structure is based on the steps of the data lifecycle (see the Data Lifecycle Fact Sheet for additional information), which are shown across the top of the diagram and defined as columns across the diagram. Vertically the diagram is broken out into rows for the categories of data, users, and data quality. The data row includes questions that should be considered in your DMP as your data move through the data lifecycle. Questions in the user's row are related to who will be responsible for data management activities throughout the data lifecycle. Questions in the data quality row are related to how data quality will be maintained throughout the lifecycle.

Figure 2. ITRC Environmental Data Management Planning Tool.

Data governance does not have a distinct row because governance policies shape all data collection and management activities. Additional information on data governance topics can be found in the following five fact sheets:

- Data Governance Fact Sheet
- Data Lifecycle Fact Sheet
- Data Access, Sharing, and Security Fact Sheet
- Data Storage, Documentation, and Discovery Fact Sheet
- Disaster Data Recovery Fact Sheet

It is important to note that while the tool is structured around the data lifecycle, the tool itself is meant to be used at the *start* of a project before data acquisition has occurred. The questions and topics shown on the diagram associated with each step of the lifecycle should be considered during the planning phase of your project so that your DMP comprehensively considers all data management activities that will occur during the data lifecycle. Alternatively, if a project is already underway, the tool can be used to plan data management related to a specific phase of the data lifecycle or to revise already existing plans, policies, and/or procedures with the data lifecycle in mind.

6 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.