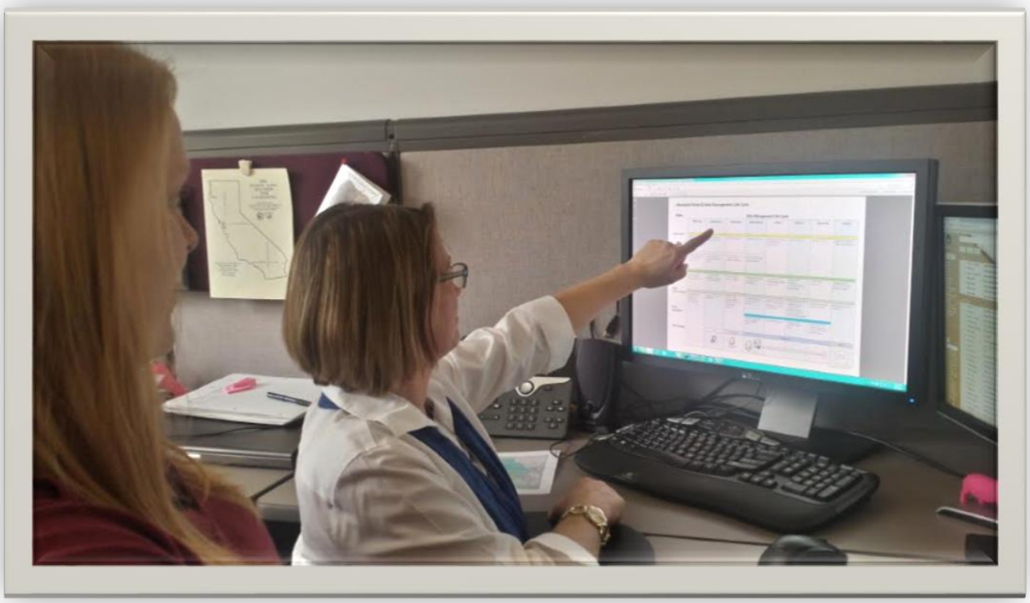
RECLAMATION
*Managing Water in the West*

# Training Materials for Data Management in Reclamation

**Research and Development Office**
**Science and Technology Program**
**(Final Report) ST-2016-ID-Y3823**



Douglas R. Clark, Denver TSC
Barbara Simpson, MP-GIS

**U.S. Department of the Interior**
**Bureau of Reclamation**
**Research and Development Office**

**June 2016**

# Mission Statements

The U.S. Department of the Interior protects America's natural resources and heritage, honors our cultures and tribal communities, and supplies the energy to power our future.

The mission of the Bureau of Reclamation is to manage, develop, and protect water and related resources in an environmentally and economically sound manner in the interest of the American public.

| REPORT DOCUMENTATION PAGE | | *Form Approved*<br>*OMB No. 0704-0188* |
|---|---|---|
| **T1. REPORT DATE** 06/2016 | **T2. REPORT TYPE** Research | **T3. DATES COVERED FY2016** |
| **T4. TITLE AND SUBTITLE Training Materials for Data Management in Reclamation** | | **5a. CONTRACT NUMBER: 16XR0680A1-RY1541DS201523840 (8)** |
| | | **5b. GRANT NUMBER** |
| | | **5c. PROGRAM ELEMENT NUMBER**<br>1541 (S&T) |
| **6. AUTHOR(S)**<br>**Douglas Clark, Denver TSC**<br>**Barbara Simpson, MP Region** | | **5d. PROJECT NUMBER**<br>Y3823 |
| | | **5e. TASK NUMBER** |
| | | **5f. WORK UNIT NUMBER**<br>86-68260 |
| **7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES): Douglas R. Clark, USBR Technical Service Center, Bldg. 67, Denver Fed. Ctr., PO 25007, 86-68260, 6th Ave and Kipling, Denver, CO 80225** | | **8. PERFORMING ORGANIZATION REPORT NUMBER** |
| **9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**<br>Research and Development Office<br>U.S. Department of the Interior, Bureau of Reclamation,<br>PO Box 25007, Denver CO 80225-0007 | | **10. SPONSOR/MONITOR'S ACRONYM(S)**<br>**R&D: Research and Development Office**<br>**BOR/USBR: Bureau of Reclamation**<br>**DOI: Department of the Interior** |
| | | **11. SPONSOR/MONITOR'S REPORT NUMBER(S): ST-2016-ID-Y3823** |

**12. DISTRIBUTION / AVAILABILITY STATEMENT**

**Final report can be downloaded from Reclamation's website: https://www.usbr.gov/research/**

**13. SUPPLEMENTARY NOTES**


**14. ABSTRACT** ***(Maximum 200 words)***

***In the wake of the Department of the Interior Data Management policy initiatives and the Open Data Executive Order Reclamation has been embarking on implementation of data management best practices within the agency. The USBR Data Stewardship Core Team (DSCT), active for nearly a decade, saw development of training materials as an important first step.***

***Ms. Barbara Simpson, a member of that team, regularly conducts data management training using three training modules developed by the U.S. Geological Survey. Her students have found the modules to be very informative, as has the DSCT. The team determined, however, that supplementary Reclamation-related materials would help make the concepts of data management provided in the modules more intelligible to USBR users. This document provides those supplementary materials.***


**15. SUBJECT TERMS: Data Management, Data Stewardship, Training Materials**

| **16. SECURITY CLASSIFICATION OF: U** | | | **17. LIMITATION OF ABSTRACT** | **18. NUMBER OF PAGES:** | **19a. NAME OF RESPONSIBLE PERSON: Douglas Clark** |
|---|---|---|---|---|---|
| **a. REPORT**<br>U | **b. ABSTRACT**<br>U | **c. THIS PAGE**<br>U | **U** | **47** | **19b. TELEPHONE NUMBER**<br>**303-445-2271** |

S Standard Form 298 (Rev. 8/98)
P Prescribed by ANSI Std. 239-18

# PEER REVIEW DOCUMENTATION

## Project and Document Information

Project Name: Fostering Data Stewardship Best Practices WOID: Y3840

Document: Training Materials for Data Management in Reclamation

Document Author(s): Douglas Clark and Barbara Simpson

Document date: June 2016

Peer Reviewer: Karri (Jan) Oliver

## Review Certification

**Peer Reviewer:** I have reviewed the assigned items/sections(s) noted for the above document and believe them to be in accordance with the project requirements, standards of the profession, and Reclamation policy.

Reviewer Karri J. Oliver Date reviewed 06/10/16
(Signature)

# Acknowledgements

The authors are grateful to James Nagode for his advice on this project and for his unflagging efforts in behalf of data management in Reclamation.

The authors would also like to acknowledge the support of the Data Stewardship Core Team, sponsored by Reclamation's R&D Office. This team has worked for many years, under the guidance of Curt Brown and Jim Nagode, to focus agency attention on the vital importance of data management.

# Executive Summary

In the wake of the Department of the Interior Data Management policy initiatives and the Open Data Executive Order Reclamation has been embarking on implementation of data management best practices within the agency. The USBR Data Stewardship Core Team (DSCT), active for nearly a decade, saw development of training materials as an important first step.

Ms. Barbara Simpson, a member of that team, regularly conducts data management training using three training modules developed by the U.S. Geological Survey. Her students have found the modules to be very informative, as has the DSCT. The team determined, however, that supplementary Reclamation-related materials would help make the concepts of data management provided in the modules more intelligible to USBR users. This document provides those supplementary materials which include:

- Discussion of the Reclamation data lifecycle
- A Reclamation data management plan template
- Examples of Reclamation data management best practice
- Lessons learned from various USB R data management efforts.

In summary, then, the USGS training modules supply vital data management information, but supplementary materials help to make their content relevant to Reclamation users. At some point, Reclamation may wish to develop its own set of training materials, tailored specifically to the needs of its scientists and engineers.

# Contents

# Tables

*Page*



# Figures

*Page*

# Background

Data management is the development of a framework to organize information and to implement policies, practices, and procedures that manage that information and its retention lifecycle. In his Data Management course "It's All About Data" for The Department of the Interior (DOI) Bureaus, Bureau of Land Management (BLM) trainer, Mr. Tom Chatfield that, teaches to successfully implement data management, an agency do the following:

- Gain upper management support of data management,
- Collect agency success stories that show the benefits of successful implementation,
- Create a sound data management policy,
- Develop a high quality training program.

Reclamation recognizes the value of sound data management practices; and is now introducing a high-quality training program to achieve and support a sound data management policy. Ms. Barbara Simpson, GIS Coordinator for Reclamation's Mid-Pacific Region, has successfully used existing U.S. Geological Survey data management training modules to inform Reclamation staff in her region concerning the fundamentals of data management. These modules include:

- The Value of Data Management
- Planning for Data Management
- Best Practices for Preparing Science Data to Share

Reclamation finds it reasonable to pursue this initial training program, given Ms. Simpson's success using these modules. Doing this will involve:

- Making use of the existing materials in the modules where appropriate.
- Adding additional materials suitable for Reclamation personnel where required.
- Revising sections of the USGS materials that are out of date or not suited to Reclamation scientists and engineers.

Members of the Reclamation Data Stewardship Core Team have undertaken this work. As a first step they have asked members of the team to review the modules and make suggestions for improvement. What follows is a summary of comments made for each of the modules along with materials that two members, Douglas Clark and Barbara Simpson have added.

# The USGS Training Modules and Reclamation Additions

## Module I: The Value of Data Management

### The Need for Data Management.

Prevention of data loss is one of the key reasons for instituting data management best practices in an agency. USGS Training Module I lists several reasons for *data loss*, including natural disasters, hardware failures, and human error (e.g. deleting the wrong file). Other causes that might be mentioned include: changes in storage media technology (tapes, zip drives, optical drives, DVDs, hard drives), corruption of storage media, loss of storage media, personnel retirements, etc.

Reclamation has encountered other problems beyond data loss, and so recognizes the value of good data management practice as highlighted in Module I. Lack of data management planning is one of these. For instance, one USBR leader of a river restoration effort has noted that unless data collections are comprehensively planned, project scientists may collect whatever data they *think* are relevant to the project using personally favored protocols, coding schemes, and quality assessment and control procedures. Unfortunately, these may ultimately be insufficient for the project goals. Such unplanned efforts may result in a disparate set of data collections gathered by various individual that have limited or no utility, because common standards, practices, and protocols were not in place when the collection began. For this reason, it is incumbent upon project managers and data stewards to ensure data collection best practices are in place from the beginning, so that data collected consistently meets the project needs.

### The Role of Scientists in Data Management

Module I then discusses the reasons why some scientists do not practice data management. A scientist may fail to manage data because she is preoccupied with her research, believing that she will "get to data management later." Or a scientist may fear that her data will be misinterpreted if released. Reclamation scientists have reported facing similar issues, which have, at times, blocked the sharing of data. However, *collaborative* data planning efforts have proven successful in alleviating this problem. For instance, according to Eric Peterson, the data manager of Reclamation's Trinity River Restoration office:

> "The collaborative *development of a master data management plan* substantially reduced concerns over misinterpretation. The meeting to draft this plan built trust and assured that data creators would be respected. In the course of the drafting meetings it was recognized that the scientific community regularly deals with differing interpretations. Deliberate mis-interpretations were recognized as unprofessional, usually obvious, and potentially damaging to a career. Accidental mis-interpretations were also seen as unprofessional and damaging. In other words, it is in the enlightened self-interest of scientists to take great care in their use and

> interpretation of acquired data. Alternate valid interpretations are a standard challenge in science and generally work out over time as further data become available. So, science itself contains a self-policing function. Within the data plan itself, the foundation of professional and respectful use of another's data was set forth, which did not, at the same time, hand-cuff data recipients from having alternate interpretations."

Reclamation scientists have reported "version control" is a problem that sometimes makes them reluctant to actively share data. As data sets or draft data sets are circulated among scientists, they are subject to change even while the original data sets are undergoing revisions; and thus, the final dataset may be hard to recover, identify, or compile. The Reclamation Trinity River Project data manager also found a solution for this problem:

> "development of a data portal provided a resource for all program participants to use to obtain the most current datasets. The program only posts completed data sets called "data packages", which include metadata documentation that clearly specify the "who, what, when, where, why, and how" of each data set and provide specific information about version."

## The Data Lifecycle

As Module I points out, scientists are in the best position to know and understand the intimate details of their data including its lifecycle; since they have either created the data or worked with it intimately. Inherently, they are most able to provide accurate data documentation. "They will also have the best idea of the lifecycle of the data, and are the logical choice for outlining a data management plan" (Module I).

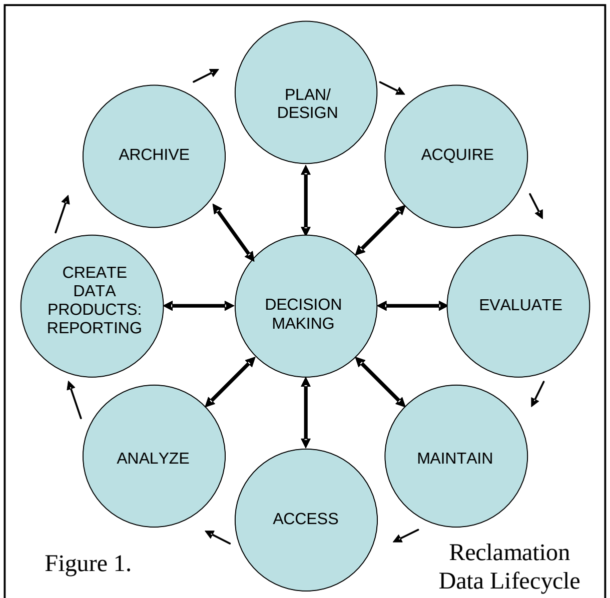



Figure 1. Reclamation Data Lifecycle

No graphic of the data lifecycle appears in any of the three modules. Lifecycles often vary by agency. The Reclamation data lifecycle is illustrated here (Figure 1), and could well be the basis of some discussion during a data management training session. This diagram is particularly designed for an organization that has more projects than programs. Projects are typically designed to serve specific mission and decision requirements. For this reason, decision making is at the center of the graphic. Data collection processes begin with a *Plan/Design* for each stage of the lifecycle:

- Acquire: What standards and protocols will be used for the actual data collection?
- Evaluate: How will the data be quality checked and *evaluated* to determine if it served the original decision requirement?
- Maintain: How will data integrity be *maintained*?
- Access: Who will have access to the data and under what conditions? How will privacy and security be protected?
- Analyze: How will the data be processed, or analyzed?
- Create Data Products and Reporting: What information products will be made available, or reported?
- Archive: Finally, how will the data be *archived* after the original purpose has been served?

These lifecycle phases apply to virtually every scientific data collection that Reclamation undertakes. These questions should be answered in a document called a Data Acquisition and Management Plant (DAMP) (Appendix A). The Information Management Group within Reclamation has a template for the preparation of a DAMP, which can be found in Appendix A. This template describes in detail what information is required for a DAMP. There will be a discussion of the Reclamation DAMP in the section on teaching Module II.

## Roles and Responsibilities for Data Management

There are a variety of roles and responsibilities associated with good data management practice, as Module I points out. For instance, data stewards "are those business subject matter experts who manage another's facts or information to ensure that they can be used to draw conclusions or make decisions." Since Reclamation is largely a project driven organization, but also manages various programs, it likely will need two distinct types of data stewards:

- Project Data Stewards – responsible for organizing and managing the data associated with a specific project
- Business Data Stewards – oversee and manage the data in a large data domain (such as water, power, real property, endangered species, dam safety, lands, facilities, etc.)

## The Benefits of Data Management

After discussing the various roles and responsibilities for data stewardship, Module I proceeds to discuss the *benefits* of managing data assets. Among these benefits are:

- Protection of the initial investment to capture and maintain the data collection.
- Creation of metadata and documentation makes the ongoing science transparent.
- Ensuring data documentation – often required for publishing in Peer-reviewed journals.
- Management and documentation make data more legally defensible
- Data management is now required by law and must be shared unless there are privacy or security concerns.

- Well-documented data can be re-used, or put to new uses, within the Big Data global market.

Reclamation has found other benefits that might be added to the list above:

- Data management ensures data is of sufficient quality for decision-makers to use in their deliberations.
- Data management helps to:
  - protect agency reputation by conserving critical information,
  - protect data assets, that if lost could result in failure of one or more business operations,
  - ensure continuity of agency operations in the event of disaster,
  - and protect vital resources such as infrastructure, endangered species, real property, water, and power.

## Data Management Case Studies

Module I introduces a case study referencing a flawed data set associated with whirling disease. A DOI employee wanted to re-publish existing data before it was taken offline because of funding cuts. She went to great lengths to obtain a copy of the data. When she did finally acquire a copy, she found significant problems using the data because of inconsistent labeling, undefined abbreviations, and undocumented sampling locations.

Similar Reclamation case studies might be researched and mentioned. Reclamation has struggled for years to develop definitive sets of locational data for its lands, its facilities, and its projects. This has been a particular problem because, during an emergency, Reclamation law enforcement needs an accurate description of where the agency's jurisdiction begins and ends. In addition, Reclamation often receives Congressional inquiries regarding what lands Reclamation owns. Different offices have had inconsistent answers to this question. The agency also receives inquiries requesting estimates of the amount of water being diverted for Reclamation projects. When project boundaries are not clearly understood, this can be difficult to determine. Reclamation offices have received queries as to where their canals are located in particular areas and how many of them run through populated and/or urban areas. These inquiries often require substantial effort to fulfill because of dated, illegible, or conflicting records, legal disputes, and/or poorly documented routing changes.

## Justification for Data Management

Module I posts the following key reasons why data should be managed:

- To maximize the effective use and value of data and information assets.
- To continually improve data quality including: data accuracy, integrity, currency, relevance, and overall usefulness.
- To ensure appropriate use and avoid improper use.

- To ensure sustainability and accessibility for long term re-use in science.

Reclamation trainers could add the following justifications to the list:

- To track the origin, chronology, or custody of the data, for example, changes in "as-built" blueprints.
- To make it possible to integrate existing data sets with others to create new data and information.
- To protect data integrity, completeness, authenticity, and reliability of data sets.
- To document the strengths and limitations of the data set for other uses.
- To provide transparency to users of the data whether they are researchers, decision-makers, managers, policy analysts, or funders of the data collection.
- To make the data shareable.
- To avoid duplicative data collection efforts.

## Module II: Planning for Data Management

### The Data Management Plan

According to USGS Module II, a data management plan is "a formal document that outlines data management considerations before work begins and how data will be handled during and after completion of the research." Planning the data collection saves time because, "It is easier to plan how data will be acquired and managed through the project lifecycle at the outset as opposed to trying to organize or retrofit data after they have been collected or acquired." Planning for a data acquisition also recognizes that data are an investment that must to be protected. In addition, planning helps to ensure that the data collector and the data consumer will be able to understand and make optimal use of the data set. Finally, planning often helps satisfy funding area requirements.

Reclamation has its own data management plan template or outline. The Reclamation Data Acquisition and Management Plan template (See Appendix A) was developed by the USBR Data Stewardship Core Team. The plan should name its author, the steward in charge of the data collection, and the date. Background is to be provided to identify the intents and purposes of the collection, i.e. what questions will be addressed or what business requirements will be served. The author will provide the design of the project which will describe how the data acquisition is to proceed (sampling protocols, sensors, data coding, etc.). The data steward will describe how data quality is to be controlled, assessed, and assured at each phase of the lifecycle. The following items are also to be addressed:

- How will data integrity be maintained during and after the initial collection?
- Who will have access to the collected data and how will data be protected or secured? How will personally identifiable information be protected?
- How will the data be processed, analyzed, and reported?
- What information products will be created?

- Finally, there should be a description of how the data will be archived after the original purpose for collecting it has been served.

The following list provides short descriptions of the terminology, elements, phases, and purposes in order to provide an explanation of the Reclamation data lifecycle cited above (See Figure 1),. Each of these elements should comprise a separate section in the Data Acquisition and Management Plan (See also Appendix A).

**Plan/Design**: This is the planning step that drives all subsequent steps. The plan/design phase provides characterization of data and information requirements of the data collection based upon the business-driven user needs, including any scientific or technical questions that must be answered to support sound decision-making. It also establishes the relevant laws, policies, standards, and documentation requirements. This phase sets the requirements for data acquisition, access, and sharing.

**Acquire**: The acquire phase describes both the collection of data from existing sources and all new field or sensor data collections. The data acquisition plan translates the business/decision needs into variable definitions, mission plans, sampling plans, collection methods, logging/recording techniques, and the like. It documents specifications, collection protocols, QA/QC procedures, and other requirements for all collected or purchased data.

This section of the DAMP should include an inventory to identify both existing information assets and data gaps. For existing data sources, it should identify the metadata needed to document the information obtained. This stage includes a review of relevant existing data standards, leading to adoption of existing standards or the potential development of new standards. It should identify any conversion or transformations of the data required for Reclamation use. Often missing from this component is a statistical power analysis, which determines the required number of data points to be taken to achieve a desired level of statistical confidence. This analysis matches collection requirements with decision requirements.

**Evaluate:** The evaluate phase of the lifecycle ensures the data collected and processed will meet the original requirements for objectivity, utility, and quality established in the Design and Acquire phases. This phase includes a periodic function to obtain feedback on data usage and analysis to determine their continued usefulness for the intended information and decision requirements.

**Maintain:** The maintain phase describes the ongoing processes, procedures, and systems for data storage, control, maintenance, security, and manipulation to ensure that the data can provide information to meet the business requirements. It includes procedures to maintain data integrity, conduct appropriate updating, and produce applicable documentation. Typically, this step involves creating a database system where the data and metadata are stored; and basic data analysis and restructuring may be performed.

**Access:** The access phase develops procedures to ensure that the data acquired are known to and retrievable by the designated user community, using documentation and discovery best practices. Processes are established to identify who can create, modify, delete, discover, view, query, download, copy, update, and report data.

**Analyze:** The analysis phase describes the analytic methods and processes that will create information products to meet the business requirements and/or to successfully address the needs of the decision-maker.

**Report:** In the reporting phase, the information products are agreed upon to meet the business and/or legal requirements. These products will be provided to decision-makers, managers, stakeholders, and/or the public.

**Archive:** The archival phase describes the data's retention and retirement into long-term storage within the agency. "Archive" refers to the disposition of data after the original business requirements have been met and the need for access will be infrequent. Once the data are archived, continued discoverability and access will be maintained. Access and maintenance processes for long-term storage are also described in this phase.

Returning to Module II, additional practical guidance is provided as to how to get started with data management. For instance, data flow diagrams can be useful for visualizing the process through which new data are gathered, processed, analyzed, and transformed into a data product. Figure 2 portrays data flows for deploying previously quality-checked data into the Reclamation DataSpace Console. Module II points out that flow models can be used to graphically connect the steps and processes required for collecting, preparing, transforming, and integrating data over the course of the project lifecycle.

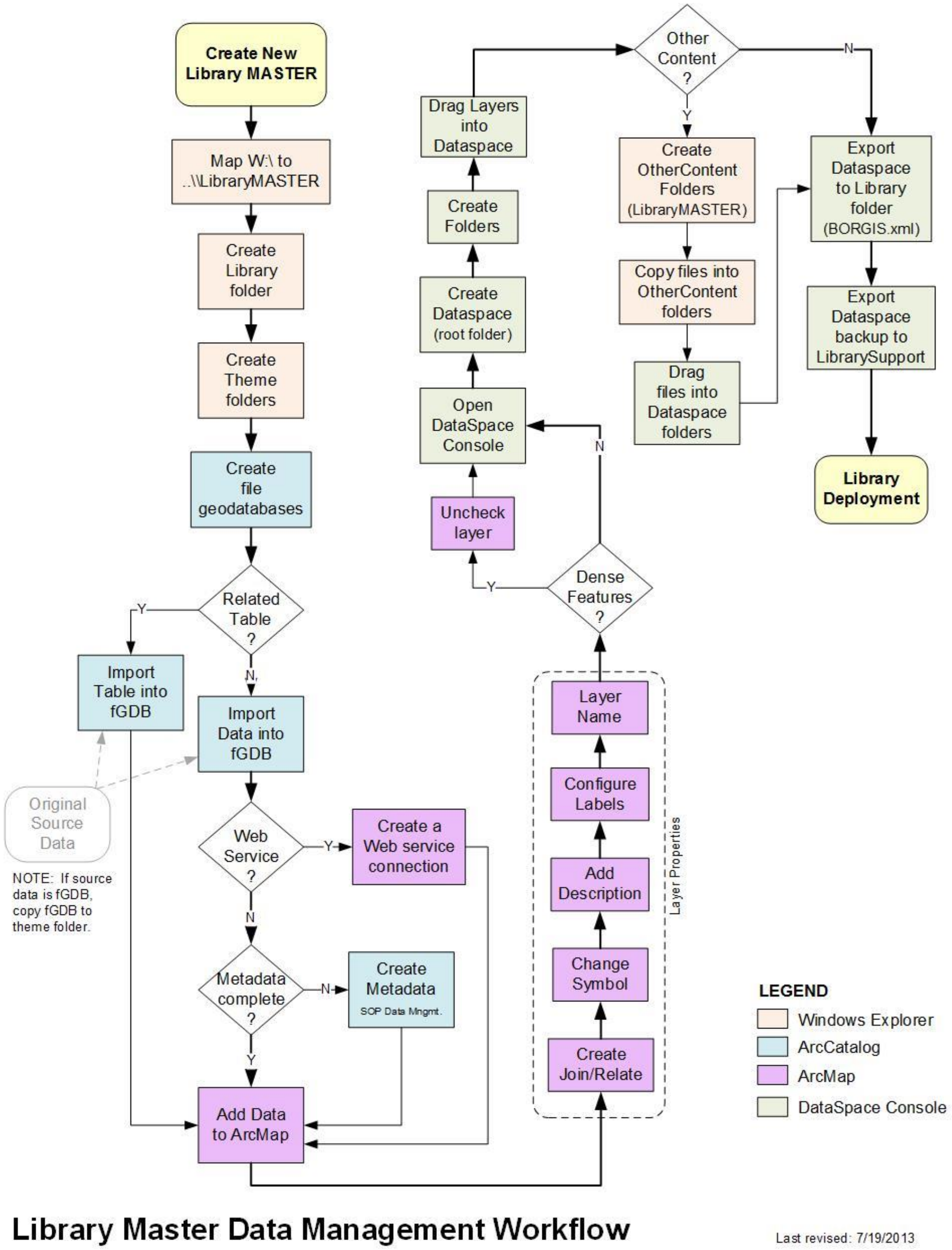



**Figure 2: Library Master Data Lifecycle/Flow for placing data into DataSpace. Source: Gregory Gault, PN Region**

Module II notes that the data management plan should, at a minimum, set forth the dataset name, description, applicable protocols, and required software. Other important information should include relevant business rules, spatial scope, format information, description of access, sharing, or reuse policies, and constraints such as scale limits for geospatial data or temporal scope. For the Reclamation DAMP this work would be done in the Design/Planning phase of the data lifecycle.

In addition, a conceptual model can be useful for visualizing the scope of the project. This would identify the primary data sets to be used and the interlinkages among them. Again, this work would be accomplished in the Design/Planning phase of the data lifecycle.

Also, a data documentation protocol would be laid out in the USBR Design/Planning phase of the lifecycle. As Module II indicates, this would define what information will be included in the metadata, how it will be captured (e.g. lab notebooks, GPS units, web forms, data logger, manual entry into a software package, etc.), and what format and standards will be used. In this section, the data managers will confront the question of what documentation must be required to make the data set(s) useable both in the present and in the future.

During the Planning, Access, and Create Data Products phases of the Reclamation Data Lifecycle (Figure 1), the data managers will address the following issues set forth in Module II:

- What are the obligations for data sharing: contractual, legal, institutional, etc.?
- How will data access be provided?
- What ethical and privacy issues may exist and how will these be addressed?
- How will intellectual property and copyright issues be addressed
- How should the data be cited when they are used?

In the Design/Planning and in the Archival sections of the Reclamation Data Acquisition and Management Plan the following issues will be addressed:

- What data will be preserved and for what term?
- Where will the data be preserved? What repository will be used?
- What metadata will be submitted along with the data sets?
- Who will be assigned the responsibility for preparing the data for preservation?
- Who will be the main contact person for data archive?

The final issues that must be addressed in the Design/Planning phase of the DAMP, are the budgetary questions surrounding the data acquisition, processing, documentation, hardware, software, personnel, and archival. This topic is covered in Module II. .

Obviously, the entire data collection and management team should be thoroughly briefed on the data management plan and the processes outlined therein. Module II also points out that documentation of the data lifecycle process works best if it is undertaken early on and supplemented on a regular schedule.

# Module III: Best Practices for Preparing Scientific Data to Share

## Data Management and Good Science

USGS Training Module III reiterates why good data management is crucial to good science:

- Allows data from the current project to be integrated successfully with data for other projects, both in the present and in the future;
- Allows others to use the data to test the reproducibility of the scientific results;
- Enables subsequent users to understand what was done in the project and why;
- Facilitates sharing;
- Establishes the origin, ownership, and history of the data provenance.
- Promotes scientific integrity by facilitating rigorous peer review.

## Living Databases

Module III explorers the utility of *living databases* – databases that are subject to orderly change as new data become available. Two case studies are examined, one for tamarisk and one for python habitat in the western hemisphere. To make the inclusion of new data and information possible for a data collection, it is essential to track the methodologies and standards used at each step of the evolving process:

- What scientists were involved?
- During what timeframe?
- What computer programs, algorithms, and other tools were used?

Careful documentation during the entire course of the project facilitates the justification of project conclusions and defense of results.

Reclamation follows such processes in its flood inundation modeling work. These might be cited in the training class. The agency regularly conducts inundation modeling of potential failures of its dams, canals, and other facilities. Modes of failure, methods used to compile topographic data and model surfaces, methods used to model resulting floods (1-D, 2-D, etc.) are all carefully documented. This documentation helps modelers in future years to evaluate previous results or processes and make subsequent improvements when and if new data and/or technology become available. Taking such steps allows science and engineering to progress in an orderly fashion over time. The selection that follows is from a modeled dam breach inundation study document. (The names of the dam and reservoir have been removed). Figure 3 shows a detail from a flood inundation study. The modeled flood boundary is in red.

***Hydraulic Models***

*The flood inundation analysis was completed using the Danish Hydraulic Institute's (DHI) [4] MIKE11 one-dimensional (1D) hydraulic model, and the DHI MIKE21 two-dimensional (2D) hydraulic model. 1D modeling was used to simulate the dam breach outflow and to route flows for the lower portion of the flooded reach from river mile 53.4 to the end of the study at the XYZ Reservoir inlet. 2D modeling was used to route flows from XYZ Dam to river mile 53.4. The reason for 2D modeling in the upper reach is related to the highly braided nature of the floodplain, for which the 2D modeling produces more reliable results. Breach simulations were performed using the National Weather Service DAMBRK formulation [5], which is available within the MIKE11 1D hydraulic model. Terrain data for the study area came from Interferometric Synthetic Aperture Radar (IFSAR) terrain data with a native 5-meter resolution. The terrain data was re-sampled to a 15-meter resolution for the 2D modeling, due to the large extent of the model and current computer modeling hardware limitations. The 1D reach used the data in its native 5-meter resolution.*

**Figure 3: Detail section of Reclamation modeled flood boundary**

## Short and Long Term Benefits of Data Management

Module III asserts that there are both short and long term benefits for good data management practices. In the short term, scientists and engineers making use of previously created data can more readily understand and use the previously created files, and therefore, they can accomplish more work toward their goal and not lose time chasing after missing information. There is another value-added advantage-- good data management and documentation expedites good peer review.

In the longer term, good data management and documentation allows external scientists and engineers to find, understand, use, and credit historic data – which protects sponsor investment.

Reclamation also could reap the substantial benefits from good data management:

- Reclamation stakeholders and the public could more readily understand what data were used in a project and how they were processed.
- Those using O&M data can have confidence that what is recorded is accurate and reflects what is actually on the ground.
- Decision makers can confidently justify the actions that they propose to take.
- Scientists and engineers using monitoring data for adaptive management can have a solid basis for guiding subsequent river operations.

## Nine Data Management Best Practices

Module III details nine fundamental data management best practices:

1. Define the contents of your data files.
2. Use consistent data organization.
3. Use stable file formats.
4. Assign descriptive file names.
5. Preserve processing information.
6. Perform basic quality assurance.
7. Provide documentation.
8. Protect your data.
9. Preserve your data.

### *Practice 1: Define the Contents of Your Data*

The module lists several straightforward principles and processes for defining the contents of a data set. This is usually done by creating a data dictionary. Best practices include:

- Use commonly accepted parameter names, descriptions, and units.
- Explicitly state the units of measure.
- Choose a format for each parameter and explain it.
- Use ISO formats wherever possible.

### *Practice 2: Use Consistent Data Organization*

Consistent data organization facilitates data integration. For instance, avoid changing or re-arranging field columns. Include header rows (first row contains file name, data set title, author, date, and companion file names). Column headings should describe the content of each column. Each row should include a complete data record.

Figure 4 shows a typical database of businesses with associated latitude and longitude locations used to determine whether an enterprise might be within a modeled flood boundary.

**Table 1: Typical USBR database for the study of flood impacts. Some businesses will be inside the flood boundary and others will be outside. Source: FEMA**

| NAME | CITY | STATE | ZIPCODE | OWNER | FUNCTION | LAT | LONG | COUNTY |
|---|---|---|---|---|---|---|---|---|
| ALTAMONT | ALTAMONT | UT | 84001 | ANR PROD CO | PROD PLANT | 40.358333 | -110.2875 | 49013 |
| BLUEBELL | ROOSEVEL | UT | 84066 | GARY ENERGY CORP | PROD PLANT | 40.380556 | -110.075 | 49013 |
| CEDAR RIM | DUCHESNE | UT | 84021 | KOCH HYDROCARBON CO | PROD PLANT | 40.205556 | -110.5625 | 49013 |
| LISBON #28 | MOAB | UT | 84532 | UNOCAL CORP | PROD PLANT | 38.1625 | -109.277778 | 49037 |
| PINEVIEW | COALVILLE | UT | 84017 | UNION PACIFIC RES CO | PROD PLANT | 40.9375 | -111.1375 | 49043 |
| YELLOW CREEK | COALVILLE | UT | 84017 | UNION PACIFIC RES CO | PROD PLANT | 40.991667 | -111.045833 | 49043 |
| MUSTANG NGL PLT | BLANDING | UT | 84511 | SAMEDAN OIL CORP | PROD PLANT | 37.625 | -109.375 | 49037 |
| CHALK CREEK | | UT | 84036 | MOUNTAIN FUEL RESOURCE | COMP STAT | 40.666667 | -111.166667 | 49043 |
| CLAY BASIN | | UT | 84023 | MOUNTAIN FUEL RESOURCE | COMP STAT | 40.981944 | -109.205 | 49009 |
| COALVILLE | | UT | 84036 | MOUNTAIN FUEL RESOURCE | COMP STAT | 40.666667 | -111.166667 | 49043 |
| BRIDGER LAKE | | UT | 84055 | PHILLIPS PETROLEUM CO | COMP STAT | 40.965 | -110.32 | 49043 |

Each row is a complete record and each column heading describes the data within that field.

Module III presents an example of poor data organization. It shows a table that has columns without headings, missing data, unlabeled numbers in inconsistent places, undefined units of measure, and notes that have meaning only to the author—certainly not suitable for subsequent data sharing.

### *Practice 3: Use Stable File Formats*

Data reuse depends upon the ability to find ("discover") and return to a data set, possibly years or decades after its creation. Best practice is to convert the data sets to "a well-documented, and non-proprietary format to maximize others' abilities to use [it]." Tabular data should be converted to ASCII or .csv format. Raster data can be converted to geotiffs, netCDF (preferably with the DF convention), HDF, and ASCII. Vector data can be converted to shapefiles and KML/GML. Module III poses one additional caution namely, protect the longevity of the data by saving it on a readable media. Remember floppy disks, zip-drives, and optical disks?

### *Practice 4: Assign Descriptive File Names*

File names should both reflect the contents of the file and uniquely identify it. File names may contain information such as the project acronym, study title, geography, data collector, time frame, data type, version, or file type. For instance, a project file for a Reclamation UAS flight might be structured as follows:

project_location_date_time_mission_sensor

In addition, following USGS procedures, a directory structure for the archival of USBR UAS data might follow this convention:

- PROJECT (name given to each project)
  - project.metadata
  - projectfiles
  - MISSION1 (name given each mission)
    - mission1.meta
    - missionfiles
    - COLLECT1 (name given each collect/flight)
    - >collect1.metadata
    - collectfiles
      - SENSOR
        - RAW DATA
          - datafiles
        - PTCLOUD DATA
          - datafiles
        - DEM DATA
          - datafiles
        - ORTHO DATA
          - datafiles
        - VIDEO DATA
          - datafiles
        - SUPPORTFILES
          - files
    - COLLECT2 (name given each collect) [same elements as above]
  - MISSION2 (name given each mission) [same elements as above]

(Please see USGS Unmanned Aircraft Systems Data Management Plan for further information).

Figure 4 shows a sample directory geospatial library structure for Reclamation DataSpace Console deployed in Reclamation Area Offices.

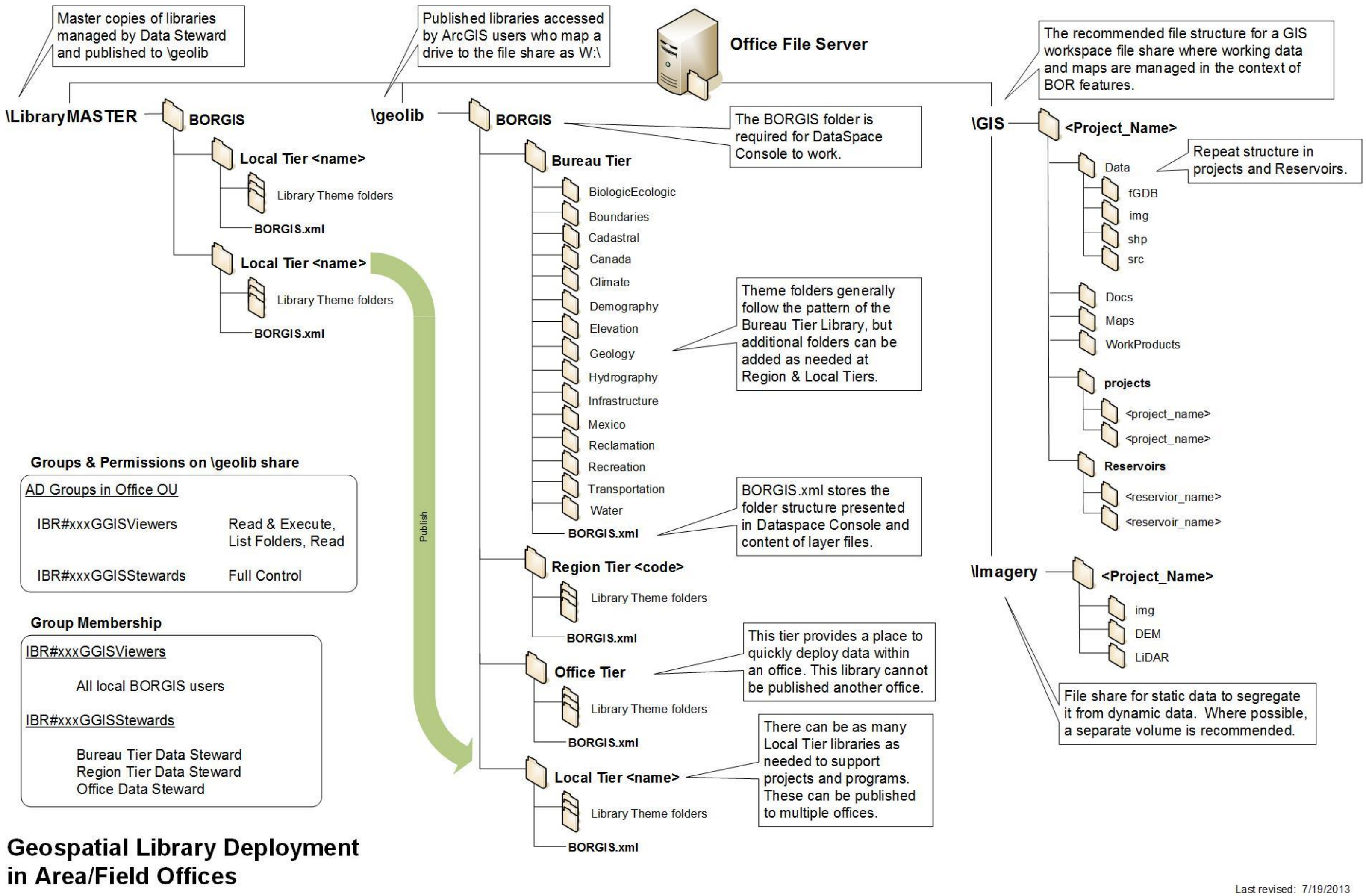



Figure 4: Geospatial Library Structure for DataSpace. Source: Greg Gault, PN Region

***Practice 5: Preserve Processing Information***

To preserve the original data and its integrity, scientists should save a read-only copy of the raw data files with no transformations, interpolations, analyses, or other manipulations. The USGS data team advises, “Use a scripted language such as “R”, “SAS”, or “MATLAB” to process data in a separate file, located in a separate directory.” These scripts are an excellent record of the data processing and can be quickly and easily revised and rerun in the event of data loss or requests for edits. They also allow for subsequent users to reproduce processing.

Within the Reclamation Remote Sensing and GIS community, ESRI Model Builder processes and Python scripts are used and can subsequently be saved and documented in the associated metadata. In addition, as shown below in Figure 5, ERDAS Imagine Model Maker processes can be documented for future examination and re-use. The script in the bottom part of the top panel is written in the software’s Spatial Modeler Language for the study of net radiation. Scripts like this become vital parts of the project documentation.

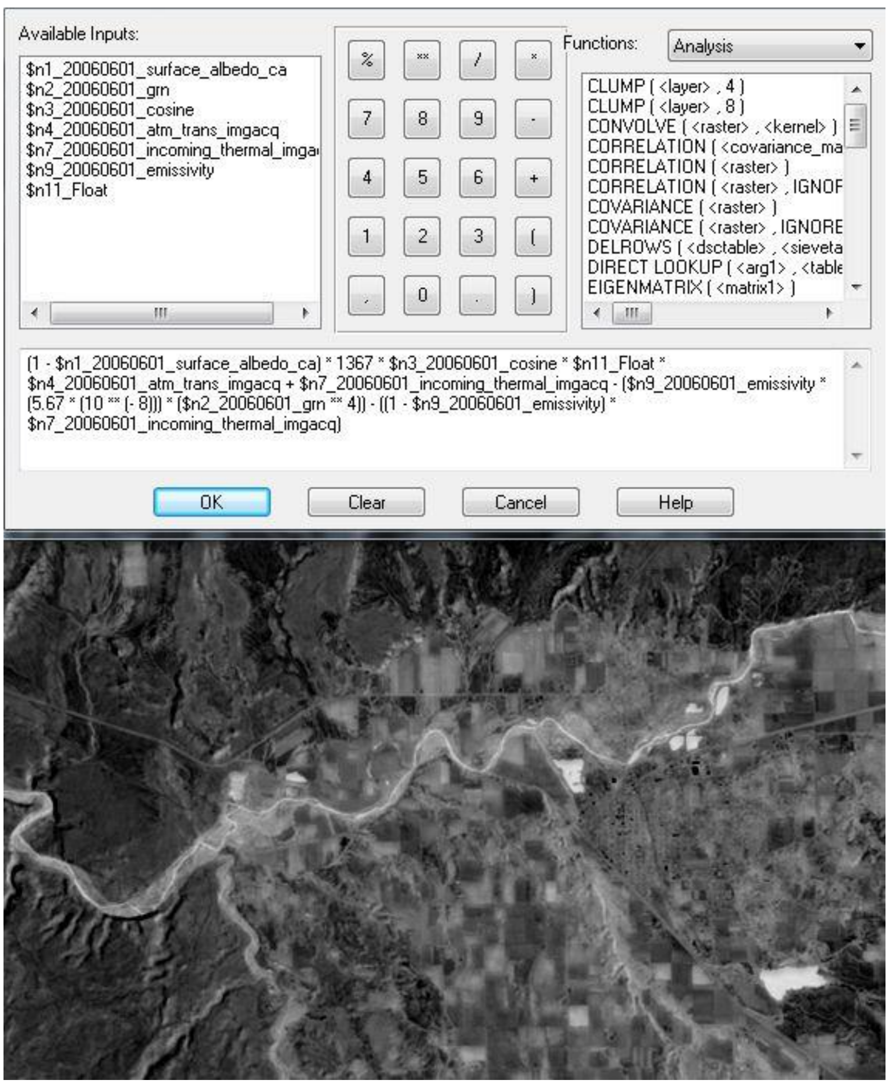



**Figure 5: ERDAS Spatial Modeler Language script for calculating net radiation. Scripts of this sort become an integral part of project documentation. Source: David Eckhardt, Denver, TSC.**

### *Practice 6: Perform Basic Quality Assurance*

Quality assurance checks should occur at every phase of the data lifecycle. During the Acquisition Phase, measurements should be checked and validated and sampling and other protocols carefully followed. Field audits and laboratory audits must be performed. Instrumentation should be carefully checked and calibrated. Data custody processes should carefully be followed. Special attention should be paid to keeping data within the database in proper columns and rows, detecting missing or outlying values, and checking for positional and thematic accuracy. In the Maintenance and Access Phases changes to the data should be carefully monitored, versions documented, security protected. Reclamation has developed *Quality Assurance Guidelines for Environmental Measurements* (2003) to guide the acquisition and management of earth science data. Reclamation also has a set of Information Quality Guidelines to ensure the "quality, objectivity, utility, and integrity of information (including statistical information)" that it disseminates. Finally, the Reclamation Manual also contains a policy on the Peer Review of Scientific Information and Assessments.

### *Practice 7: Provide Documentation*

Sound documentation practice requires keeping the formal metadata with the project data. Metadata, of course, describes the "who, what, where, why, when, and how" of the data in a project. This documentation becomes a "user's guide" for that data. As Module III indicates, it should be written for a person who may know nothing about the project, possibly decades into the future. And, as with the project data, metadata should be saved using stable, non-proprietary formats. Associated images, figures, and pictures should be stored as individual .gif or .jpeg files. Figures 6-8 show screens from tcEME, an EPA metadata editor modified for and frequently used in Reclamation. One screen is for gathering basic information; a second screen is for collecting data quality, spatial extent, and attribute information; and a third screen is for instructional information about distribution. Three metadata applications used in Reclamation are described and compared in the next section of this report.

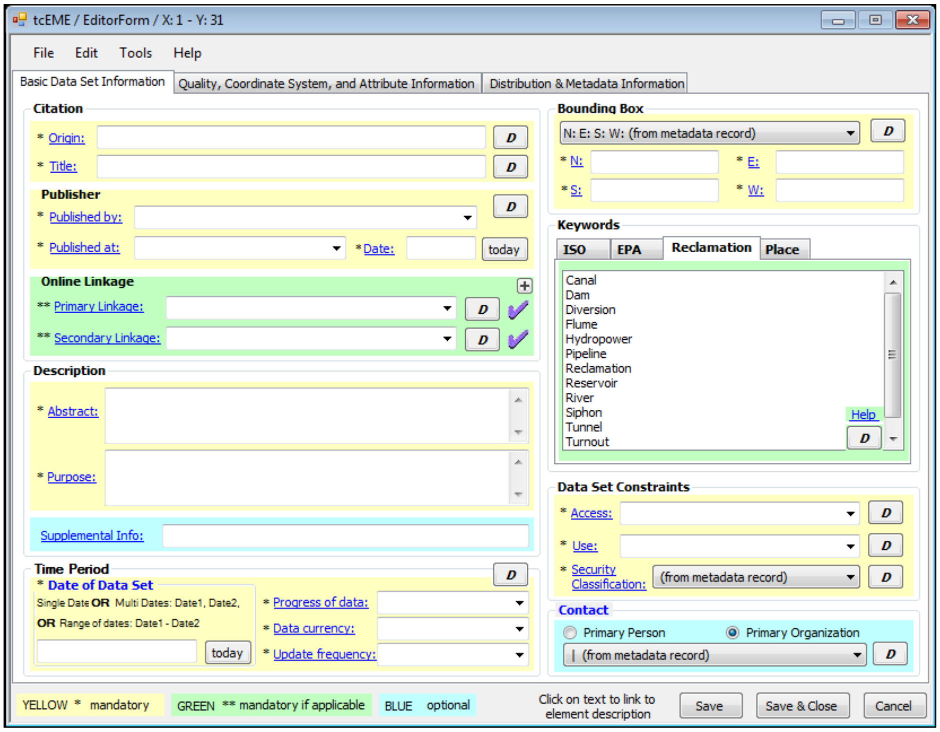



**Figure 6: tcEME Metadata Editor, Screen for basic information.**

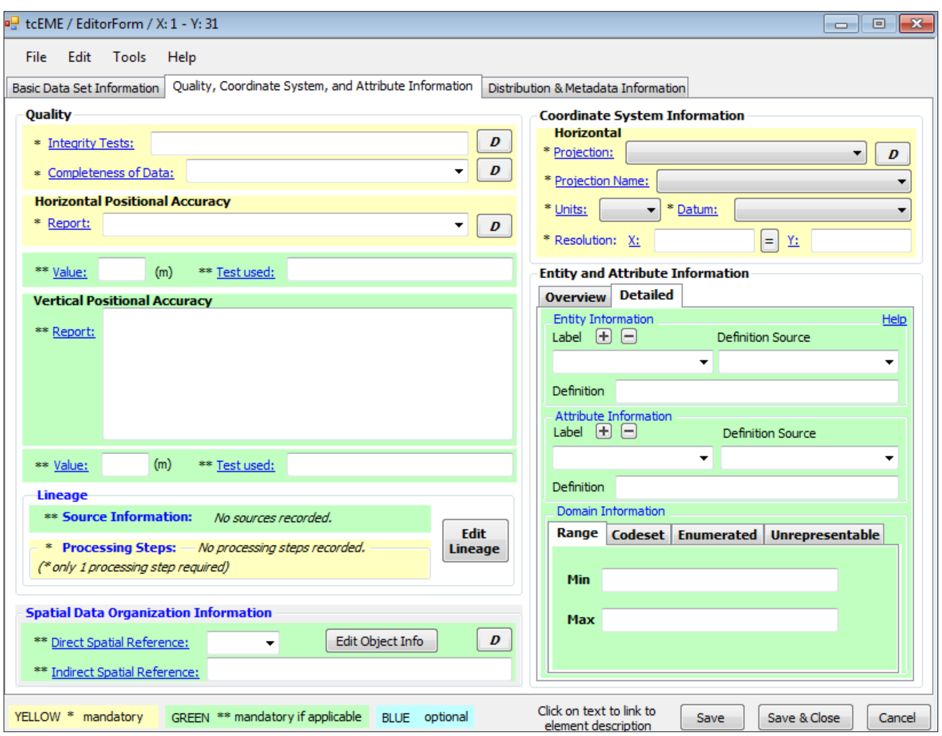



**Figure 7: tcEME Metadata Editor, screen for quality, coordinate, and attribute information.**

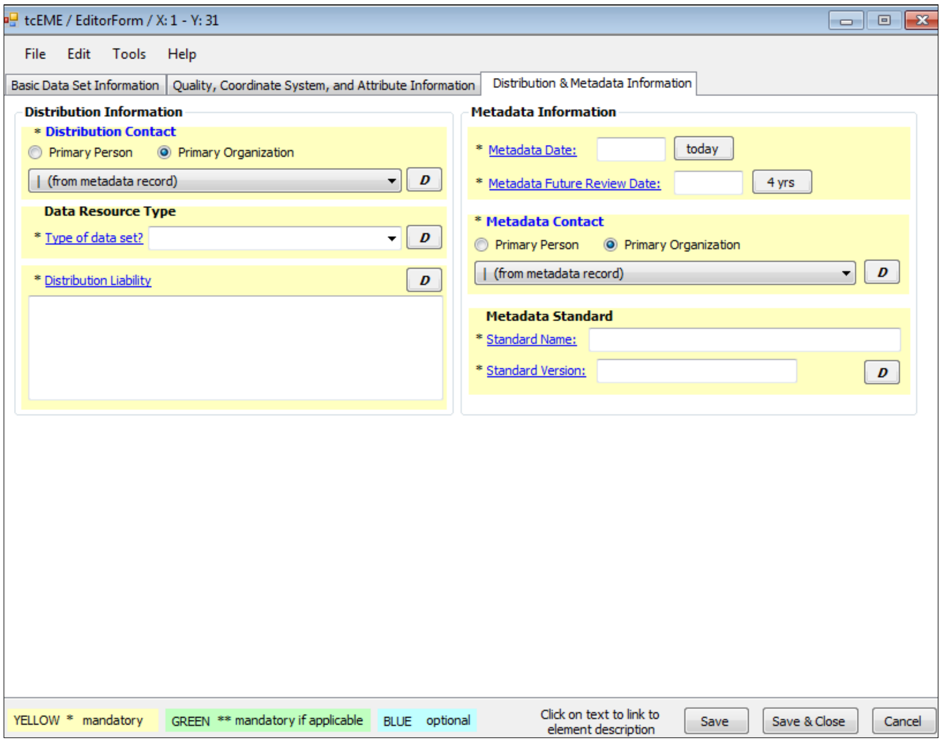



**Figure 8: tcEME Metadata Editor, screen for distribution and metadata information. (For more information, please contact Joel Murray at jmurray@usbr.gov.)**

### *Practice 8: Protect Your Data*

A best practice for protecting data is to maintain three copies of project data: the original, an on-site but external backup, and an off-site back up in case of a disaster. Cloud storage offers another alternative. Another best practice is to periodically test backups.

### *Practice 9: Preserve Your Data*

In general, well-structured and well-defined data files should be preserved. Units of measure should be documented. A data dictionary should be provided, along with metadata. Additional information "providing context is also desirable including materials from wiki/websites, files describing the project, protocols, or field sites, including photos, and any publications using the data."

A variety of options exist for archiving Reclamation project data. For engineering projects, the Reclamation Electronic Document System exists. Remote sensing data may be archived with the Earth Resources Observation and Science (EROS) Data Center. The National Archives and Records Administration is the official repository for Federal data. NARA Regulations are published in 36 CFR Chapter XII. Electronic records management is published under Part 1234 of the 36 CFR Chapter XII. Data used to make governmental decisions are especially important to archive.

# Metadata Editors: Cross-Comparisons

There are a variety of metadata editors available to use. What follows is a strength comparison between three that are most often used within the government: the USGS Metadata Editor, the USDA Metavist, and the EPA Metadata editor. A detailed, attribute by attribute tabular comparison can be found in Appendix B.

USGS Metadata Editor strengths are in collecting information about the database structure (it includes column names and a data dictionary), descriptions and purpose of the data set, key word information, geospatial reference information, data accuracy, instructions to facilitate final distribution of the data, publication of the data, and the most information about the project methodology compared with the two other editors.

The USDA Metavist Editor strengths are in collecting information about time frames of the data set, biological information (more so than the other editors), key word information, and data privacy/security information.

The EPA Metadata Editor strengths are in collecting information about the metadata standard, publication of the data, descriptions and purpose of the data set, geospatial reference information, and data accuracy.

# Discussion

The USGS Training Modules I, II, and III provide an admirable introduction to the fundamentals of data management for scientists and engineers. These modules could, however, be supplemented with applicable Reclamation content to make the data management training more relevant and intelligible to USBR personnel. This paper has provided some potential examples of that content. Obviously, other materials might also be added. Eventually, Reclamation may find it more useful to develop its own training materials however in the meantime, utilizing available USGS materials is a workable first step in making this vital information available.

# APPENDIX A

## *DATA ACQUISITION AND MANAGEMENT PLAN (DAMP) TEMPLATE FOR LARGE-SCALE COLLECTIONS OF MISSION-CRITICAL RESOURCES DATA*

**DAMP TITLE:**

Plan author (name, title, email, organization code, phone).

Date.

Data steward for this data collection (name, title, email, organization code, phone).

Background of data collection effort.

**DEFINE/PLANNING PHASE OF THE DATA LIFECYCLE (Figure 1).**

This section identifies the reasons and purpose for the data collection. This should be based on statutory/regulatory requirements, business needs, or a future decision that must be made. The purpose of this phase is to establish with managers and decision makers the principal information and data needed to meet Reclamation requirements.

The data and information requirements should be specific about:

- The intended use and application of the data. Describe briefly the principal management questions or business objectives that the data will need to address.
- The level of resolution, quality, and accuracy of the data for the application and use (data quality objectives).
- Concerns about access or use of the data and products of analysis external to Reclamation.
    - Identify any portion of the data or analysis that may be **sensitive** or **proprietary** in nature.
    - Identify any portion of the data or analysis products that will publicly available or served to the public.
- Identify the intended users of the data and products of analysis.

There are a variety of other items that that should be considered during the Planning/Design phase. While not required in this template, it is good practice to document the disposition of these items. A checklist is provided below:

- ☐ Acquiring appropriate data management expertise.
- ☐ Acquiring expertise in the acquisition of data management software and hardware.
- ☐ Understanding the laws and regulations that pertain to the data collection.
- ☐ Identifying relevant business rules.
- ☐ Determining the spatial and temporal scope of the project.
- ☐ Obtaining collaboration, buy-in, and support for the data management effort from partners, researchers, principal investigators, advisors, reviewers, and stakeholders.
- ☐ Managing personnel, equipment, and related budget issues.
- ☐ Building in program flexibility in the event that the project changes scope.
- ☐ Assigning specific responsibilities for the various aspects of data management in such areas as technology acquisition, data collection, data evaluation, access, archiving, etc.
- ☐ Developing clear data standards, protocols, naming conventions, and metadata. Making sure that all data collectors (contractors, partners, stakeholders, other agencies, etc.) are using the same standards, naming conventions, protocols, etc.
- ☐ Developing processes for possibly managing multiple databases and computers.
- ☐ Working with IT personnel to understand what resources will be necessary to maintain the data collection.
- ☐ Putting together a data management training program so that everyone involved in the data collection will be following a common set of standards and collecting data using common processes.

**ACQUIRE PHASE OF THE DATA LIFECYCLE. The purpose of this phase is to translate the decision or business requirements into a plan for data collection that will produce data of sufficient relevance, scope, accuracy, repeatability, and credibility.**

This section identifies the data acquisition plan to meet the objectives identified in the **Define Phase**. These requirements apply to new data as well as the acquisition of existing (legacy) data. As appropriate, this section should be specific about:

- Acquisition plan, sampling design, flight plan, mission plan, etc.
- Identification of primary data sets to be used and their interlinkages.

- Data acquisition protocols.
- Existing standards to be followed in data acquisition, sample handling, and data processing procedures.
- Metadata: standards, how it will be captured.
- Documentation requirements during the **Acquisition Phase** (chain-of-custody provisions).
- Managing legacy/historical data sets.
- Managing field access, permitting, and other difficulties such as traveling to remote or dangerous locations where the data will be collected.
- Making determinations about format.

**<u>EVALUATE PHASE OF THE DATA LIFECYCLE.</u> The purpose of this phase is to evaluate the actual data to ensure that they meet the objectives of the Define Phase.**

This section identifies the methods to identify the quality of the data acquired or collected for addressing the objectives of the **Define Phase**. This quality assessment should be specific for each category of data. This assessment should be as quantitative as possible, and should identify the level of uncertainty associated with a group of data.

- Methods used to evaluate data quality.
- Accuracy tests performed, if any.
- Documenting limitations in use of the data for meeting the objectives stated in the **Define Phase**.
- Procedure for the management of data gaps, outliers, and null values.
- Standards for the review process for ensuring that data continues to meet objectives or intended use, including data currency or timeliness.

**<u>MAINTAIN PHASE OF THE DATA LIFECYCLE</u>. The purpose of this phase is to ensure that systems and procedures are in place to protect the integrity of the data, to store the data and associated metadata securely, and to facilitate needed access, analysis, and reporting.**

This section identifies the maintenance requirements for the data and information to support access, analysis, and reporting. This section should be specific as to:

- Storage location or systems for the data and associated metadata.

- Security requirements for the data, including provisions for the prevention of data loss and recovery from data loss.
- Processes for updating or revising records, including documentation requirements for changes.
- Authorization of persons or positions performing functions of Create, Read, Update, and Delete for data.
- Processes for certifying completeness and accuracy of the content.
- Processes for archiving data.
- Identifying any system requirements.
- Providing a database manual with the following minimum requirements:
    - Contact information for the data steward or database manager.
    - Description of the database structure, user roles, security, instructions (for how to use the system: data entry, data management, etc.), standards (for documenting updates/changes to the system), and other relevant information.
    - Training requirements for system use.
    - Method for ensuring the integrity of data entry or data revisions.

**ACCESS PHASE OF THE DATA LIFECYCLE. The purpose of this phase is to define data access control procedures and system functions, which ensure that all parties have appropriate access to the data, which may range from full access to no access.**

This section identifies access rights for the data and information generated. A key consideration in determining data access or protection requirements is to first recognize which data contains sensitive or proprietary information, versus which data, or portions thereof, can be made available to the public (as identified in the **Define Phase)**. This section should address the following:

- The access control process, including associated identification of roles and responsibilities in providing access;
- Data access requirements for Reclamation, cooperators, and the public;
- Protocols for providing public access to data intended for release;

- Protocols for retention of, and security for protecting/limiting access to sensitive or proprietary data;
- Investigation of privacy and ethical issues that might impinge on the data set;
- Access rights for all users of the data (Read, Write, Delete, etc.) by category of user;
- Protocols for managing data requests;
- Protocols for data release (for example: inter-agency, available to the public, FOIA, or how the intellectual and copyright issues will be addressed);

**<u>ANALYSIS PHASE OF THE DATA LIFECYCLE</u>. The purpose of this phase is to ensure that the data can be reliably, credibly, and accurately converted into information that answers the decision or project objectives identified in the Define Phase.**

For each of the principal management questions or business objectives listed in the DEFINE phase, describe how the data will be analyzed or synthesized to meet the manager's information needs. Include known analytical standards or protocols that are a key part of the analyses. Describe how levels of accuracy or uncertainty in data or analyses will be portrayed.

**<u>REPORTING PHASE OF THE DATA LIFECYCLE</u>. The purpose of this phase is to ensure that managers and decision makers receive the information needed to support business requirements or decisions in a timely fashion with appropriate levels of documentation and explanation.**

Based on the objectives identified in the **Define Phase**, identify the expected principal reports or reporting methods. In addition, this phase should identify:

- The peer review process for information reported, if applicable;
- Audiences or users of information being reported;
- Processes for discovery of data and information, along with methods of access;
- Venues for reports, and information;
- The expected process for releasing information intended for public release and/or distribution;
- Protocols for objectively interpreting findings.

**<u>ARCHIVE PHASE OF THE DATA LIFECYCLE.</u> The purpose of this phase is to set forth the plans for retention or disposal of the data set.**

Describe the schedule and implementation plan for data, samples, and reports for retention/archive. This should consider the fate or disposition of data, samples/specimens acquired as well as the analysis performed and reports generated. Data formats and system requirements for accessing the data and results of analysis should be identified. Storage plans

should consider long term storage requirements based on archival standards. This section should identify:

- Reclamation's Records Management requirements for the data and reports.
- National Archive and Records Administration (NARA) requirements that may (or may not) pertain to this data collection.
- EROS Data Center requirements.
- Other relevant laws, standards, and policies that affect disposition and archiving.
- Reclamation rules for retaining/archiving that pertain to this data collection.
- Provision of contact information for the data collection archival site.
- Distilling which data will be preserved and for what term.
- Who will have the responsibility for preparing the data for archive?
- Who will be the contact person for information about the data set?

# APPENDIX B:

## Metadata Editor Cross-Comparisons:

## (Pages 30-32)

| Metadata Tool Name | USGS Online Metadata Editor (OME) | USDA Metavist | EPA Metadata Editor |
|---|---|---|---|
| Website | https://www1.usgs.gov/csas/ome/newForm.htm | http://www.nrs.fs.fed.us/pubs/2737 | https://edg.epa.gov/EME/ |
| **METADATA STANDARD** | | | |
| Metadata Standard Name and Version | | | Y |
| **DATABASE INFORMATION** | | | |
| Data Set Summary: Explanation of Column Values | Y | | |
| Reference for a Data Dictionary | Y | | |
| **DISTRIBUTION INFORMATION** | | | |
| Contact Name | Y | Y | Y |
| Contact Title | Y | | |
| Contact Organization | Y | | Y |
| Contact Position | Y | | |
| Contact Phone | Y | | |
| Contact Email | Y | | |
| Contact Address | Y | | Y |
| Author Name | Y | Y | Y |
| Author Title | Y | Y | Y |
| Author Organization | Y | | Y |
| Author Position | Y | | |
| Author Phone | Y | | |
| Author Email | Y | | |
| Author Address | Y | | |
| Data Can be Shared (Y/N)? Availability. | Y | | |
| Access Constraints | | Y | |
| Use Constraints | | Y | |
| Distribution Information: contact, order process, technical pre-requisites for use, dates available | | Y | |
| Data Set Credit | | Y | |
| Graphics | | Y | |
| Cross-reference information | | Y | |
| Type of Data Set: applications, clearing house, downloadable, geographic activities, geographic services, map files, offline maps, static map images,etc. | | | Y |

| PUBLICATION INFORMATION | | | |
|---|---|---|---|
| Metadata Tool Name | USGS Online Metadata Editor (OME) | USDA Metavist | EPA Metadata Editor |
| Publication Date | Y | Y | Y |
| Website or Other Location where more Information can be Obtained | | | Y |
| Abstract | Y | Y | Y |
| Is Data Set Part of a Series Y/N | Y | | |
| Data Published Y/N | Y | | |
| Status of Data (Planned, In Work, or Completed) | Y | | Y |
| Progress: Planned, In Work, or Complete | | Y | Y |
| Primary Location of Accessing the Data | Y | | Y |
| Frequency Data Will be Updated (eg. weekly, yearly) | Y | | Y |
| Updating: Continually, Daily, Weekly, Monthly, As Needed, None Planned, or Unknown | | Y | Y |
| **DESCRIPTION AND PURPOSE OF THE DATA SET** | | | |
| Description of the Data Set. | Y | | Y |
| Purpose of the Data Set. | Y | Y | Y |
| **TIME FRAMES INCLUDED** | | | |
| Beginning and Ending Time(s) Period of the Data | Y | Y | Y |
| Time Periods | Y | Y | Y |
| Single Dates, Multiple Dates, Range of Dates | | Y | |
| Gregorian Dates or Geologic Dates | | Y | |
| **KEY WORDS** | | | |
| List Project Science Key Words | Y | | Y |
| List Project Place Key Words | Y | Y | Y |
| List Project Stratum Key Words (eg. troposphere, benthic) | Y | Y | |
| List Project Temporal Key Words (eg. Holocene) | Y | Y | |
| **GEOSPATIAL INFORMATION** | | | |
| Spatial Organization (Finding the data without coordinates) | Y | | |
| Geographic EYtent (Description) and Bounding Coordinates & Elevations | Y | Y | Y |
| Indirect Spatial Reference: eg. county, address, PLSS, place | Y | Y | Y |
| Direct Spatial Reference: Type of Spatial Data (Point, Line, raster) | Y | Y | Y |
| Do the Data Contain a Horizontal Coordinate System (Y/N) | Y | | Y |
| Do the data Contain a Vertical Coordinate System (Y/N) | Y | | Y |

| ATTRIBUTE INFORMATION | | | |
|---|---|---|---|
| Metadata Tool Name | USGS Online Metadata Editor (OME) | USDA Metavist | EPA Metadata Editor |
| List Entities and Attributes (eg. Themes like road width, road type). | Y | Y | Y |
| Citation of Source Define Entity Type with Attribute and Associated Values | Y | Y | Y |
| **DATA ACCURACY** | | | |
| Attribute Accuracy Report | Y | Y | Y |
| Quatitative Accuracy Assessment | Y | Y | Y |
| Positional Accuracy Report | Y | | Y |
| Logical Accuracy and Completeness Report | Y | | Y |
| **Methods** | | | |
| Data Processing Model or Tool Citations | Y | | |
| Field and Lab Methods Descriptions | Y | | |
| Methods to Create Final Data Set | Y | | |
| Description of Data Processing Models or Tools | Y | | |
| Analytic Tools: e.g. spreadsheets, databases, statistical software, GIS software, models, etc. | | Y | |
| **BIOLOGICAL INFORMATION** | | | |
| General taxonomic Coverage | | Y | |
| Taxonomic Information | Y | Y | |
| Keywords/taxon | Y | Y | |
| Taxonomic Classification | Y | Y | |
| Taxonomic Classification System | Y | Y | |
| Native Data Set Environment | | Y | |
| **DATA SECURITY** | | | |
| Security Classification System | | Y | |
| Security Classification: Unclassified, Sensitive, Restricted, Confidential, Secrete, Top Secret | | Y | |
| Metadata access constraints, sensitivity classification system and class, and security handling. | | Y | |

# References

Gacke, Carolyn. 2015. USGS Unmanned Aircraft Systems Data Management Plan, Version 1. US Geological Survey. http://uas.usgs.gov/publications.shtml

QA/QC Implementation Group, Bureau of Reclamation. 2003. Quality Assurance Guidelines for Environmental Measurements. Bureau of Reclamation.

Viv Hutchison, Thomas Burley, Michelle Chang, Thomas Chatfield, Robert Cook, Heather Henkel, Carly Strasser, Lisa Zolly. 2013. USGS Data Management Training Modules. USGS Community for Data Integration. http://dx.doi.org/10.5066/F7RJ4GGJ

# Data Sets that support the final report

If there are any data sets with your research, please note:

- Share Drive folder name and path where data are stored:  N/A.  No data sets.

- Point of Contact name, email and phone:  Douglas R. Clark, drclark@usbr.gov, 303-445-2271

- Short description of the data:  (types of information, principal locations collected, general time period of collection, predominant files types, unusual file types.)  N/A, no data sets.

- Keywords: Data Management, Data Stewardship

- Approximate total size of all files:  N/A No data sets.