

The Environmental Data Management Plan (EDMP)

Why do we need one and how do we get there?

If you have ever worked on an environmental project involving data collection, you have probably found yourself in need of a dataset— and after finally discovering who had it and where, you found out it existed only in a spreadsheet or on a CD in someone’s desk drawer. Our need for using or reporting data raises many questions: Is this all there is? Has the data been changed? What is the quality of the data? Who owns the data?

This paper will explain why an Environmental Data Management Plan (EDMP) is important for avoiding those issues and will guide you in writing your own EDMP. The process of writing your own EDMP will highlight potential workflow issues, relative importance of different datasets, and communication gaps between data users, data generators, and data managers. The EDMP can guide ongoing and future work, and its development with consideration to all aspects of environmental data management throughout a project’s lifecycle is a benefit in and of itself. Also, the EDMP highlights that the data have intrinsic value beyond their initial purpose. This paper is presented in two parts: a paper describing why you need an EDMP, and an outline that guides the creation of your own EDMP.

Environmental Data Management Plan: Purpose and Necessity

Introduction

An Environmental Data Management Plan (EDMP) governs the flow, storage, use, and archiving of environmental data gathered during a project. The EDMP serves clients, data owners, data users, and stakeholders who need to access data to meet organizational, programmatic, or regulatory requirements.

The EDMP:

- **Defines** standards for data structure, content, format, and data-management practices and policies;
- **Ensures** that data-management practices support project Data Quality Objectives (DQOs) and defensible decision making;
- **Improves** performance, accuracy, and efficiency;

- **Supports** effective dissemination of the data by clearly defining the owners and recipients of the data and how they will use the data.

In addition, the acquisition of scientifically and legally defensible data may dictate the need for an EDMP. By implementing an EDMP, data owners, users, and stakeholders are assured that their data are available, accessible, and of known quality. Also, the EDMP may be mandated by program or regulatory requirements, considered a best practice by an organization, or used on a project-specific basis.

An EDMP is most effective if all data owners, users, and stakeholders are aware of its existence and importance. Buy-in from all members of a project team—including the user with the CD in the desk drawer—is imperative. The reasons behind (i.e. “why”) and the specific rationale for each element of the EDMP content are discussed in the paragraphs below.

Defines Standard Structure, Standard Content, and Format

The EDMP is important because it defines the criteria for the standard structure, content and format of the data for a given project. Environmental data should be normalized and organized within a relational database management system (RDBMS) to reduce redundancy, improve data integrity, and facilitate reporting. Specifically, the EDMP:

Defines the standard structure for the data

The EDMP defines the database schema—the data tables, relationships, and referential integrity requirements between the tables. Commercial Off the Shelf (COTS) software with proprietary schemas can be described in terms of key field relationships, and any customizations that affect the standard schema should be documented in detail. Field definitions and other information pertaining to content requirements are typically documented in the “data dictionary,” which may require vendor-specific, or project-specific documentation. The database schema and the data dictionary provide information on where and how the data is stored. This facilitates access for the duration of the project, and reuse of the data in the longer term. For example, standard reports or database views can be created and reused at a project or organization level, with clear understanding of the query definitions required (the schema), and the criteria selections available (data dictionary content and valid values). Detailed knowledge of the standard structure also allows reports to be exported from the database to other applications.

Defines standard content and format

The EDMP defines content requirements for the data within the standard structure. For example, valid values for standard units (date, time, concentration, etc.) and standard terms (chemical name, method, etc.) should be documented and maintained in the EDMP. Also, the origin and requirements for metadata, and how the database tables containing metadata, (e.g., calculated values) should be presented. The standard content and formats can be controlled and checked during the importing process, before the data enter the database structure. The EDMP should define the standard formats used for loading data, and, for each standard format, the unique (primary key), required, field length, and data-type field requirements for the incoming data records. Any organizational “business rules,” or conditional checks (e.g., nondetect data must have a reporting limit value populated) should be described, and any

corrective actions identified. After the content is checked, the data are “mapped” to the database, consistent with the “data dictionary.” Why is this important? The data in the database are consistent among data sets and conforming with business requirements. Knowing what your data are, or having a defined standard content and format, improves the accuracy, efficiency, and flexibility of reporting.

Ensures That Data-Management Practices Support Project Data-Quality Objectives (DQOs)

The EDMP is important because it describes vital data-acquisition practices, describes the data-management practices, and provides the other documentation needed to ensure that a project database and associated data-management processes clearly support effective, relevant, and defensible project decision making. Specifically, the EDMP:

Precludes collection and management of nonessential data

It is important for an EDMP to clearly identify the parameters necessary for project decision making along with any precision/accuracy requirements for those parameters. These considerations help define the tools and/or data collection and laboratory methodologies required to meet DQOs. This knowledge helps project managers avoid the acquisition of nonessential or poor-quality data. In turn, this avoidance keeps data managers from having to manage such data, which can clutter the project database. By preventing the collection of nonessential data or poor-quality data (and any need to manage those data), data-management practices can focus on content that most cost-effectively contributes to essential and foreseeable project decision making.

Supports navigation and access by data users

An effective EDMP provides data users with a roadmap of the content of a database, the pedigree of that content, the organization of that content, and a guide for how to store and retrieve the essential data (and supporting documentation) to meet project objectives. An EDMP that defines data-management practices (and the criteria for essential project data)—which facilitate defensible decision making by data users and project stakeholders—increases their confidence in those practices and the outcome.

Supports project/program analysis and decision making

An EDMP is essentially an interdisciplinary work product that defines a data-management process (created in large part by the data manager) to support meeting overall project goals (defined by the project stakeholders and project manager). As data management becomes more complex and specialized; an effective EDMP requires that the data manager and project manager be able to communicate across their respective disciplines to create an EDMP that supports project specific data analysis and decision making. For this communication to occur, data managers may need to consider how to best keep project managers informed about the evolution of data-management practices.

Supports project evolution and/or long term monitoring

Data needs (in terms of essential data) and decision-making criteria can evolve over the life of a project. This evolution can occur, for example, when a project transitions into post-remediation or long-term monitoring. An EDMP must anticipate phases of the project and any associated changes in DQOs that occur as a project moves from phase to phase. If there are phase-specific data-management practices

that need to come into play, the criteria that trigger those changes should be incorporated in the EDMP.

Defines Data-Management Practices and Policies

The EDMP is important because it defines the data-management practices and policies, giving a consistent and defined framework within which to collect, process, and distribute data for the project. Specifically, the EDMP:

Ensures timely data posting, processing, reporting, and data defensibility

Data specifications in an EDMP ensure that data collectors provide data that are readily transferrable into the data management system. Accountability, provided by an EDMP, allows data collectors to know exactly what steps are required for task completion. Likewise, defined workflows provide for task scheduling with reasonable posting deadlines. Defined and repeatable processes ensure that data stewards can readily perform a task.

Once data are processed in a timely manner, they may then be reported to users for interpretation and analysis. Like data-posting and -processing, reporting that is consistent with EDMP standards follows defined workflows that allow for scheduling and determining deadlines. Defensible data requires traceability. An EDMP defines the roles, responsibilities, and standards for each person and data set in the system. In the event a third-party review is necessary, data are easily traced through the defined processes in the EDMP. Traceability and defensibility must be carefully planned for throughout a project's lifespan. An EDMP is just one part of a foundation to ensure defensibility.

Guides data use, dissemination, storage life, and overall best practices

Data users must know the scope and limits of each data set provided to them. This knowledge reduces the possibility of incorrect or misguided analysis. Likewise, data users must know what methods are appropriate for data interpretation and presentation. Cradle-to-grave data-management approaches provide for complete traceability for data sets. These require careful planning of standards to ensure that data posting, processing, use and dissemination are all consistent with EDMP requirements. Likewise, providing detailed timelines for data archiving ensures compliance with the guiding principles. The principles and philosophy defined in the EDMP informs any unforeseen events, corrective action, or new data types.

Defines Data-Management Roles and Responsibilities

The EDMP is important because it documents workflow processes that involve people with competing priorities. Any individual or group involved in the generation, documentation, processing, updating, or transfer of data should be explicitly defined, along with their respective responsibilities. Typical roles defined in the EDMP may include:

- *Data Manager/Steward* - People responsible for supervising the flow of data throughout the management process, typically from pre-field planning through to reporting and archiving.
- *Quality Control Manager* - People responsible for ensuring that quality objectives are met in accordance with the QAPP. Responsibilities may include setting EDD content and functional rules, performing or supervising QC activities, and documenting all QC procedures, activities, and

results.

- *Data Management Systems Administrator/Custodian* - People responsible for the daily maintenance and function of the environmental data storage system and for knowledge of data workflow processes into and out of the system. Responsibilities may include providing technical assistance and processing, updating, or transferring data to internal and external data users, managers, and stakeholders.

That the EDMP defines these roles is vital to ensure accountability, streamline data-management tasks, and eliminate redundancy and/or duplication of effort. To ensure accountability, a complete understanding of who is responsible for each step in the workflow is necessary to ensure that sources of error or breaks in procedures are quickly identified and rectified. Data-management tasks can be streamlined by roles that are clearly defined and documented and that can be carried out with maximum efficiency and consistency. An understanding of the relationship between roles, responsibilities, and the workflow can lead to opportunities for automation and other efficiency gains. Also, redundancy and/or duplication of effort can be eliminated by clearly defining roles in the EDMP. Data-management personnel often wear multiple hats, including providing QC and reporting services to users on an as needed basis. Clearly assigned roles reduce or eliminate the potential for duplication across tasks and allows users to quickly identify their support staff.

Improves Performance, Accuracy, and Efficiency

The EDMP is important because it improves performance, accuracy, and efficiency, and ultimately reduces cost for the data owners, managers, and stakeholders. Specifically, the EDMP:

Reduces data reformatting and Eliminates Inconsistencies, Errors, and Omissions

Collecting, loading, and maintaining data using the processes specified by an EDMP reduces the need for subsequent data reformatting and minimizes inconsistencies, errors, and omissions. Clearly defined and repeatable processes discourage random or 'one-off' processing events, during which steps can be missed, and errors can occur. Historical data usually requires manual or automated processing in order to be merged with an existing project dataset, and is a primary source of error propagation. However, if incorporated in accordance with an EDMP, and audited as per [LINK TO HISTO PAPER](#), the data becomes accessible to users and of documented quality. Providing traceable processes, whether automated or manual, allows for faster error identification, resolution, and corrective action.

Reduces institutional data loss with loss of personnel

The EDMP reduces the dependence on institutional knowledge as the flow, structure, and use of the data is documented and maintained during the project. The loss or gain of project personnel should not negatively affect the management of the data beyond workload balancing and task-reassignments.

Increases data use effectiveness and efficiency and ensures reproducible results

Clearly defined reporting processes in an EDMP provide guidance and help to users who need to retrieve data in a consistent and repeatable format. Users are more efficient when common or standardized query criteria are well documented and available, and output is both consistent with expectations, and

comparable with previous and future reporting activities. One of the most frustrating reporting aspects for users is in understanding the effect of apparently minor query alterations to the reproducibility of results. For example, the assumption of analytical method name in a query for Volatile Organics may exclude methods and therefore analytes for certain laboratories where a variation on the method name is present. This seemingly minor choice on the part of the user can make a significant difference in terms of the data retrieved for analysis. The EDMP can maximize reproducibility of results by defining standard query terms and processes.

Maximizes project value/benefit of data management process

Consistent, defined data-management workflows ensure that fast, repeatable processes are achieved from data creation to storage, and opportunities for automation are maximized. In this manner, project data-management budgets are predictable and accountable. Likewise, the cost-benefit ratio is defined and consistent, improving the ability for project managers to plan, direct, and benefit from the data-management process.

Supports Effective Dissemination of Data

The EDMP is important because it provides the information needed by those who have not participated in the design and implementation of the EDMS to navigate the EDMS and to retrieve data in a reproducible manner.

Serves as database "roadmap" for other prospective users

Staff turnover and stakeholder changes are likely, particularly over long periods of time and for large projects for which a more complex environmental database may be needed. An effective EDMP serves as a roadmap for current and future users regarding database structure, content, relationships, roles and responsibilities, rules, processes, and supporting documentation. Being able to clearly and efficiently orient new users with an effective EDMP is essential if those users (and the stakeholders they represent) are to get the maximum benefit and value from an existing database.

Contributes to perception of transparency by any affected parties

When a database is used to support data-driven or data-supported decision making, an effective EDMP makes that database clear and accessible to parties on all sides of any issues. Such clarity and accessibility provides all parties with the potential to repeat analyses and recreate data-driven decisions that can potentially affect them.

Can protect restricted or sensitive data (such as personal information)

Access to specific information contained in a database (e.g., municipal water-supply well locations, operational details, property-ownership details, personal information) may need to be limited because of potential risks, security concerns, and identity-theft concerns. Managing restricted or sensitive data so that they are a modular component in a database will facilitate controlling access. Clearly identifying and managing restricted and sensitive data and having well-defined and effective rules in an EDMP that spell out who can have access to those data (and the steps they need to take to gain that access), can provide this protection.

Broadens access to and increases value of data

Environmental data have intrinsic and lasting value. An EDMP that clearly and effectively describes an appropriately well-constructed and well-documented database adds value to a project. An EDMP is valuable to all project stakeholders by making relevant and defensible data readily accessible to all for tracking environmental conditions, for analysis, and for decision making. Recognizing that environmental conditions, concerns, and regulations are dynamic and subject to change over time; a well-constructed and well-documented (in an EDMP) database of relevant information can also have lasting value to future stakeholders.

Ensures Data Meet Quality, Defensibility, and Regulatory Requirements (Requirements and Guidance)

Accurate and well-documented data increase legal and scientific defensibility. Such data may reduce the potential for and cost of litigation and appeals involving the agencies and organizations that provide funding for projects that rely on data, data statistics, and data analysis. Consequently, these agencies and organizations typically require submission of a data-management plan for project funding. They want to know what data will be created, what policies apply to the data, who will own and access the data, what data-management practices will be used, what tools will be used to store data, and who will be responsible for each activity.

Additionally, as a mechanism to provide transparency for data used in funded research and analysis, the U.S. Office of Management and Budget (OMB) in 1999 amended Circular A-110 to require federal awarding agencies to ensure that all data produced under an award will be made available to the public through the procedures established under the Freedom of Information Act' (FOIA). This amendment defines research data as "the recorded factual material commonly accepted in the scientific community as necessary to validate research findings" (University of Montana, 2017).

Most important, data management *is required by law*, including legislation and Executive Orders, such as:

- *Information Quality Act* - USGS Guidelines (Section 515 of The Treasury & General Government Appropriations Act for FY 2001) allows the public to examine and challenge the data disseminated by the USGS and provides review procedures for those challenges.
- *Clinger-Cohen Act* (IT Management Reform Act) established the position of Chief Information Officer to oversee information quality and IT implementation. It mandates that agencies develop enterprise-wide information architectures to improve business performance and data portability. See the USGS Information Technology Strategic Plan: Fiscal Years 2007-2011 [PDF] (USGS, 2006).
- *Privacy Act* establishes a Code of Fair Information Practice that governs the collection, maintenance, use, and dissemination of personally identifiable individual information that is maintained by federal agencies.
- *Government Performance & Results Act* requires agencies to engage in project management tasks such as setting goals, measuring results, and reporting their progress. To comply with GPRA, agencies produce strategic plans including EDMPs.

- *Paperwork Reduction Act* provides the basis for managing information as a resource. It mandates that agencies take steps to improve their data-quality and data-sharing capabilities.
- *Executive Order 12906* (Geospatial Data) directed the Federal Geographic Data Committee (FGDC) to establish a National Spatial Data Infrastructure (NSDI) to acquire, process, store, distribute, and improve use of geospatial data. (USGS, 2017).

Often, funding agencies and organizations that require submission of a data-management plan also provide some level of guidance for preparation and submission of the plan. Examples of agencies and organizations that require a data-management plan for project funding, and provide online guidance for preparation and submission of those data-management plans include:

- Centers for Disease Control and Prevention (CDC);
- National Science Foundation (NSF);
- National Endowment for the Humanities (NEH);
- Institute for Museum and Library Services (IMLS);
- U.S. National Oceanic and Atmospheric Administration (NOAA);
- U.S. Department of Energy (DOE);
- U.S. Environmental Protection Agency (EPA);
- U.S. Geological Survey (USGS);
- U.S. Department of Agriculture (USDA);
- U.S. Department of Education (ED);
- U.S. Department of Defense (DOD);
- U.S. National Aeronautics and Space Administration (NASA);

Table 1 lists additional sources for standards or guidance in the preparation of data-management plans, including website links and brief descriptive comments.

Highlights the understanding that data have intrinsic value, beyond their initial purpose.

An EDMP is important because it highlights what data users, managers, and stakeholders inherently know: data have intrinsic value beyond their initial purpose, and should be treated as an asset. Well-documented data can be copied and reused without loss in value. Reusing the same dataset for different purposes within the allowed boundaries of ‘intended use’ maximizes the value of that data and return on investment for the data owner. The reuse of data in other studies (e.g., meta analysis), or in longitudinal projects is reduced or eliminated in the absence of appropriate data documentation, such as an EDMP.

REFERENCES

University of Montana, 2017: Online at: <https://www.umt.edu/research/ORSP/propdev/datamgmt.php>.

US Geological Survey (USGS), 2017: Online at: <https://www2.usgs.gov/datamanagement/why-dm/dmimportance.php>.

USGS, 2006: Information Technology Strategic Plan: Fiscal Years 2007-2011, Circular 1304. Online at: <https://pubs.usgs.gov/circ/2006/1304/pdf/Circular%201304.Frondorf.pdf>