

A Data Management Plan Framework

Stanley Smith¹, Steven Tessler^{2,3}, Michael McHale³

January 13, 2011

U.S. Geological Survey

¹Alaska Science Center, ²New Jersey Water Science Center, ³Climate Effects Network

In January of 2010 the authors began a team effort to develop and implement a data management plan (DMP) that could serve both the Climate Effects Network (CEN) and the Alaska Science Center (ASC). The ASC had already embarked on a state-wide, multi-agency effort to integrate their climate-related information (Alaska Climate Change Executive Roundtable – ACCER), and that effort would serve as a model and initial surrogate for the CEN effort to come.

Throughout the year the team met regularly to discuss the general principles of data management as they pertain to scientific research organizations. The team also evaluated data management plans, practices, and related documentation from other federal agencies, research organizations, and academia. In addition, visits were made to a number of USGS offices and other organizations to discuss and examine their “best practices” with regard to large-scale, long-term, data management.

As the year progressed, we created three documents that captured our synthesis and conclusions regarding research data management. These concepts were formulated in the organization of a generic DMP framework (DMPf) for research organizations. We were influenced by the experiences and excellent work of our colleagues, new contacts, and the literature, and we ultimately proposed a very general plan for how to proceed toward our goal of producing DMPs for both organizations while facilitating data integration.

This paper combines the three original, separate documents into one package to make them more convenient to share with others and edit. The documents describe: (1) concepts for a generic data management plan ‘framework’ that recognizes and respects agency and program policies and procedures, and the importance of inheritance of standards within an organization; (2) the scope of the major DMPf sections to be written and who should be charged with the responsibility of writing them; and, (3) a starting point table-of-contents (ToC) for the research organization DMPs that respects the concepts and scope defined by the previous two documents. We consider this an open process and invite all discussion and comment.

1. Constructing a Data Management Plan Framework

This document described three overarching concepts that improve the language and clarity of discussions on scientific data management, and sets the boundaries for a general ‘framework’ for writing DMPs (the DMPf).

2. Scope of DMPf Sections

This document, although written last chronologically, defines the scope and expected authorship participation for writing a generalized DMPf outline that follows a ‘layered’ approach.

3. DMP ToC

This is a detailed outline of a DMPf that follows the concept principles and scope, is presented as a ‘table of contents’, and is intended for USGS research projects and programs.

Constructing a Data Management Plan Framework

Stanley Smith¹, Steven Tessler^{2,3}, Michael McHale³

U.S. Geological Survey

¹Alaska Science Center, ²New Jersey Water Science Center, ³Climate Effects Network

The Alaska Science Center (ASC) and the Climate Effects Network (CEN) share the need to create a Data Management Plan (DMP) that provides guidance for researchers, facilitates data integration between multiple programs, and provides for the long-term preservation and availability of our valuable data assets. Over the past year the ASC and CEN researched current operations, participated in draft writings, and held numerous discussions to identify the ‘best practices’ in data management for research organizations. This proposed concept for developing a DMP framework (DMPf) is the result of that work. The intent of this document is to describe a framework for data management plans that will accomplish all the stated objectives and additionally be easily adaptable for other programs.

During the past 10 months the ASC-CEN data management team has researched and collected numerous data management documents in print and on the web from various federal agencies, academia, and other organizations both domestically and internationally. The team also made site visits to discuss data management best practices currently in operation at the National Environmental Observing Network (NEON), the National Park Service’s Inventory and Monitoring Program (NPS I&M), the USGS Western Ecology Research Center (WERC) in San Diego, the USGS Fort Collins Science Center, the USGS Earth Resources Observation and Science (EROS) Center, and others. Although many data management practices are shared by all, the approach taken to document these practices has been unique within each organization. Each organization produced DMPs that supported their mission, but none are sufficiently generalized in context, coverage, or document structure to be re-used by other programs without significant modification.

From this beginning and informed by our new partnerships and literature review, we recognized the need for a generalized data management plan framework for research-based organizations. The challenge is to construct a basic framework for DMPs that could be customized for specific programs or projects, while at the same time accommodating the data management and data integration goals of the organization. This basic framework must incorporate the best practices used by the USGS and similar organizations, and be designed to be easily adaptable by multiple projects and programs within the USGS. The construction of the DMPf will make considerable use of the library of existing documentation we have collected.

For the above objectives to be realized, the DMPf must recognize three over-arching concepts - or ‘dimensions’. These concepts are: Dual Data Management Cycles, Data Levels, and Document Layering.

Dual Data Management Cycles

The first framework concept is called “Dual Data Management Cycles”. This concept grew from our direct observation within the USGS of the disparate data management needs between researchers and traditional data management professionals. Research organizations such as the USGS need to manage data for two separate purposes: to support the immediate and flexible needs of researchers, and to provide for the long-term preservation and wide accessibility of data. The ‘dual cycle’ concept acknowledges the legitimate needs of both perspectives and expresses this as two separate but tightly linked data cycles: the research data cycle and the long-term preservation data cycle. These two cycles serve different needs and require different data management skills and approaches.

Research Data Cycle

In the research data cycle researchers need substantial control over the data, flexibility for handling and reshaping data, and the ability to use familiar tools (such as spreadsheets) for manipulating and managing their data. They also need to ensure that the data fit the requirements of their chosen analytical tools, and that confidentiality is maintained pending publication.

Preservation Data Cycle

In contrast, the long-term preservation data cycle is more the concern of the parent organizations, investigators, and future researchers. The preservation cycle is generally overseen by data management professionals who build and maintain data systems, adhere to standards, maintain documentation, transition data with changing technologies, and whose goals are to keep the data in its original context while providing accessibility to both current and future users.

The contrasting missions between research and long-term data management can create tension in an organization when the needs of one cycle are pitted against the needs of the other – especially in a funding-limited environment. A DMPf must recognize and address the needs of both of these cycles, and by doing so create a near seamless transition from the short-term research data cycle to the long-term preservation and delivery of those data.

Data Levels

The second concept needed to build a robust data management framework is that of “Data Levels”. A generalized DMPf must support data values and structure transforming from as a result of processing. In this context data levels identify the degree of transition from raw data to data products to published science documents. Each change in state brings with it a change in data management requirements.

This concept was borrowed from NEON (1) who in-turn borrowed it from NASA (2). These organizations designed their ‘Levels’ of transition to meet their unique needs, and the levels defined below are those we propose for USGS. These levels were defined to optimally correspond with changes in data management requirements as the data – and their progeny products - transition from initial measurement to research products. Without the benefit of this differentiation, discussions on data management, metadata, data formats, and data access often result in misunderstanding or disagreement because the topics themselves are level-specific. Recognizing and documenting data at different levels also provides the language necessary to trace higher level data products back to their source data. In the proposed DMPf we recognize data transition as comprising five Levels:

Level-0 - Raw Data

Level-0 or 'raw' data refers to data which are not modified except as necessary to allow for the long-term storage in an electronic format. Level-0 data may exist as discrete values recorded by a field worker, data streamed from instruments, photographs, strip-charts, or lab reports. The critical indicator is that the data values are unaltered as collected at the source. Raw data is not usually available for direct use. However, Level-0 data are a high priority for stable, long-term storage because data from all other levels are derived from Level-0 data.

Level-1 - Base Data

Level-1 or 'base' data refers to Level-0 data which have been reviewed, edited for quality when necessary, and certified for general use. Base data are effectively data that have been certified as ready-to-use. Activities that transition Level-0 to Level-1 data include flagging values for quality issues; removing outliers; handling missing data as appropriate; standardizing field names, units and formats; rearranging data; aggregating data into collections with similar data; and making collections of data accessible. In some cases well-formed Level-0 data may simply become certified as Level-1 quality and undergo no change at all. In other cases high frequency instrument generated data may be reduced to a surrogate value appropriate for scientific research. The result of these activities produces the high-quality, accepted, base facts used in all higher level interpretations of the original data. Base data may exist in many forms, such as discrete facts in relational databases, processed scenes from satellites, or LIDAR delivered by a contractor.

Level-2 - Data Products

Level-2 data are 'data products' resulting from further transformation of Level-1 data, or combinations of Level-1 and Level-2 data. Level-2 data are produced when base data are transformed by some degree of processing into formats and derived products that are more convenient for specific groups of data consumers. Level-2 transformations can range from the simple computation of stream discharge from stage data, to the complex calculations needed to transform Level-1 or other Level-2 data into seamless images, grids, or NetCDF files.

Level-3 - Project Databases

Level-3 or 'project databases' are the databases built by researchers as they accumulate data from all other levels to conduct their research. The main criterion defining Level-3 data is that the data have been integrated into a combined dataset for project use. This could include spatial and temporal registration to standards, blending data from different sources internal and external, and other similar activities. Proper data management at lower levels can mitigate some of the requirements for Level-3 data integration, but will not eliminate them.

Level-4 - Project Products

Level-4 or 'project products' are any product, publication, or document produced from the Level-3 research database. Much of data management at this level recognizes the essential nature of 'records' or 'document' management in a comprehensive data management framework. Project products may also include project databases made available to support specific data products such as a journal articles, USGS reports, or National Science Foundation projects.

Inclusion of Levels in the DMPf is important because it requires us to consider data management throughout the full evolution of data from its initial collection to publication, and provides traceability from higher level data products back to their original source data. Data Levels also help to focus our

thinking as we address broad concepts such as metadata, operational procedures, and data structures because the requirements change from level to level.

Document Layering

The third dimension for the DMPf is “Document Layering”. This concept was borrowed from the NPS Inventory and Monitoring program’s guidelines for National Park DMPs (3). We define layering as the process of differentiating Agency principles, standards, and requirements from those unique to a specific program or project. Layering creates a framework that sets the foundation for data integration across an agency or program. Science Centers, programs, and projects will be able to use a common framework to produce a DMP tailored to their specific requirements with minimal effort, while acknowledging and adhering to the broader requirements of an organization or data-sharing program.

Four layers are currently defined in our DMP document stack. The first two layers are common to all DMPs and the next two are specific to individual DMPs.

Base Layer

The first or ‘base’ DMP Layer provides the common DMP structure pre-populated with a Table-of-Contents and all of the common sections completed. The text of this layer provides the purpose, theory, and objectives of the DMP including the definition of roles and responsibilities, common vocabulary, case studies, references, etc.

Standards and Integration Layer

The second DMP Layer includes standards and procedures that will facilitate the integration and exchange of data across all DMPs written using this framework. These include sections for data standards, references to collection protocols, metadata standards, data model standards, naming conventions, data exchange formats, web service guidelines and other processes or requirements necessary for data integration.

Local Implementation Layer

The third DMP Layer documents requirements unique to a science center, program, or project. This is the only Layer an entity using the framework would need to complete in order to create their program-specific DMP because the ‘base’ and ‘standards’ layers are already complete and define the basic requirements shared by all data partners. The Local Implementation Layer documents any extensions to the DMP that the entity writing their program-DMP would require above and beyond base requirements. Deviations from the framework standards are also documented and defended in this layer.

Tactical Planning Layer

This last document Layer contains the year-by-year tactical plans for a data management program. This Layer is optional based upon the needs of the program to have an implementation schedule outlined within the DMP.

Summary

In summary, the ‘Dual Management Cycle’ acknowledges and clarifies the different roles researchers and data management professionals play in the development and preservation of data resources, ‘Data Levels’ defines a context for discussing handling and processing of data and ensures that data are

managed and traceable throughout their entire life span, while DMP ‘Document Layering’ creates a uniform template for reusability across many separate implementations, while facilitating data integration.

Development of the DMPf

An economy and broader adoption can be achieved by collaborating in the development of the DMPf with data management professionals across the USGS. The ASC and CEN have committed to jointly developing the DMPf just described. Our desire, however, is to expand the team to include contributors from other USGS programs and potentially other DOI Bureaus with similar needs. A broader team would help to create a truly generalized and portable DMP framework.

Much of the material for this proposed framework already exists in data management plans developed by research organizations and federal agencies from the U.S. and abroad. Wording from these existing plans can be adopted, adapted, and used to produce a general framework that will serve much of the USGS and other federal partners. After all, data management in support of research is not something new.

Works Cited

1. **NEON.** *NEON Cyberinfrastructure Architecture*. Cyberinfrastructure Team, National Ecological Observatory Network. 2010. Version B Draft (Not Published). NEON.DSDV.CYI.003024.DSN.
2. **NASA.** Earth Science Data Terminology and Formats - Data Processing Levels for EOSDIS Data Products. *NASA Science - Earth*. [Online] <http://science.nasa.gov/earth-science/earth-science-data/earth-science-data-terminology-and-formats/>.
3. **NPS.** *Data Management Guidelines for Inventory and Monitoring Networks*. Natural Resource Program Center, National Park Service - Department of the Interior. 2008. Natural Resource Report NPS/NRPC/NRR-2008/035. [Online] http://science.nature.nps.gov/im/datamgmt/docs/DMPlans/National_DM_Plan_v1.2.pdf.

Scope of DMPf Sections

Stanley Smith¹, Steven Tessler^{2,3}, Michael McHale³

U.S. Geological Survey

¹Alaska Science Center, ²New Jersey Water Science Center, ³Climate Effects Network

I. *Data Management (DM)*

Scope: Universe

Purpose: Describe and define the general principles and scope of professional data management as it applies to scientific research, resource management, long-term monitoring, decision support, and the perpetual preservation, maintenance, and accessibility of data assets for future use.

II. *Standards*

Scope: Enterprise (USGS)

Purpose: Define or provide links to all USGS data standards. Depending on the source and/or dynamic nature of the adopted standard, the standards may be included in a printed *Standards* section, an attached appendix, or identified and accessed through a web link to an external site or document.

III. *Research Data Management Plan (RDMP)*

a. *Enterprise RDMP*

Scope: Enterprise (USGS)

Purpose: Define the procedures and standards required by the *USGS* for each step in the *Research Data Management Cycle*. The Enterprise defines ‘what’ must be performed or managed by the Programs and Projects leaving it to the discretion of the Programs and/or Projects to articulate ‘how’ they are to perform these actions. The Enterprise should create set of standards, procedures (SOPs), and guidelines for the RDMP that constrain the degrees of freedom allowed the Programs and Projects. The Enterprise can lock a step so Programs cannot deviate, allow deviations and extensions, or leave a step to the discretion of the Project. The Enterprise layer includes applicable laws, policies, and standards from the Federal government, OMB, and the Department of Interior. For each topic in the RDMP the Enterprise will provide the following:

- Topic description
- Explanation of requirement
- List applicable standards
- Set degree of freedom for Programs

b. Program RDMP

Scope: Program

Purpose: Define the procedures and standards required by the *Program* for each step in the *Research Data Management Cycle* when they deviate from the *Enterprise* guidelines. A Program must defend their reason for extension or deviation. A Program can lock a step so Projects cannot deviate, allow deviations and extensions, or leave a step to the discretion of the Project.

For each topic in the RDMP a Program will provide the following:

- Explanation of deviations or extensions to the Enterprise processes
- List extensions or deviations to Enterprise standards
- Justify any deviations from Enterprise rules
- Assign persons to roles
- Set degree of freedom for Project

c. Project RDMP

Scope: Project

Purpose: Define the procedures and standards required by the *Project* for each step in the *Research Data Management Cycle* when they deviate from *Enterprise* or *Program* guidelines. A Project must defend their reason for extension or deviation.

For each topic in the RDMP the Project will provide the following:

- Explanation of deviations or extensions to the Program processes
- List extensions or deviations to Program standards
- Justify any deviations from Enterprise or Program rules
- Assign persons to roles

IV. Preservation Data Management Plan (PDMP)

a. Enterprise PDMP

Scope: Enterprise (USGS)

Purpose: Define the procedures and standards required by the *Enterprise* for each step in the *Preservation Data Management Cycle*. The Enterprise defines ‘what’ must be performed or managed by the Programs leaving it to the discretion of the Programs to articulate ‘how’ it will perform these actions. The Enterprise should create set of standards, procedures (SOPs), and guidelines for the PDMP that constrain the degrees of freedom allowed the Programs and Projects. The Enterprise layer includes applicable laws, policies, and standards from the Federal government, OMB, and the Department of Interior. For each topic in the PDMP the Enterprise will provide the following:

- Topic description
- Explanation of requirement
- List applicable standards
- Set degree of freedom for Program

b. Program PDMP

Scope: Program

Purpose: Define the procedures and standards required by the *Program* for each step in the *Preservation Data Management Cycle* when they deviate from *Enterprise* guidelines. For each topic in the PDMP the Program will provide the following:

- Explanation of deviations or extension to the Enterprise procedures
- List extensions or deviations to Enterprise standards
- Justify any deviations from Enterprise rules
- Assign persons to roles

c. Project PDMP

Scope: Project

Purpose: Define the procedures and standards required by the *Project* for each step in the *Preservation Data Management Cycle* when they deviate from *Enterprise* or *Program* guidelines. A project may be ongoing and require active maintenance of data archives and reshaped data prior to the transition to corporate preservation standards. For each topic in the PDMP the Project will provide the following:

- Explanation of deviations or extension to the Enterprise or Program procedures
- List additional standards or extensions unique to the Project
- Justify any deviations from Enterprise rules
- Assign persons to roles

V. Exposure and Delivery Plan (EDP)

a. Enterprise EDP

Scope: Enterprise (USGS)

Purpose: Set broad guidelines and requirements for how products and results of Enterprise Programs and Projects will be made accessible and understandable to a broader audience that includes other researchers, policy-makers, decision-makers, the media, and the public. Provide contact information for policy at the Enterprise level. The Enterprise layer includes applicable laws, policies, and standards from the Federal government, OMB, and the Department of Interior.

b. Program EDP

Scope: Program

Purpose: Explain how the products and results of this Program will be made accessible and understandable to a broader audience that includes other researchers, decision-makers, the media, and the public. Provide contact information for policy at the Program level. Set criteria for any Projects operating under this Program.

c. Project EDP

Scope: Project

Purpose: Explain how the products and results of this Program will be made accessible and understandable to a broader audience including other researchers, decision-makers, and the public. [Policy for dealing with the media is not in the purview of individual projects.]

Provide contact information. The EDP will include links to publications, presentations, and other non-data products that resulted from the research Project.

VI. Appendices

Definitions for terms used in the DMPf and this document

Degree of Freedom:

This term is used to acknowledge the flexibility allowed a Program or Project under the business rules and constraints imposed by a 'higher' level. It is expected that higher 'degrees of freedom' will equate to a lack of 'upper level' standards covering a data-related activity carried out by a Program or Project.

EDP:

The Exposure and Delivery Plan (EDP) is a document written to address the processes and activities associated with the use and dispersal of metadata to make data discoverable, and those separate technical systems and procedures that make the data accessible in formats that are convenient to users while maintaining the original context and provenance. Unlike the RDMP and PDMP, the EDP is not a data cycle as much as it is a set of associated activities and technologies that can apply to many kinds of data (= standardized). Cyclical events, such as update frequencies and storage/delivery reevaluations, are optional and not necessary for static data resources.

Enterprise:

An enterprise is an entire organization including all of its component organizations. The enterprise is not considered to be not part of another organization within the scope of this context. For this DMP USGS is considered to be the Enterprise.

Exposure and Delivery:

These terms represent separate-but-related activities that overlap and are timed to coincide with significant milestone events in the RDMP and PDMP cycles. Usually they occur after the research has been completed (at least for a fiscal cycle), and the data have been preserved and archived. The first part, Exposure, is the utilization and dispersal of metadata in various forms, and various marketing approaches (fact sheets, web sites). The Delivery side is a mechanism for users to obtain select data in a digital format that includes support documentation.

PDMP:

The Preservation Data Management Plan (PDMP) is a document written to address the processes and activities associated with the Preservation Data Management Cycle.

Preservation:

This term represents those activities associated with data storage and maintenance for perpetuity. The goal is to make all research data discoverable and available for use today and into the unforeseeable future.

Program:

Wikipedia: A research program is a coordinated set of projects undertaking related research, often at a national or even international level, with government funding.

Project:

Wikipedia: A project in business and science is a collaborative enterprise, frequently involving research or design that is carefully planned to achieve a particular aim.

RDMP:

The Research Data Management Plan (RDMP) is a document written to address the processes and activities associated with the Research Data Management Cycle.

Standards:

Data standards are a model or example established by authority, custom or general consent purposed to specify certain properties of the data, metadata, or data diagrams which apply consistency to the representation and documentation of data. Data standards manage naming conventions, data formats, table design, key structure, data definitions, and diagramming preferences.

Data Management Plan Framework

(Outline for a Data Management Plan Framework for Research Projects and Organizations)

Stanley Smith¹, Steven Tessler^{2,3}, Michael McHale³

U.S. Geological Survey

¹Alaska Science Center, ²New Jersey Water Science Center, ³Climate Effects Network

- i. My Program or Project
- ii. Acknowledgements
- iii. How to use this document

Table of Contents

i. My Program or Project	2
ii. Acknowledgements	2
iii. How to use this document	2
1 Introduction	1-2
1.1 Why Manage Data?.....	1-2
1.2 Objectives of the DMP framework	1-2
1.2.1 Scope of DMP Framework	1-2
1.3 Nomenclature	1-2
1.4 Data Management Roles.....	1-2
1.5 Metadata.....	1-2
1.6 Data Ownership	1-2
1.7 Dual Management Cycle for Research Organizations.....	1-2
1.7.1 Managing data during research	1-2
1.7.2 Long-term data management	1-2
1.8 Data Evolution.....	1-2
1.8.1 Planning.....	1-2
1.8.2 Collection	1-2
1.8.3 Raw data	1-2
1.8.4 Base data.....	1-2
1.8.5 Reshaped data	1-2
1.8.6 Project data.....	1-2
1.8.7 Project products.....	1-2

Data Management Plan Framework

1.9	Best Practices	1-2
1.10	Cost of Data Management	1-2
2	Roles and Responsibilities	2-2
2.1	Researcher	2-2
2.2	Systems Engineer	2-2
2.3	Systems Administrator	2-2
2.4	Data Manager	2-3
2.5	Data Modeler (Data Administrator).....	2-3
2.6	Database Administrator (DBA).....	2-3
2.7	Metadata Specialist.....	2-3
2.8	Data Resource Specialist	2-3
2.9	Application Developer.....	2-3
2.10	GIS Specialist	2-3
2.11	Web Developer	2-3
3	Laws and Policies	3-3
3.1	Federal	3-3
3.2	Department of Interior	3-3
3.3	U. S. Geological Survey.....	3-3
4	Sources of Data	4-3
4.1	Research Data	4-3
4.2	USGS Enterprise Data Stores.....	4-3
4.3	Data External to USGS.....	4-3
5	Data Quality.....	5-3
5.1	QA and QC.....	5-3
5.2	Quality Assurance Protection Plan (QAPP)	5-3
5.3	Data Versioning Control.....	5-3
5.4	References	5-3
6	Infrastructure and Architecture	6-4
6.1	Anatomy of Architecture	6-4
6.1.1	Use of Standards	6-4
6.1.2	Use of Translators	6-4
6.1.3	Use of Web Services.....	6-4
6.1.4	Use of Collaboration	6-4
6.2	Hardware Architecture	6-4
6.3	Software Architecture.....	6-4
6.4	File Architecture.....	6-4
6.5	Data Architecture.....	6-4

- 7 Integration through Standards7-4**
 - 7.1 Objective 7-4
 - 7.2 Define Data Sharing Standards 7-4
 - 7.3 Local Data Conformance to Standards 7-4
 - 7.4 Regional Consolidation of Data..... 7-4
 - 7.5 Establish new Enterprise Data Store..... 7-4
- 8 Establishing Standards8-4**
 - 8.1 Objectives..... 8-4
 - 8.2 Internal Standards..... 8-4
 - 8.2.1 Assemble SME Teams 8-4
 - 8.2.2 Define Standards..... 8-4
 - 8.2.3 External Review..... 8-4
 - 8.2.4 Test against existing Standards..... 8-5
 - 8.2.5 Publish Standard 8-5
 - 8.3 External Standards 8-5
 - 8.3.1 Form Partnerships..... 8-5
 - 8.3.2 Determine Delivery Architecture 8-5
 - 8.3.3 Project Metadata Exchange Standard 8-5
 - 8.3.4 Data Metadata Standard..... 8-5
 - 8.3.5 Data Exchange Standard 8-5
- 9 Standards.....9-2**
 - 9.1 Laws and Policies 9-2
 - 9.2 Data Modeling..... 9-2
 - 9.3 Data Dictionary 9-2
 - 9.4 Naming 9-2
 - 9.5 Domains 9-2
 - 9.6 Vocabularies and Ontologies 9-2
 - 9.7 Data Types..... 9-2
 - 9.8 Data Structure Types..... 9-2
 - 9.9 Data Formats..... 9-2
 - 9.9.1 Spatial Data 9-2
 - 9.9.2 Temporal Data 9-2
 - 9.10 File Formats and Organizations 9-2
 - 9.11 Metadata..... 9-2
 - 9.12 Documentation 9-2
- 10 Enterprise Research Data Management Plan10-2**
 - 10.1 Objective 10-2
 - 10.2 Planning..... 10-2

Data Management Plan Framework

10.2.1	Laws and Policies	10-2
10.2.2	Scoping	10-2
10.2.3	Project Metadata	10-2
10.2.4	Roles.....	10-2
10.2.5	Business Requirements.....	10-2
10.2.6	Data Requirements	10-2
10.2.7	Architectural Plan.....	10-2
10.2.8	Data Management Work Plan.....	10-2
10.2.9	Planning for non-digital object preservation	10-2
10.2.10	Data Management Budgeting	10-2
10.2.11	Approval to Proceed	10-2
10.3	Design and Testing.....	10-2
10.3.1	Design Data Capture Methodologies.....	10-2
10.3.2	Hardware Architecture	10-2
10.3.3	Software Architecture.....	10-2
10.3.4	Define Metadata Requirements	10-2
10.3.5	Data Architecture.....	10-2
10.3.6	Define QA/QC Requirements.....	10-3
10.3.7	Chain of Custody Requirements	10-3
10.3.8	Define Version Control Requirements.....	10-3
10.3.9	Define Data Access Restrictions.....	10-3
10.3.10	Test the Infrastructure.....	10-3
10.3.11	Approval to Proceed	10-3
10.4	Data Collection.....	10-3
10.4.1	Metadata.....	10-3
10.4.2	Protocols	10-3
10.4.3	Manage Data Source Files.....	10-3
10.4.4	Chain of Custody Recording.....	10-3
10.4.5	Data Quality Notation	10-3
10.5	Raw Data Management	10-3
10.5.1	Metadata.....	10-3
10.5.2	Load Raw Data	10-3
10.5.3	Digitize Forms.....	10-3
10.5.4	Apply Access Control Policy	10-3
10.5.5	Backup and Archive.....	10-3
10.6	Base Data Management.....	10-3
10.6.1	Metadata.....	10-3
10.6.2	Record Raw Data Source.....	10-3
10.6.3	Data Validation Checks	10-3
10.6.4	Quality Control Checks.....	10-3

Data Management Plan Framework

10.6.5	Identify Data Gaps.....	10-3
10.6.6	Handle Data Gaps.....	10-3
10.6.7	Apply Access Control Policy	10-3
10.6.8	Backup and Archive.....	10-4
10.6.9	Certify Data for Release	10-4
10.6.10	Schedule for Hand-off to Preservation Cycle.....	10-4
10.7	Derived Data Products	10-4
10.7.1	Product Description	10-4
10.7.2	Metadata.....	10-4
10.7.3	Data Sources	10-4
10.7.4	Data Transform Process	10-4
10.7.5	Quality Control Procedures.....	10-4
10.7.6	Certify for Release.....	10-4
10.7.7	Backup and Archive.....	10-4
10.7.8	Recalculation Schedule	10-4
10.8	Project Database Management	10-4
10.8.1	Database Description.....	10-4
10.9	Project Products.....	10-4
10.10	Hand-off to Data Preservation Cycle	10-4
11	Program Research Data Management Plan	11-5
11.1	Research Program and Objectives	11-5
11.2	Planning.....	11-5
11.2.1	Laws and Policies	11-5
11.2.2	Scoping.....	11-5
11.2.3	Project Metadata	11-5
11.2.4	Roles.....	11-5
11.2.5	Business Requirements.....	11-5
11.2.6	Data Requirements	11-5
11.2.7	Architectural Plan.....	11-5
11.2.8	Data Management Work Plan.....	11-5
11.2.9	Planning for non-digital object preservation	11-5
11.2.10	Data Management Budgeting.....	11-5
11.2.11	Approval to Proceed	11-5
11.3	Design and Testing.....	11-5
11.3.1	Design Data Capture Methodologies.....	11-5
11.3.2	Hardware Architecture	11-5
11.3.3	Software Architecture.....	11-5
11.3.4	Define Metadata Requirements	11-5
11.3.5	Data Architecture.....	11-5
11.3.6	Define QA/QC Requirements.....	11-6

Data Management Plan Framework

11.3.7	Chain of Custody Requirements	11-6
11.3.8	Define Version Control Requirements	11-6
11.3.9	Define Data Access Restrictions	11-6
11.3.10	Test the Infrastructure	11-6
11.3.11	Approval to Proceed	11-6
11.4	Data Collection	11-6
11.4.1	Metadata	11-6
11.4.2	Protocols	11-6
11.4.3	Manage Data Source Files	11-6
11.4.4	Chain of Custody Recording	11-6
11.4.5	Data Quality Notation	11-6
11.5	Raw Data Management	11-6
11.5.1	Metadata	11-6
11.5.2	Load Raw Data	11-6
11.5.3	Digitize Forms	11-6
11.5.4	Apply Access Control Policy	11-6
11.5.5	Backup and Archive	11-6
11.6	Base Data Management	11-6
11.6.1	Metadata	11-6
11.6.2	Record Raw Data Source	11-6
11.6.3	Data Validation Checks	11-6
11.6.4	Quality Control Checks	11-6
11.6.5	Identify Data Gaps	11-6
11.6.6	Handle Data Gaps	11-6
11.6.7	Apply Access Control Policy	11-6
11.6.8	Backup and Archive	11-7
11.6.9	Certify Data for Release	11-7
11.6.10	Schedule for Hand-off to Preservation Cycle	11-7
11.7	Derived Data Products	11-7
11.7.1	Product Description	11-7
11.7.2	Metadata	11-7
11.7.3	Data Sources	11-7
11.7.4	Data Transform Process	11-7
11.7.5	Quality Control Procedures	11-7
11.7.6	Certify for Release	11-7
11.7.7	Backup and Archive	11-7
11.7.8	Recalculation Schedule	11-7
11.8	Project Database Management	11-7
11.8.1	Database Description	11-7
11.9	Project Products	11-7

Data Management Plan Framework

11.10	Hand-off to Data Preservation Cycle	11-7
12	Project Research Data Management Plan	12-8
12.1	Research Project and Objectives	12-8
12.2	Planning.....	12-8
12.2.1	Laws and Policies	12-8
12.2.2	Scoping.....	12-8
12.2.3	Project Metadata	12-8
12.2.4	Roles.....	12-8
12.2.5	Business Requirements.....	12-8
12.2.6	Data Requirements	12-8
12.2.7	Architectural Plan.....	12-8
12.2.8	Data Management Work Plan.....	12-8
12.2.9	Planning for non-digital object preservation	12-8
12.2.10	Data Management Budgeting.....	12-8
12.2.11	Approval to Proceed	12-8
12.3	Design and Testing.....	12-8
12.3.1	Design Data Capture Methodologies.....	12-8
12.3.2	Hardware Architecture	12-8
12.3.3	Software Architecture.....	12-8
12.3.4	Define Metadata Requirements	12-8
12.3.5	Data Architecture.....	12-8
12.3.6	Define QA/QC Requirements.....	12-9
12.3.7	Chain of Custody Requirements	12-9
12.3.8	Define Version Control Requirements.....	12-9
12.3.9	Define Data Access Restrictions.....	12-9
12.3.10	Test the Infrastructure.....	12-9
12.3.11	Approval to Proceed	12-9
12.4	Data Collection.....	12-9
12.4.1	Metadata.....	12-9
12.4.2	Protocols	12-9
12.4.3	Manage Data Source Files.....	12-9
12.4.4	Chain of Custody Recording.....	12-9
12.4.5	Data Quality Notation	12-9
12.5	Raw Data Management	12-9
12.5.1	Metadata.....	12-9
12.5.2	Load Raw Data	12-9
12.5.3	Digitize Forms.....	12-9
12.5.4	Apply Access Control Policy	12-9
12.5.5	Backup and Archive.....	12-9
12.6	Base Data Management.....	12-9

Data Management Plan Framework

12.6.1	Metadata.....	12-9
12.6.2	Record Raw Data Source.....	12-9
12.6.3	Data Validation Checks	12-9
12.6.4	Quality Control Checks.....	12-9
12.6.5	Identify Data Gaps.....	12-9
12.6.6	Handle Data Gaps.....	12-9
12.6.7	Apply Access Control Policy	12-9
12.6.8	Backup and Archive.....	12-10
12.6.9	Certify Data for Release	12-10
12.6.10	Schedule for Hand-off to Preservation Cycle.....	12-10
12.7	Derived Data Products	12-10
12.7.1	Product Description	12-10
12.7.2	Metadata.....	12-10
12.7.3	Data Sources	12-10
12.7.4	Data Transform Process	12-10
12.7.5	Quality Control Procedures.....	12-10
12.7.6	Certify for Release.....	12-10
12.7.7	Backup and Archive.....	12-10
12.7.8	Recalculation Schedule	12-10
12.8	Project Database Management	12-10
12.8.1	Database Description	12-10
12.9	Project Products.....	12-10
12.10	Hand-off to Data Preservation Cycle	12-10
13	Enterprise Data Preservation Plan.....	13-2
13.1	Objective	13-2
13.2	Data Transition Planning.....	13-2
13.2.1	Laws and Policies	13-2
13.2.2	Roles.....	13-2
13.2.3	Supporting Documentation.....	13-2
13.2.4	Evaluation of Existing Data Model	13-2
13.2.5	Validation of Data	13-2
13.2.6	Transformation of Data Types	13-2
13.2.7	Cost of Data Incorporation.....	13-2
13.3	Long Term Data Storage Planning.....	13-2
13.3.1	Laws and Policies	13-2
13.3.2	Roles.....	13-2
13.3.3	Data Volume Estimation	13-2
13.3.4	Impacts to Architecture	13-2
13.3.5	Frequency of Acquisition	13-2
13.3.6	QA/QC Requirements.....	13-2

Data Management Plan Framework

13.3.7	Purge Policy.....	13-2
13.3.8	Backup and Recovery Policy.....	13-2
13.3.9	Access Restriction	13-2
13.3.10	Versioning Requirements.....	13-2
13.3.11	Programming Requirement	13-2
13.3.12	Estimated Recurring Costs	13-2
13.4	Data Acquisition	13-2
13.4.1	Data Loading Programs	13-2
13.4.2	Load Data	13-2
13.5	Base Data Management.....	13-2
13.5.1	Certify Data for Release	13-3
13.5.2	Refresh Maintenance Program	13-3
13.5.3	Data Maintenance Program.....	13-3
13.6	Digital Data Maintenance	13-3
13.6.1	Media Rotation	13-3
13.6.2	Versioning Policy	13-3
13.6.3	Backup and Recovery.....	13-3
13.7	Data Archiving.....	13-3
13.7.1	Raw Data	13-3
13.7.2	Tabular Databases.....	13-3
13.7.3	Data Products.....	13-3
13.7.4	Records Management.....	13-3
13.8	Annual Work Plans.....	13-3
14	Program Data Preservation Plan	14-4
14.1	Program Objectives.....	14-4
14.2	Data Transition Planning.....	14-4
14.2.1	Laws and Policies	14-4
14.2.2	Roles.....	14-4
14.2.3	Supporting Documentation.....	14-4
14.2.4	Evaluation of Existing Data Model	14-4
14.2.5	Validation of Data	14-4
14.2.6	Transformation of Data Types	14-4
14.2.7	Cost of Data Incorporation.....	14-4
14.3	Long Term Data Storage Planning.....	14-4
14.3.1	Laws and Policies	14-4
14.3.2	Roles.....	14-4
14.3.3	Data Volume Estimation	14-4
14.3.4	Impacts to Architecture	14-4
14.3.5	Frequency of Acquisition	14-4
14.3.6	QA/QC Requirements.....	14-4

Data Management Plan Framework

14.3.7	Purge Policy.....	14-4
14.3.8	Backup and Recovery Policy.....	14-4
14.3.9	Access Restriction	14-4
14.3.10	Versioning Requirements.....	14-4
14.3.11	Programming Requirement	14-4
14.3.12	Estimated Recurring Costs	14-4
14.4	Data Acquisition	14-4
14.4.1	Data Loading Programs	14-4
14.4.2	Load Data	14-4
14.5	Base Data Management.....	14-4
14.5.1	Certify Data for Release	14-5
14.5.2	Refresh Maintenance Program	14-5
14.5.3	Data Maintenance Program.....	14-5
14.6	Digital Data Maintenance	14-5
14.6.1	Media Rotation	14-5
14.6.2	Versioning Policy	14-5
14.6.3	Backup and Recovery.....	14-5
14.7	Data Archiving.....	14-5
14.7.1	Raw Data	14-5
14.7.2	Tabular Databases.....	14-5
14.7.3	Data Products.....	14-5
14.7.4	Records Management.....	14-5
14.8	Annual Work Plans.....	14-5
15	Enterprise Data Exposure and Delivery Plan.....	15-2
15.1	Communication.....	15-2
15.2	Marketing.....	15-2
16	Program Data Exposure and Delivery Plan	16-3
16.1	Communication.....	16-3
16.2	Marketing.....	16-3
17	Project Data Exposure and Delivery Plan.....	17-4
17.1	Communication.....	17-4
17.2	Marketing.....	17-4
18	Appendices	18-2
18.1	Glossary.....	18-2
18.1.1	Acronyms	18-2
18.1.2	Terms and Definitions	18-2
18.2	Bibliography	18-2
18.3	Forms	18-2

Data Management Plan Framework

18.4	Data Collection Protocols.....	18-2
18.5	Variable Description Sheets.....	18-2
18.6	Skills and Training Guidelines.....	18-2

PART I

MANAGING DATA

1 Introduction

1.1 Why Manage Data?

1.2 Objectives of the DMP framework

1.2.1 Scope of DMP Framework

1.3 Nomenclature

1.4 Data Management Roles

1.5 Metadata

1.6 Data Ownership

1.7 Dual Management Cycle for Research Organizations

1.7.1 Managing data during research

1.7.2 Long-term data management

1.8 Data Evolution

1.8.1 Planning

1.8.2 Collection

1.8.3 Raw data

1.8.4 Base data

1.8.5 Reshaped data

1.8.6 Project data

1.8.7 Project products

1.9 Best Practices

1.10 Cost of Data Management

2 Roles and Responsibilities

2.1 Researcher

2.2 Systems Engineer

2.3 Systems Administrator

2.4 Data Manager

2.5 Data Modeler (Data Administrator)

2.6 Database Administrator (DBA)

2.7 Metadata Specialist

2.8 Data Resource Specialist

2.9 Application Developer

2.10 GIS Specialist

2.11 Web Developer

3 Laws and Policies

3.1 Federal

3.2 Department of Interior

3.3 U. S. Geological Survey

4 Sources of Data

4.1 Research Data

4.2 USGS Enterprise Data Stores

4.3 Data External to USGS

5 Data Quality

5.1 QA and QC

5.2 Quality Assurance Protection Plan (QAPP)

5.3 Data Versioning Control

5.4 References

6 Infrastructure and Architecture

6.1 Anatomy of Architecture

- 6.1.1 Use of Standards
- 6.1.2 Use of Translators
- 6.1.3 Use of Web Services
- 6.1.4 Use of Collaboration

6.2 Hardware Architecture

6.3 Software Architecture

6.4 File Architecture

6.5 Data Architecture

7 Integration through Standards

7.1 Objective

7.2 Define Data Sharing Standards

7.3 Local Data Conformance to Standards

7.4 Regional Consolidation of Data

7.5 Establish new Enterprise Data Store

8 Establishing Standards

8.1 Objectives

8.2 Internal Standards

- 8.2.1 Assemble SME Teams
- 8.2.2 Define Standards
 - 8.2.2.1 *Adopt a Standard*
 - 8.2.2.2 *New Standards*
- 8.2.3 External Review

Data Management Plan Framework

8.2.4 Test against existing Standards

8.2.5 Publish Standard

8.3 External Standards

8.3.1 Form Partnerships

8.3.2 Determine Delivery Architecture

8.3.2.1 *Exchange Technology*

8.3.2.2 *Roles and Responsibilities*

8.3.2.3 *Long-term Maintenance Approach*

8.3.3 Project Metadata Exchange Standard

8.3.4 Data Metadata Standard

8.3.5 Data Exchange Standard

Data Management Plan Framework

PART II

STANDARDS

9 Standards

9.1 Laws and Policies

9.2 Data Modeling

9.2.1.1 Business Model

9.2.1.2 Conceptual Data Model Standards

9.2.1.3 Logical Data Model Standards

9.2.1.4 Physical Data Model Standards

9.3 Data Dictionary

9.3.1.1 Use of Central Data Repository

9.4 Naming

9.5 Domains

9.6 Vocabularies and Ontologies

9.7 Data Types

9.8 Data Structure Types

9.9 Data Formats

9.9.1 Spatial Data

9.9.2 Temporal Data

9.10 File Formats and Organizations

9.11 Metadata

9.12 Documentation

PART III

RESEARCH DATA MANAGEMENT

10 Enterprise Research Data Management Plan

10.1 Objective

10.2 Planning

- 10.2.1 Laws and Policies
- 10.2.2 Scoping
 - 10.2.2.1 *Expected uses of data*
- 10.2.3 Project Metadata
- 10.2.4 Roles
- 10.2.5 Business Requirements
- 10.2.6 Data Requirements
 - 10.2.6.1 *Conceptual Data Model*
 - 10.2.6.2 *Sampling Protocols*
 - 10.2.6.3 *Identify Data Sources*
 - 10.2.6.4 *Estimate Data Volume*
- 10.2.7 Architectural Plan
- 10.2.8 Data Management Work Plan
- 10.2.9 Planning for non-digital object preservation
- 10.2.10 Data Management Budgeting
 - 10.2.10.1 *New Infrastructure*
 - 10.2.10.2 *Data Management Personnel Costs*
- 10.2.11 Approval to Proceed

10.3 Design and Testing

- 10.3.1 Design Data Capture Methodologies
- 10.3.2 Hardware Architecture
- 10.3.3 Software Architecture
- 10.3.4 Define Metadata Requirements
- 10.3.5 Data Architecture
 - 10.3.5.1 *Logical Data Model*
 - 10.3.5.2 *Data Standards*
 - 10.3.5.3 *Data Collection Forms*

Data Management Plan Framework

10.3.5.4 *Physical Data Model*

10.3.5.5 *Map Form Fields to Data Models*

10.3.5.6 *File Architecture*

10.3.6 Define QA/QC Requirements

10.3.7 Chain of Custody Requirements

10.3.8 Define Version Control Requirements

10.3.9 Define Data Access Restrictions

10.3.10 Test the Infrastructure

10.3.11 Approval to Proceed

10.4 Data Collection

10.4.1 Metadata

10.4.2 Protocols

10.4.3 Manage Data Source Files

10.4.4 Chain of Custody Recording

10.4.4.1 *Instrumentation Logging*

10.4.4.2 *Non-digital Object Location Tracking*

10.4.5 Data Quality Notation

10.5 Raw Data Management

10.5.1 Metadata

10.5.2 Load Raw Data

10.5.3 Digitize Forms

10.5.4 Apply Access Control Policy

10.5.5 Backup and Archive

10.6 Base Data Management

10.6.1 Metadata

10.6.2 Record Raw Data Source

10.6.3 Data Validation Checks

10.6.4 Quality Control Checks

10.6.5 Identify Data Gaps

10.6.6 Handle Data Gaps

10.6.7 Apply Access Control Policy

Data Management Plan Framework

10.6.8 Backup and Archive

10.6.9 Certify Data for Release

10.6.10 Schedule for Hand-off to Preservation Cycle

10.7 Derived Data Products

10.7.1 Product Description

10.7.2 Metadata

10.7.3 Data Sources

10.7.4 Data Transform Process

10.7.5 Quality Control Procedures

10.7.6 Certify for Release

10.7.7 Backup and Archive

10.7.8 Recalculation Schedule

10.8 Project Database Management

10.8.1 Database Description

10.9 Project Products

10.10 Hand-off to Data Preservation Cycle

11 Program Research Data Management Plan

11.1 Research Program and Objectives

11.2 Planning

- 11.2.1 Laws and Policies
- 11.2.2 Scoping
 - 11.2.2.1 *Expected uses of data*
- 11.2.3 Project Metadata
- 11.2.4 Roles
- 11.2.5 Business Requirements
- 11.2.6 Data Requirements
 - 11.2.6.1 *Conceptual Data Model*
 - 11.2.6.2 *Sampling Protocols*
 - 11.2.6.3 *Identify Data Sources*
 - 11.2.6.4 *Estimate Data Volume*
- 11.2.7 Architectural Plan
- 11.2.8 Data Management Work Plan
- 11.2.9 Planning for non-digital object preservation
- 11.2.10 Data Management Budgeting
 - 11.2.10.1 *New Infrastructure*
 - 11.2.10.2 *Data Management Personnel Costs*
- 11.2.11 Approval to Proceed

11.3 Design and Testing

- 11.3.1 Design Data Capture Methodologies
- 11.3.2 Hardware Architecture
- 11.3.3 Software Architecture
- 11.3.4 Define Metadata Requirements
- 11.3.5 Data Architecture
 - 11.3.5.1 *Logical Data Model*
 - 11.3.5.2 *Data Standards*
 - 11.3.5.3 *Data Collection Forms*

Data Management Plan Framework

11.3.5.4 Physical Data Model

11.3.5.5 Map Form Fields to Data Models

11.3.5.6 File Architecture

11.3.6 Define QA/QC Requirements

11.3.7 Chain of Custody Requirements

11.3.8 Define Version Control Requirements

11.3.9 Define Data Access Restrictions

11.3.10 Test the Infrastructure

11.3.11 Approval to Proceed

11.4 Data Collection

11.4.1 Metadata

11.4.2 Protocols

11.4.3 Manage Data Source Files

11.4.4 Chain of Custody Recording

11.4.4.1 Instrumentation Logging

11.4.4.2 Non-digital Object Location Tracking

11.4.5 Data Quality Notation

11.5 Raw Data Management

11.5.1 Metadata

11.5.2 Load Raw Data

11.5.3 Digitize Forms

11.5.4 Apply Access Control Policy

11.5.5 Backup and Archive

11.6 Base Data Management

11.6.1 Metadata

11.6.2 Record Raw Data Source

11.6.3 Data Validation Checks

11.6.4 Quality Control Checks

11.6.5 Identify Data Gaps

11.6.6 Handle Data Gaps

11.6.7 Apply Access Control Policy

Data Management Plan Framework

11.6.8 Backup and Archive

11.6.9 Certify Data for Release

11.6.10 Schedule for Hand-off to Preservation Cycle

11.7 Derived Data Products

11.7.1 Product Description

11.7.2 Metadata

11.7.3 Data Sources

11.7.4 Data Transform Process

11.7.5 Quality Control Procedures

11.7.6 Certify for Release

11.7.7 Backup and Archive

11.7.8 Recalculation Schedule

11.8 Project Database Management

11.8.1 Database Description

11.9 Project Products

11.10 Hand-off to Data Preservation Cycle

12 Project Research Data Management Plan

12.1 Research Project and Objectives

12.2 Planning

- 12.2.1 Laws and Policies
- 12.2.2 Scoping
 - 12.2.2.1 *Expected uses of data*
- 12.2.3 Project Metadata
- 12.2.4 Roles
- 12.2.5 Business Requirements
- 12.2.6 Data Requirements
 - 12.2.6.1 *Conceptual Data Model*
 - 12.2.6.2 *Sampling Protocols*
 - 12.2.6.3 *Identify Data Sources*
 - 12.2.6.4 *Estimate Data Volume*
- 12.2.7 Architectural Plan
- 12.2.8 Data Management Work Plan
- 12.2.9 Planning for non-digital object preservation
- 12.2.10 Data Management Budgeting
 - 12.2.10.1 *New Infrastructure*
 - 12.2.10.2 *Data Management Personnel Costs*
- 12.2.11 Approval to Proceed

12.3 Design and Testing

- 12.3.1 Design Data Capture Methodologies
- 12.3.2 Hardware Architecture
- 12.3.3 Software Architecture
- 12.3.4 Define Metadata Requirements
- 12.3.5 Data Architecture
 - 12.3.5.1 *Logical Data Model*
 - 12.3.5.2 *Data Standards*
 - 12.3.5.3 *Data Collection Forms*

Data Management Plan Framework

12.3.5.4 *Physical Data Model*

12.3.5.5 *Map Form Fields to Data Models*

12.3.5.6 *File Architecture*

12.3.6 Define QA/QC Requirements

12.3.7 Chain of Custody Requirements

12.3.8 Define Version Control Requirements

12.3.9 Define Data Access Restrictions

12.3.10 Test the Infrastructure

12.3.11 Approval to Proceed

12.4 Data Collection

12.4.1 Metadata

12.4.2 Protocols

12.4.3 Manage Data Source Files

12.4.4 Chain of Custody Recording

12.4.4.1 *Instrumentation Logging*

12.4.4.2 *Non-digital Object Location Tracking*

12.4.5 Data Quality Notation

12.5 Raw Data Management

12.5.1 Metadata

12.5.2 Load Raw Data

12.5.3 Digitize Forms

12.5.4 Apply Access Control Policy

12.5.5 Backup and Archive

12.6 Base Data Management

12.6.1 Metadata

12.6.2 Record Raw Data Source

12.6.3 Data Validation Checks

12.6.4 Quality Control Checks

12.6.5 Identify Data Gaps

12.6.6 Handle Data Gaps

12.6.7 Apply Access Control Policy

Data Management Plan Framework

12.6.8 Backup and Archive

12.6.9 Certify Data for Release

12.6.10 Schedule for Hand-off to Preservation Cycle

12.7 Derived Data Products

12.7.1 Product Description

12.7.2 Metadata

12.7.3 Data Sources

12.7.4 Data Transform Process

12.7.5 Quality Control Procedures

12.7.6 Certify for Release

12.7.7 Backup and Archive

12.7.8 Recalculation Schedule

12.8 Project Database Management

12.8.1 Database Description

12.9 Project Products

12.10 Hand-off to Data Preservation Cycle

Data Management Plan Framework

PART IV

PRESERVATION DATA MANAGEMENT

13 Enterprise Data Preservation Plan

13.1 Objective

13.2 Data Transition Planning

- 13.2.1 Laws and Policies
- 13.2.2 Roles
- 13.2.3 Supporting Documentation
 - 13.2.3.1 *Project Metadata*
 - 13.2.3.2 *Data Models*
- 13.2.4 Evaluation of Existing Data Model
- 13.2.5 Validation of Data
- 13.2.6 Transformation of Data Types
- 13.2.7 Cost of Data Incorporation

13.3 Long Term Data Storage Planning

- 13.3.1 Laws and Policies
- 13.3.2 Roles
- 13.3.3 Data Volume Estimation
- 13.3.4 Impacts to Architecture
- 13.3.5 Frequency of Acquisition
- 13.3.6 QA/QC Requirements
- 13.3.7 Purge Policy
- 13.3.8 Backup and Recovery Policy
- 13.3.9 Access Restriction
- 13.3.10 Versioning Requirements
- 13.3.11 Programming Requirement
- 13.3.12 Estimated Recurring Costs

13.4 Data Acquisition

- 13.4.1 Data Loading Programs
- 13.4.2 Load Data

13.5 Base Data Management

Data Management Plan Framework

- 13.5.1 Certify Data for Release
- 13.5.2 Refresh Maintenance Program
- 13.5.3 Data Maintenance Program

13.6 Digital Data Maintenance

- 13.6.1 Media Rotation
- 13.6.2 Versioning Policy
- 13.6.3 Backup and Recovery

13.7 Data Archiving

- 13.7.1 Raw Data
- 13.7.2 Tabular Databases
- 13.7.3 Data Products
- 13.7.4 Records Management

13.8 Annual Work Plans

14 Program Data Preservation Plan

14.1 Program Objectives

14.2 Data Transition Planning

- 14.2.1 Laws and Policies
- 14.2.2 Roles
- 14.2.3 Supporting Documentation
 - 14.2.3.1 *Project Metadata*
 - 14.2.3.2 *Data Models*
- 14.2.4 Evaluation of Existing Data Model
- 14.2.5 Validation of Data
- 14.2.6 Transformation of Data Types
- 14.2.7 Cost of Data Incorporation

14.3 Long Term Data Storage Planning

- 14.3.1 Laws and Policies
- 14.3.2 Roles
- 14.3.3 Data Volume Estimation
- 14.3.4 Impacts to Architecture
- 14.3.5 Frequency of Acquisition
- 14.3.6 QA/QC Requirements
- 14.3.7 Purge Policy
- 14.3.8 Backup and Recovery Policy
- 14.3.9 Access Restriction
- 14.3.10 Versioning Requirements
- 14.3.11 Programming Requirement
- 14.3.12 Estimated Recurring Costs

14.4 Data Acquisition

- 14.4.1 Data Loading Programs
- 14.4.2 Load Data

14.5 Base Data Management

Data Management Plan Framework

- 14.5.1 Certify Data for Release
- 14.5.2 Refresh Maintenance Program
- 14.5.3 Data Maintenance Program

14.6 Digital Data Maintenance

- 14.6.1 Media Rotation
- 14.6.2 Versioning Policy
- 14.6.3 Backup and Recovery

14.7 Data Archiving

- 14.7.1 Raw Data
- 14.7.2 Tabular Databases
- 14.7.3 Data Products
- 14.7.4 Records Management

14.8 Annual Work Plans

Data Management Plan Framework



PART V

DATA EXPOSURE AND DELIVERY

15 Enterprise Data Exposure and Delivery Plan

15.1 Communication

15.2 Marketing

15.3 Metadata Repositories

15.4 Data Delivery

15.4.1 Data Upload to Repositories

15.4.2 Portals and Websites for Data Download

15.4.3 Web Services

16 Program Data Exposure and Delivery Plan

16.1 Communication

16.2 Marketing

16.3 Metadata Repositories

16.4 Data Delivery

16.4.1 Data Upload to Repositories

16.4.2 Portals and Websites for Data Download

16.4.3 Web Services

17 Project Data Exposure and Delivery Plan

17.1 Communication

17.2 Marketing

17.3 Metadata Repositories

17.4 Data Delivery

17.4.1 Data Upload to Repositories

17.4.2 Portals and Websites for Data Download

17.4.3 Web Services

Data Management Plan Framework

PART VI

APPENDICES

18 Appendices

18.1 Glossary

18.1.1 Acronyms

18.1.2 Terms and Definitions

18.2 Bibliography

18.3 Forms

18.4 Data Collection Protocols

18.5 Variable Description Sheets

18.6 Skills and Training Guidelines