Environmental Data Management Plan (EDMP): Template and Content Guidelines

Environmental Data Management Plan (EDMP): Template and Content Guidelines

Date Created	[date]	[authors]
Date Revised	[date]	[notes]
Revision History	[date]	[notes]

1 Introduction

1.1 Purpose and Scope of the EDMP

This section should introduce the EDMP and note any key limitations in terms of scope to lead the reader to the appropriate section. For example, if the project is highly confidential with limitations to data use, note the limitation in this section and direct the reader to more information. Note also the scope in terms of period covered by the EDMP, and reference related documentation if not discussed in the sections below.

1.2 Data Management Standards and Principles

The Data Management System (DMS) relies upon a consistent and logical framework to be laid out in

the Environmental Data Management Plan (EDMP). Using a consistent framework allows growth and flexibility of the DMS. This framework provides a broadly similar, defined, and repeatable process that supports all data types. Defined logical workflows ensure that a process is traceable and repeatable from data acquisition to data reporting. Maintaining consistent and logical workflows allows the DMS to provide project and organization managers with efficient access to accurate information. Defining overarching standards and principles helps guide the remainder of process of creating the EDMP.

1.3 Relation to Quality Assurance Systems

This section should note the location and direct relationship of existing or planned documents or systems that relate to the EDMP. Particularly, the EDMP and EDMS are subordinate systems to a Quality Management Plan (QMP) and Quality Management System (QMS), respectively. Defining the relations of these systems will provide framework for creating and administering the EDMP.

1.3.1 Data Quality Objectives (DQO)

As for any process, setting goals allows users to identify successes and needs for improvement. Data Quality Objectives (DQO) guide the refinement of macro and micro processes. An overview of company-wide data quality objectives will establish a principled approach to project-level objectives. Subsequent project-specific data quality objectives should be outlined in sufficient detail to improve project-level interpretation and decision making.

Note: If guidance is needed, the U.S. Environmental Protection Agency (EPA) has developed a guidance document for the DQO Process to determine the type, quantity, and quality of data needed to support decision making. The guidance is the culmination of experiences in applying DQOs in different Program Offices at the EPA. (Guidance on Systematic Planning Using the Data Quality Objectives Process, EPA QA/G-4, online at: https://www.epa.gov/sites/production/files/2015-06/documents/g4-final.pdf).

1.3.2 Relationship to QAPP, QMP, Sampling and Analysis Plan

Note the location and direct relationship to the Quality Assurance Project Plan (QAPP), Quality Management Plan (QMP), and Sampling/Analysis Plan (S&AP). Any unifying or divergent aspects of the project related to these systems should be explicitly defined here. For example, analytical data will normally be associated with a QAPP, but exceptions to established laboratory quality processes may be assessed on specific datasets. Clearly identifying these datasets or data types will reduce confusion for data management personnel and users.

1.4 Review and Revisions

The EDMP is a 'living document' subject to change during the life of a project, and even thereafter if data is reused in subsequent studies. Traceability in the data processes requires traceability in revisions

to the DMP. Thus, it is important to define schedules and scopes to EDMP revisions. Whenever a new data type or workflow is added to the DMS, the EDMP should be revised. Any revisions to the EDMP should be logged in a corresponding REVISION HISTORY section at the beginning of the EDMP, and should include the author and date.

1.5 Implementation

Implementing the DMS requires documenting and inventorying of all data for a DMS compatibility review. Any data to be included must adhere to the standards and principles defined in Section 1.2, with guidance provided throughout this document. Any data not to be included should be identified with guidance found in Section 4.2.8.

2 Personnel

2.1 Management and Quality Assurance

Identify key project personnel responsible for project management, direction, and quality assurance, if unrelated to data-management staff. Identification of personnel should include name, role, organization, and contact information.

2.2 Roles and Responsibilities

Establish and document any employee, laboratory, and consultant roles with responsibility for data collection, submission, management, quality control, reporting, and archiving. Clearly defining the function and boundaries of each role will reduce duplication of effort, ensure there are no workflow gaps, and allow for a clear transition of data throughout the workflow. Note points within the workflow where 'hand off' occurs from one party to another. For example, the transition of responsibility from the field staff to the laboratory occurs when a series of samples is submitted along with a Chain of Custody (COC) to the laboratory. The procedure for ensuring that data-management personnel are in possession of the COC after this process should be clearly understood and defined.

2.3 Training

To ensure consistent that data management protocols are followed, training expectations for existing and new employees relating to the DMS should be outlined in this section. Where possible, direct users to required and recommended training resources based on the user role identified and note key personnel for leadership and assistance in designated functions.

Under most circumstances, new employees or new project participants should first familiarize

themselves with the QAPP, QMP, and EDMP to establish a framework to begin operating in the DMS. Any documentation, presentations, or courses provided by relevant software developers whether online, in person, or as documentation, should be considered for training purposes before the new employee begins working in the DMS. Additionally, a demo facility with demo EDDs and data provides an opportunity to become familiar with a company-specific EDD work process without jeopardizing production data quality or confidentiality. The EDMP provides an opportunity to explicitly define training programs like the examples provided above.

Current employees should also be encouraged to attend industry conferences and regularly review documentation, presentations, and courses provided by software developers. Ensuring a dynamic and efficient DMS requires on-going training to acquire new tools and processes. Scheduling new and refresher training in the EDMP provides consistently skilled data managers and stewards, and increases the quality of the DMS.

3 Data Management Systems

3.1 Physical Data

Physical data may comprise field forms, logs, printed photographs, hand-drawn or hand-notated maps and diagrams, stored samples, and preparations from samples. These materials need to be stored in a manner compatible with the security and preservation mechanisms described elsewhere in this document. Descriptions of these materials (i.e., their metadata) must be stored with them, along with records of data provenance (who collected and who took possession of the physical data). This section should be used to describe the procedures that will be used to satisfy these objectives.

3.2 Digital Systems

Use this section to describe the Information system or hardware and software used to manage the data. Clearly document the DMS vendor, version, platform, underlying technologies (e.g., SQL Server 2012, Microsoft Windows Server 2008), infrastructure (e.g., SaaS Cloud, on-premises, or in a data center), and licensing details in the appropriate sections below. The change-management procedures associated with these systems are described in section 3.3.

3.2.1 Hardware

A description of the hardware components of the data-management system should include servers, storage devices, network infrastructure and firewalls (intrusion protection), power supplies, and internetworking connectivity. If hardware components are off-site (i.e., data center or cloud), insert details of the off-site infrastructure here or in an appendix. Include emergency contact information and any contractual support information.

3.2.2 Software

3.2.2.1 Licensed Software

Describe the operating systems, database server applications, and web-server applications that support the DMS. Other applications that support components of data storage and processing, such as spatial-data management systems, image analysis, field-data collection, and other data-processing tasks should also be described. Licensing information should be included and explicitly described in terms of user requirements and terms. Include a contact from within the EDMP organization and the software vendor for licensing queries or resolution. If licensing explicitly limits the sharing of proprietary database information such as data models or report code, clearly document those limitations here, and repeat in sections associated with related DMS functions.

3.2.2.2 Developed Software

Custom applications may be developed to support project data management if publicly available (either commercial or open-source) systems do not meet the project's requirements or are not viable financially, among other reasons. This could include spreadsheet formulas and macros, or whole applications based on one or more programming languages. These applications should be sufficiently documented so that users understand how they are used in the DMS, and what, if any, ramifications there are for the managed data. This section should describe procedures that are used to manage, test, and document custom-developed applications. Include appropriate development information in an appendix, such as requirements, analysis documentation, functional analysis, and testing logs.

3.3 Change Management of Software and Hardware

Configuration management defines the procedures that will govern the maintenance of systems over time. Changes made to computer processors, firmware, operating systems, database-management software or geodatabase software, universal resource locators, passwords, and storage systems may affect the integrity, numeric processing (calculations), and accessibility to the data. A plan to document and control such changes should be developed and documented in this section. The choice of appropriate data-management systems (both hardware and software) will affect and be affected by the amount and diversity of project data, and requirements for data access.

Project datasets and the systems that contain them must exist in an accessible form for long periods. It can be expected that, over time, software, hardware, and networking infrastructures will change. Planning for such change should begin at the outset of a project, and should be documented in this section along with any changes as they occur. The choice of systems and software should consider the development path of those systems. If data are put into a system that will soon become obsolete or no will longer supported by its manufacturer, which will present potential issues for accessibility and data

migration in the future. Open source systems can present similar issues, for example because of stalled development or changes in development paths. A solution suitable for a particular project will be influenced by project needs, and documentation of the state of the information systems will support future upgrades when those needs change.

3.4 Archival Systems and Formats

All DMSs require regular and consistent backup processes. Backup schedules and procedures may be determined by the project requirements or may be related to the recommendations of the software vendor. In either case, this section should document the content, timing, and location of any and all backup and archival information, including reference to expiration and obsolescence.

3.5 Age of Data and Systems

EDMP generally assumes that all data will be treated as useful and beneficial. In some rare cases, data may be deemed no longer useful or beneficial. Since this is a departure from the EDMS, the method of exit should be specified for each occurrence. Use this section to document or reference any 'end of life' policies implemented by the EDMP organization or EDMS vendor.

3.6 Data Dictionary

Provide the "data dictionary," which describes the contents of each field comprising each data table (i.e., where certain types of data should reside). The data dictionary provides titles of each data field, the relationships between fields or objects, the type of data (such as text or image or binary value), possible predefined values, and a brief text description.

4 Data Management Procedures

Procedures for the management of project data should be described in this chapter. Some aspects of the documentation of provenance, methods, data quality, and accuracy may be governed by regulatory frameworks, legal requirements, or professional best practices associated with a project. Governing requirements should be addressed in the introduction. This chapter addresses both the tracking systems and the handling of various data types. Tracking systems are those systems and procedures that are used for documenting the ownership, custody, workflow for data transmitted from one party to another, and change management. The mechanisms, formats, processing, and quality assurance will be described by data type. Document security and retention can be described in greater detail in their respective chapters within the EDMP.

4.1 Documentation or Tracking Systems

Documentation is the key to creating a defensible dataset, effectively sharing data with others, and to facilitating reuse beyond the initial intended purpose. Tracking systems may be implemented separately for different management processes, however, access to these systems should be described and facilitated via the documentation in this section.

4.1.1 Origin

Describe the source of the data to be managed during the project. For example, if the data originates from another consultant, provide details as to the originating DMS, organization, and transfer processes. Clearly documenting the origin of the datasets may facilitate reuse in cases of questionable quality. For example, an analytical table that is noted as originating at a named laboratory is more readily consumed than an analytical table of questionable origin, provenance, and interim processing.

4.1.2 Ownership

Any and all ownership information should be clearly defined in this section. Data collected in the field on behalf of a client is typically owned by that client. Any interpretive data or code may be the intellectual property of the consultant or contractor, depending on the situation. Clearly defining ownership within the plan will help resolve conflict at project close, and may avoid confusion with respect to data access and sharing.

4.1.3 Users

Define users and roles in terms of DMS function and in terms of software and hardware access and use. See section 2 (Personnel) of this document and relevant sections in the *'VALID VALUES' paper - reference.*

4.1.4 Revisions and change management

Define the methodology for determining, implementing, and capturing change in the DMS and associated workflow processes. Document any workflow changes in the EDMP.

4.1.5 Generalized Data Workflow

Create a workflow document or diagram (e.g., swim-lane diagram) that identifies the task assignments and goals at each step of the data workflow. An example of a general workflow is attached; however, most projects or internal EDMS systems will include more complicated relationships. Wherever possible, include role definitions and highlight key tasks within the workflow.

4.2 Data Types

4.2.1 Hardcopy and Data Collection

Describe the handling and storage of hardcopy documents such as field notebooks and handwritten logs. Describe any procedures for cataloging documents and recording their source. Describe procedures for document retrieval, reproduction, and for determining access permissions. This subsection should describe procedures for the archival preservation of hardcopy documents.

4.2.2 Laboratory

Laboratories may provide certifications, certificates of authority, standard operation procedures for preparation and analysis methods, sample receipts, analytical reports, and electronic data files (also known as electronic data deliverables (EDDs)). This subsection is used to describe how such documents will be handled. For more information on laboratory EDDs see the *Laboratory EDD Whitepaper.*

4.2.3 Field Instrumentation and Sensors

This subsection addresses the documentation of specific instrumentation, calibration, maintenance, deployment, and demobilization. Such information should be recorded for each individual instrument in use during the project, using serial numbers or other asset-tracking identifiers. This subsection is not intended to be used for describing calibration, installation, and maintenance procedures themselves.

Field-instrument data are recorded measurements of physical or chemical characteristics of air, water, or soil, collected and measured at the site being characterized. The collection and measurement is performed with an instrument specifically intended for the type of media and constituent or parameter being collected.

Examples of field instrument physical data include:

- Air
    - Airflow
    - Temperature, humidity, and pressure
- Water
    - Water flow
    - Temperature, pH, dissolved oxygen
    - Oxidation/reduction potential, conductivity, salinity
- Soil
    - Bulk density, moisture, compaction
    - Elasticity
    - Porosity

In addition to physical data, advanced analytical field instruments are sometimes used to perform on-

site collection and analysis of chemical or physical constituents, to assess air quality, water quality, or soil quality.

Field instrument data can be recorded by hand-written notes in a field log book, an on-site electronic tablet or laptop computer program that uploads the hand-inputted data directly to a database, or directly by the field instrument, for download later at the office.

4.2.4 Spatial Data

Spatial data can include physical and political boundaries, locations, topographic or bathymetric features, structures, and georeferenced imagery. These features may be sourced from non-project repositories or through generation during the project. This subsection is used to describe the requirements and mechanisms for recording the source of spatial data, developing metadata, and specifications for transformations and projections. This subsection should describe requirements and the recording of the accuracy for each spatial data product.

Spatial data are any type of data distributed over a given two-dimensional area or three-dimensional volume. For the purpose of assessment, the location of each datum may be referred to as a point, node, or position, to name a few. The database for assessment of spatial data must include the x, y, and, as applicable, z coordinates, in a defined, standard coordinate system.

Spatial data may be analyzed at a specific point in time, or at various points in time over a given time frame (spatiotemporal analysis). Spatial data analysis often involves how data varies (or doesn't vary) in magnitude from one area to another (distribution analysis), such as estimation of heterogeneity.

4.2.5 Photographic and Videographic

It is assumed that hardcopy photographs and film media will be addressed in the hardcopy document subsection. This subsection describes the file-naming system, logging, formats, and storage of electronic imagery. This subsection should address the workflow from recording to transfer and storage of imagery. Geolocation and image resolution requirements and recording should also be addressed. If necessary, any calibration materials (e.g., for color correction) should be addressed here. Image processing can be summarized here, referencing other project documentation; however, the tracking of pre-processed and post-processed imagery should be discussed here.

Data in the form of photographs and videos provide information that may not otherwise be obtained from sample population data. Historical aerial photographs, underground pipe and wellbore videos, and contaminant plume aerial photos are just a few examples of types of photographic and videography data. Photos and videos may also be used to corroborate or question conclusions derived from sample population data.

The database for photographs and videos may be organized by date or geographic location, or some other parameter. Accuracy and reliability of aerial photos and videos relies on the visual quality, an

accurate scale of measurement that can be applied to each observation, accurate dates, and (in the case of video used for time-lapse analyses) accurate knowledge of real time-lapse from point to point.

4.2.6 Data Loading and Storage

Data loading and storage should follow documented guidelines to ensure consistent transfer of information from EDD to EDMS. If a controlled loading environment (like Earthsoft's EQuIS ™ EDP) is to be used, format files and mapping documents must be provided and maintained in a manner consistent with policies within. Uncontrolled environments (direct importing to tables) may provide inconsistently traced data. As such, any data loading in an uncontrolled environment must be thoroughly documented in this section.

4.2.7 Data Verification and Validation

Verification and validation refer to distinctly different processes. Upon data loading, data is verified by data owners to ensure a loaded data set accurately reproduces the provided data set. Defining a process will ensure consistent data verification before use.

4.2.8 Other data

Information generated during a project that does not fit into the categories above may still have relevance and value to the long term reuse of the data as a whole. Interpretations and summations are usually managed separately within a document management system or form part of a final report for the project, leading to a separation of this key information from the source data. The Data Management Plan should include references to such documentation at the least, and at best include path and storage protocols to ensure that this information can be accessible to data consumers.

5 Data Security

This section is used to describe procedures that support the security and integrity of the systems that house project data, and the data themselves. Depending on the client or funding sources, there may be specific minimum requirements for security, confidentiality, or privacy. The source of those requirements or guidance can be described in the Introduction section and reviewed in more detail here. In terms of the long-term integrity and availability of data, archival concepts may be applicable to security and systems planning, applicable to both hardcopy and digital records.

5.1 Access Controls and Authentication

Access controls and authentication are used to protect the integrity of the data and documents and to specifically define who may view, insert, delete, or change data records and documents. Access controls should be described in terms of physical barriers and locking mechanisms, computer and network privileges, the means of authentication, and user change management procedures. These security procedures should address the protection of both physical documents and information contained within computer systems. Roles should be described, such as who may have administrative access (full access

to the systems and documents), who may insert, update or change, and delete information, who may view information, and who may have access to the various functionalities of a system. These roles may be described in general for a given project and system, and then addressed to specific persons or entities that are involved in the project. Change-management in the context of access controls refers to the manner in which new people are added to project-specific roles and people who are no longer involved are removed from access rights.

5.2 File Storage

Part of the data from any environmental program will exist in hardcopy format. In addition, there may be physical media that contain original data transmittals or backup copies of the project database. These materials need to be stored in a secure environment that includes resistance to fire, water, extreme temperatures, or other physical conditions that can degrade or destroy the media.

The storage plan should also address accessibility to the stored materials, describing how access will be controlled and monitored.

Except for archival paper, all data storage media have relatively short life-spans (i.e., on the order of a decade or less). The devices that produce and read computer backup media will also have an inherent life span. The planning process should address how media storage will be compatible with desired or required retention times.

5.3 Privacy and Confidentiality

6 Data Retention and Distribution

6.1 Retention requirements and archiving

Most organizations have an established protocol for retaining information that may be pertinent to regulatory or operational needs. This section of the DMP should include a clear definition of any and all policies affecting the data, including internal (consultant) policies, Client-mandated policies, and regulatory or governing body policies. For example, a regulatory policy relating to the data collected may require a longer retention time than that mandated by the client. A clear understanding of these requirements and their implications to data storage and archival is necessary both to protect future access to the data and to provide legal protection to the originating consultant.

6.1.1 Meta-data

Metadata is key to providing context or supporting information relevant to a managed dataset and may

constitute data in its own right. Examples of metadata include field logs – the original books or transcriptions, laboratory reports, field photographs, and other information used to interpret data. Retention and archiving of metadata should be clearly defined and ideally should match policies relating to analytical and other primary data.

6.1.2 Original documents

This section should outline the location and retention policy surrounding any original documents that are pertinent to the interpretation of data.

6.1.3 Data Products

Need to do some additional research here. What are the common practices for registering data products? How to address cloud storage, archival, backups?

6.2 Reporting and End-Users

6.2.1 Quality Control Data

The application of quality control (QC) processes that affect how data are reported can be described here. The DMP should reference project quality assurance (QA) and QC processes if these are documented elsewhere, or describe QA and QC procedures that will be used to assess the accuracy and precision of data. Field or laboratory data may have qualifiers applied to them after quality-control assessments have been performed. This section should reference or describe such qualifications and how qualified data are intended to be used for assessment and interpretation.

6.2.2 Data Transformations and Calculations

Data may have transformations applied to them after they are collected in the field or reported from an analytical laboratory. This application of transformations, which may occur within the database or upon export from the database, should be described here. This section should be used to describe the procedures used to calculate derived values stored in the database or as used in the reporting of raw data after export from the database. Examples of data treatments and calculations include the averaging of replicate measures from the same sample or from replicate samples, the treatment of non-detected values for assessments and statistics and in the calculations of the sums of chemical concentrations within a chemical group. Other situations that involve transformations or calculations include moving averages or filtering of high-frequency time-series data, such as from continuously recording instruments or the calculation of equilibrium and toxicity values based on observed values.

6.2.3 Documentation for Distribution

The workflows, methods, and relationships of sampling events, locations, samples, field observations, and analytical results (chemistry or other laboratory-based measurements) should be provided in other project documents, such as sampling and analysis plans. How these methods and relationships are indicated in the database and in data exports should be described here so that a data user will have sufficient information about the datasets to perform data analysis and interpretation. This section should include a data dictionary that provides the structure of data and that describes any codes or identifiers used in the data. This documentation should include a guide for how a user can navigate the database, particularly if a relational database is provided as an export, or if a web-based user interface is provided to filter the database for various data products.

A change log or read-me file that is included in data distribution can document any changes to records in the database or notes on specific issues with data records or datasets.

6.3 Use and Access Metrics

Recording who requests and receives data or how a web-based data portal is used to either request or display various datasets can be useful for determining the actual use of the data. Recording usage and receiving user feedback may suggest priorities for improving accessibility or for developing new data products. Usage metrics may be used to evaluate the need for long-term support for access to the data.

6.4 Long-term Data Change Management

7 References

8 Glossary