DATA MANAGEMENT POLICY AND PROCEDURES DEVELOPED FOR THE PCOR PARTNERSHIP'S BELL CREEK STUDY

Plains CO_2 Reduction (PCOR) Partnership Phase III
Value-Added Report

Prepared for:

Andrea M. Dunn

National Energy Technology Laboratory
U.S. Department of Energy
626 Cochrans Mill Road
PO Box 10940
Pittsburgh, PA 15236-0940

DOE Cooperative Agreement No. DE-FC26-05NT42592

Prepared by:

Nicholas S. Kalenze
Ryan J. Klapperich
John A. Hamling
Charles D. Gorecki
Edward N. Steadman
John A. Harju

Energy & Environmental Research Center
University of North Dakota
15 North 23rd Street, Stop 9018
Grand Forks, ND 58202-9018

Nicholas A. Azzolina

The CETER Group, Inc.
1027 Faversham Way
Green Bay, WI 54313-7491

March 2015
2015-EERC-03-15 Approved

EERC DISCLAIMER

LEGAL NOTICE This research report was prepared by the Energy & Environmental Research Center (EERC), an agency of the University of North Dakota, as an account of work sponsored by the U.S. Department of Energy (DOE) National Energy Technology Laboratory (NETL). Because of the research nature of the work performed, neither the EERC nor any of its employees makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement or recommendation by the EERC.

ACKNOWLEDGMENT

This material is based upon work supported by DOE NETL under Award No. DE-FC26-05NT42592.

DOE DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

NDIC DISCLAIMER

This report was prepared by the EERC pursuant to an agreement partially funded by the Industrial Commission of North Dakota, and neither the EERC nor any of its subcontractors nor the North Dakota Industrial Commission (NDIC) nor any person acting on behalf of either:

(A) Makes any warranty or representation, express or implied, with respect to the accuracy, completeness, or usefulness of the information contained in this report or that the use of any information, apparatus, method, or process disclosed in this report may not infringe privately owned rights; or

(B) Assumes any liabilities with respect to the use of, or for damages resulting from the use of, any information, apparatus, method, or process disclosed in this report.

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the NDIC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the NDIC.

TABLE OF CONTENTS

LIST OF FIGURES iii

LIST OF TABLES iii

PREFACE 1

INTRODUCTION 2
The Process 3
Data Migration 4
Challenges 4
Lessons Learned 5

BELL CREEK DATA MANAGEMENT STANDARD OPERATING PROCEDURE 7

1.0 INTRODUCTION 8
1.1 Purpose 8
1.2 Legacy Versus New Project Data 8
1.3 Data Life Cycle 9
1.4 Organization of the Data Management SOP 10

2.0 FILE DIRECTORY STRUCTURE 12
2.1 Exceptions 13
2.1.1 Specialized Software 13
2.1.2 Photographs 13
2.1.3 Temporary Transfer Folder 13
2.2 Folder Shortcuts 13
2.3 Detailed File Path Standards 14

3.0 FILE-NAMING CONVENTIONS 15
3.1 File Types 16
3.2 File-Naming Format 16
3.2.1 Descriptor Element 16
3.2.2 Date Element 17
3.2.3 Version Element 17
3.3 Importance of File Naming and Version Control 18
3.4 File Naming for Document Development 18
3.5 Document Version Control 20

4.0 QUALITY ASSURANCE AND QUALITY CONTROL 22
4.1 Quality Management and the Data Life Cycle 22

Continued . . .

TABLE OF CONTENTS (continued)

4.2 QA Before Data Collection 23
4.3 QA/QC During Data Acquisition 23
4.4 QA/QC During Data Storage 24
4.5 QA/QC During Data Processing 27
4.5.1 Processed Data Subfolders 27
4.5.2 File-Naming Conventions 28

5.0 SHARED FOLDER ACCESS CONTROLS 29

6.0 SHARED FOLDER BACKUP POLICY 30
6.1 Current Risk of File Loss 30
6.2 Fixed-Frequency Automated Backup 30
6.3 Supplemental Manual Backup 30
6.4 Backup Details 32

7.0 POLICY ADHERENCE INITIATIVE 33
7.1 Purpose 33
7.2 Folder Managers 33
7.3 Monitoring Schedule 34
7.4 Monitoring Plan 35
7.5 Folder Inspection Checklist 35
7.6 Feedback to Project Team Members 39

8.0 CHANGE MANAGEMENT PLAN 40
8.1 Change in the File Directory Structure or Access Controls 40
8.2 Other Changes 40
8.3 Project Management Team Review 41

9.0 SOURCES 42

LIST OF FIGURES

1-1 The six stages of the data life cycle 10

2-1 Bell Creek Confidential Shared Folder top-level folder directory 12

3-1 Example D66 document development highlighting the use of version control 21

4-1 QA/QC procedures from the point of data acquisition to the point of data storage 26

6-1 Bayesian statistical approach for quantifying the daily and weekly rate of files that were saved to the current shared folders ("bellconf" and "bconf_mva") between November 6, 2013, and March 21, 2014 (135 days – top panel) and updated with new information collected between March 21 and March 28, 2014 (7 days – bottom panel) 31

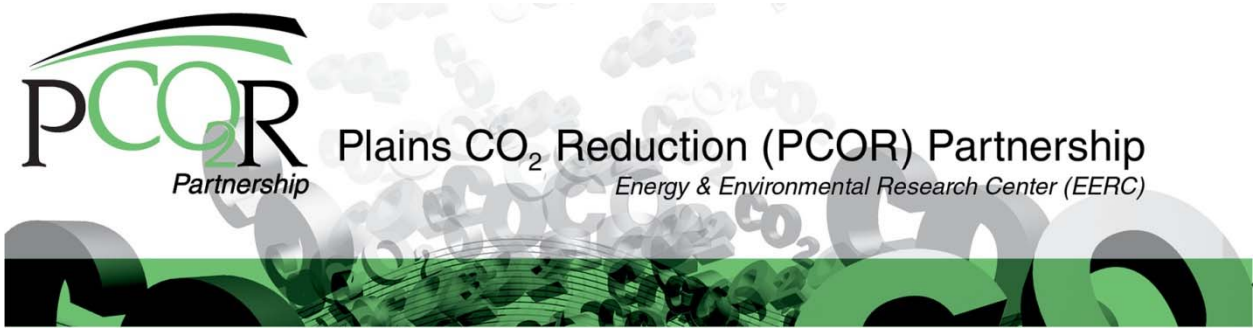
LIST OF TABLES

7-1 List of specific top-level directories and associated subdirectories for the Bell Creek Confidential Shared Folder and the Folder Managers to whom responsibility has been assigned 34

7-2 Phased schedule for monitoring compliance with the Data Management SOP 34

7-3 Example blank checklist for monitoring compliance with the Data Management SOP 37

7-4 Example completed checklist for monitoring compliance with the Data Management SOP for the 200_Soil_Gas folder 38



DATA MANAGEMENT POLICY AND PROCEDURES DEVELOPED FOR THE PCOR PARTNERSHIP'S BELL CREEK STUDY

PREFACE

This document presents a data management policy and procedures developed for the Plains CO_2 Reduction (PCOR) Partnership's Bell Creek Study. The policies and procedures represented within are designed to be a living document that allows for flexibility, adaptability, and accommodation of expanding and diversifying data sets and the development and implementation of improved data handling as technologies progress and/or for additional requisites as the project evolves. The objective of the policy and pursuant procedures are to 1) enhance and ensure data accessibility and collaboration among Project Team members, 2) ensure consistency of data sets and versioning control, 3) maximize data security and to mitigate the potential for data loss, 4) protect against personnel turnover, 5) minimize redundancy of data sets across multiple users, 6) provide consolidation of data to minimize computer resources and aid in planning of future computer resource needs, and 7) ensure efficient end of project closure and data transfer to stakeholders.

The project has generated, and continues to generate, large and diverse data sets ranging from laboratory analysis of geologic and fluid samples, reservoir surveillance data, geophysics data, geochemistry data, well log data, oilfield operations data, completions data, production and injection data, well log data, lidar data, soil gas and water chemistry data, technical reports and presentations, public outreach materials, and project management materials to name just a few. Additionally, each data type often contains both raw data and various iterations of processed data necessary for quality assurance/quality control (QA/QC), and chain of custody. Each data type can also contain a combination of public, licensed, confidential, finalized and prefinalized data. To illustrate the magnitude of the issue, from project inception in 2010 through 2014, over 100 TB and over 1,000,000 unique files have been generated through the course of the project.

The size and diversity of the Project Team utilizing, interacting with, and communicating the data also create a variety of separate but significant challenges. In 2014, the Project Team consists of over 60+ individual technical and administrative staff which have spent time working with the data generated throughout the course of this project. Data access, simultaneous interfacing, versioning control, and data redundancy are significant challenges to ensuring efficient use of computer resources, minimizing nonproductive time, and ensuring correct up-to-date data are accessible and being utilized and employed by all Project Team members.

Project data are valued assets that need to be managed through their entire life cycle. As the nature of large-scale research projects evolve and technology advances, diverse teams and individuals with specialized skills are more frequently needed. While this document was written specifically for the Bell Creek Study, the challenges associated with it and the solutions implemented are not unique. Technical managers for large-scale research projects who have had involvement with this project have indicated that they have derived significant value from this document as a guide for adapting and implementing data management policies and procedures for a wide range of research projects. By implementing a clear consistent data management policy at project initiation, many of the challenges discussed within can be circumvented. The authors are, therefore, presenting the main text in an unabridged format from the current version at the time of publication with the hopes that it may aid data management practices for other research projects.

The PCOR Partnership is led by the Energy & Environmental Research Center (EERC) at the University of North Dakota and is one of the seven regional partnerships under the U.S. Department of Energy's National Energy Technology Laboratory Regional Carbon Sequestration Partnership (RCSP) Initiative. The goal of the RCSP Program is to determine the best geologic and terrestrial storage approaches and to apply and demonstrate viable technologies to safely and permanently store carbon dioxide (CO_2).

The Bell Creek Study is a development-phase RCSP project being conducted by the PCOR Partnership to investigate commercial-scale injection of CO_2 and to demonstrate safe viable storage of millions of tons of CO_2 associated with a commercial enhanced oil recovery (EOR) project being conducted by Denbury Onshore LLC in the Bell Creek oil field. The project is also tasked with understanding the necessary regulatory, economic, liability, ownership, and public outreach efforts needed for successful carbon capture, utilization, and storage (CCUS) and to develop the necessary human capital, knowledge base, and experience necessary to implement future CCUS operations. The Bell Creek Study is a multiyear, multicomponent project (2010–2017) focused on many interrelated aspects of the CCUS life cycle and includes a rigorous approach to site characterization, geologic modeling and simulation, risk assessment, and monitoring, verification, accounting, and reservoir surveillance to demonstrate safe, viable storage of CO_2 associated with EOR.

INTRODUCTION

With the Bell Creek Data Management Policy established, work could begin on the Data Management Standard Operating Procedure (SOP). The SOP was developed to help organize and facilitate efficient data retrieval as well as improve file naming, document versioning, and standardized QA/QC procedures for Project Team members. Since a wide variety of data types are used for the project, improved data accessibility and usability (defined here as compatible data sets) are vital to the project group. Improved data security was also a goal in establishing the Data Management SOP to limit "write access" to only those members who work intimately with particular data sets, while "read-only access" is provided to everyone.

To accomplish this, a dedicated server and confidential shared folder system (with access limited to only Project Team members) was created. The folder structure contained on this server is organized in a hierarchical format to allow for easier data searching, identification, and retrieval among the various data sources (i.e., Modeling, Simulation, Core Analysis, Monitoring data sets, etc.). Significant effort has gone into consolidating data sources in this format and has already saved the Project Team significant time in data retrieval, allowing the Project Team to better communicate and collaborate. The establishment of this dedicated, project-specific storage system will also allow for sharing data with external stakeholders in a streamlined fashion.

The Data Management SOP is an effort from the PCOR Partnership to establish a standardized and centralized database for shared data. This effort sets a precedent for future EERC project adoption in terms of organizing and streamlining data sources.

The Process

The steps included the following:

- Identifying, categorizing, and parsing key data classes based on data type and data usage. This was carried out with extensive input from the Project Team members who were familiar with and regularly used each data type. This process was necessary to create the previously mentioned shared folder system.

- Assigning read/write access controls to specific project folders for specific Project Team members.

- Incorporating a backup policy for the Confidential Shared Folder to limit the risk of data and file loss between backups.

- Identifying, assigning, and training "data champions" responsible for organizing and moving specific data folders or files from the existing shared folders to the new file directory.

- Assigning "folder managers" responsible for monitoring top-level folders/subfolders in the Confidential Shared Folder on a recurring basis.

- Developing a Policy Adherence Initiative to ensure compliance with the Data Management SOP.

- Consolidating of existing data to remove as much duplicate and otherwise unneeded data as possible prior to the date of transferring data to the new server.

- Prepared a staff-training initiative to help Project Team members adhere to the Data Management Policy.

Throughout each stage of the rollout process, feedback and input were sought from all Project Team members allowing for refinement to the Data Management SOP.

Data Migration

An essential task associated with the rollout was data migration to a single, centralized, shared folder, the new Confidential Shared Folder, on a dedicated server. To effectively accomplish the data migration, which was scheduled during a weekend where server activity was minimal, three important subtasks were assigned:

- *Folder assignments*: Each subfolder that was located on the old shared folders was assigned to specific individuals at the EERC. These assignments were based on the project-specific knowledge of the individual. These individuals were then assigned responsibility for dictating where existing files should be stored in the new file directory structure in accordance with the Data Management SOP.

- *Itemized port plan:* A detailed "port plan" was developed which outlined 1) the specific folders located on the previously shared folders, 2) the individual who is responsible for that folder, 3) the specific folder(s)/subfolder(s) destination for porting to the Confidential Shared Folder, and 4) the schedule for porting these files.

- *Data ported to new server:* This is the implementation of the itemized port plan. The port occurred during the weekend of May 16–18, 2014. All files were successfully ported (100% effective transfer). An up-front file inventory step served as the official file count prior to the port. A subsequent file-checking process occurred after the port to validate that all files were successfully transferred to the new Confidential Shared Folder, which went live on Monday, May 19, 2014.

Challenges

Challenges associated with rollout activities included the following:

- Consolidating project data onto a single shared server. The previous shared network drives consisted of a variety of file types and formats, for example, site characterization data, models, monitoring data acquisitions, presentations, literature and reports, presentations, and photos. Many of these were found to be stored on multiple drives and individual user systems as not all Project Teams were utilizing or had access to the same network drives.

- Assignment of data sources to key data champions. The data champions were in charge of preparing their respective data sets for migration to the Confidential Shared Folder on the new server during the port weekend. The sheer volume and diverse nature of the various data sets necessitated distributing this task among multiple Project Team members.

- Identification of duplicate files and data sets. Duplicate files were identified using the program "Duplicate File Finder Pro Edition" from *Ashisoft, Inc*. At the start of the data evaluation process, approximately 50% of the entire file count was initially identified as being potentially duplicative data. Many "duplicates" were found to be files which

are autogenerated by the various software programs that the PCOR Partnership uses. These autogenerated files had identical file names but unique content. Nevertheless, a substantial number of data types and even some data sets were found to have been duplicated and stored in multiple locations. This led to a manual investigation of the data sets whereby duplicate files and folders were identified by various Project Team members (under the supervision of their respective data champion).

- Removal of duplicate files and data sets. Duplicate files and outdated data sets were manually purged by Project Team members under the supervision of their respective data champion. In the end, approximately 17% of the original files were purged prior to the data-porting process.

- Full communication efforts via meetings between the Project Team members. These meetings were held to discuss the issues with select shared data sets including duplicates, outdated data sets, and identifying the latest versions prior to data migration. In addition, these meetings helped all Project Team members contribute valuable input regarding unique data requirements to the development of the data migration and consolidation plans.

- Developing an SOP that would suit the Project Team member/management staff needs while still being flexible enough to adapt to changes in those needs.

- Development of consistent parent and subfolder structure to accommodate the needs of all shared data sources.

- Utilizing a universal file-naming system that allows the Project Team to understand a file's descriptor, date, and version elements and be intuitively searchable. This allows for more rapid file identification and more efficient communication between Project Team members.

Lessons Learned

The Data Management SOP document was created to be adaptive to the needs of the project. As data sources evolve over the course of the project, so will the Data Management SOP. It is for this reason that the SOP document was created to be a "living document" where updates/changes can be made to reflect the present data management practices. The Project Team is, therefore, able to view the most up-to-date version of the SOP through a change procedure, which is posted to the main directory of the Confidential Shared Folder. The feedback process from the Project Team facilitates any necessary changes to the SOP, although there have been relatively minor changes since the Confidential Shared Folder went live (May 19, 2014). Active management is needed to ensure that the SOP is fully adopted within an organization.

Data management for the project has been improved substantially since implementing the SOP. In July 2014, a questionnaire was submitted to team members for feedback. Responses from the group were positive in ensuring that the Project Team is experiencing an improved ability to locate files efficiently as well as the latest versions to those files. Feedback meetings

are scheduled periodically with the Project Team to ensure the Data Management SOP is being followed and that the current SOP is applicable to future data acquisitions. If a concern is brought up at any time during this review process, then that concern can be addressed and an immediate change to the Data Management SOP may be enacted through a change procedure.

A primary goal of the Data Management SOP is to make the system as functional as it can be for both the Project Team and the EERC. To this point, the EERC feels this effort has been successful. Utilizing the processes established in the Data Management SOP has enabled the Project Team to better communicate among itself, as well as efficiently locate and identify shared data resources. The Project Management Team has instituted an "open-door policy" to address any issues of concern from the Project Team. Buy-in from the Project Team has been key for allowing successful implementation of the Data Management SOP.

Project data are valued assets that need to be managed over their entire life cycle beyond the immediate need. Effective organization and the presence of secure long-term storage of all project data sets are essential not only for the current efficiency of the project, but also for the transfer of data to project stakeholders. Data storage QA/QC, saving, file naming, and versioning practices conducted from the Data Management SOP have allowed the Project Team to effectively track, organize, and maintain the highest possible quality of data. The current Data Management SOP that details this strategy is published in full as an example to other projects and operators. The appendices referenced in the Data Management SOP have not been included with this document. Rather, they have been withheld by the EERC for internal use because of the presence of potentially confidential and business-sensitive data.

For more information, please contact the authors of this document, or visit the PCOR Partnership Web site at www.undeerc.org/pcor/.

Bell Creek Data Management Standard Operating Procedure

Bell Creek Data Management Policy

All key data and materials generated for the Bell Creek Study, as well as all interpretations generated from that data (including raw data necessary to recreate those products), will be stored and maintained in an up-to-date secured singular location in an intuitive, organized manner that promotes efficient ease of use and retrieval, maximizes data accessibility between the Project Team, promotes versioning control, minimizes redundancy of data sets across multiple users, and prevents unauthorized access or dissemination of project data.

Data are considered a valued and protected resource to the EERC, the PCOR Partnership, and the Bell Creek Study. Adherence to the Data Management SOP is mandatory for all personnel who utilize Bell Creek data. It is the personal responsibility and prerogative of all personnel who utilize Bell Creek data to ensure compliance with the policy as outlined in the Data Management Standard Operating Procedure; violation may result in disciplinary action, up to and including possible termination.

1.0 Introduction

This *Bell Creek Data Management Standard Operating Procedure* (Data Management SOP) presents the data management SOPs for the Bell Creek Carbon Dioxide (CO_2) Carbon Capture and Storage (CCS) and Enhanced Oil Recovery (EOR) Demonstration Project (hereafter referred to as the "Bell Creek Study"). *This document pertains to **all** Bell Creek data.*

1.1 Purpose

Project data are valued assets that need to be managed over their entire life cycle beyond the immediate need. Ultimately, most if not all of the Bell Creek data will be delivered to the U.S. Department of Energy (DOE) National Energy Technology Laboratory (NETL) and/or Denbury Resources Inc. (Denbury). Therefore, the organization and secure long-term storage of all project data are essential not only for the current efficiency of the project, but for the eventual transfer of data to the client.

A new confidential project shared folder has been created on the "zfs" server at the Energy & Environmental Research Center (EERC) (hereafter referred to as the "Bell Creek Confidential Shared Folder"). This Data Management SOP discusses the organization of the Bell Creek Confidential Shared Folder and the SOPs for naming, saving, modifying, accessing, or backing up project data in this folder.

1.2 Legacy Versus New Project Data

All legacy project data were migrated to the Bell Creek Confidential Shared Folder, and all new project data must be saved to the Bell Creek Confidential Shared Folder using the conventions and processes outlined in this Data Management SOP.

Legacy data included all Bell Creek data saved on the former project shared folders up to and including May 16, 2014. These data were migrated to the Bell Creek Confidential Shared Folder in a systematic transfer according to the "Port Plan" that was developed for May 16–18, 2014. Most of the legacy data were not renamed; they were simply relocated from the old locations to the new Bell Creek Confidential Shared Folder and saved according to the new file directory structure. Under special circumstances, a particular task lead may request that legacy data be renamed according to the new file-naming conventions to facilitate accessibility for certain common data sets. Additionally, when legacy data are accessed through routine project activities, users should evaluate file names for ambiguity. If a file is determined to be ambiguous or nonintuitive or if multiple versions of legacy data sets exist, then the following steps should be taken:

- Perform a quality check on the data (Are the data still relevant? Are the data accurate? Are data types within clearly labeled and presented, etc.?).

- Combine/update/correct the files in question.

- Create a new version of the file using the new file-naming convention.

- Delete or create a recycle folder to store the out of date version(s) of the file that have been superseded until confident of deletion (except raw data).

The Port Plan is separate and distinct from this Data Management SOP.

New project data include any project data acquired after May 16, 2014. These data must be saved to the Bell Creek Confidential Shared Folder using the conventions and processes outlined in this Data Management SOP.

1.3 Data Life Cycle

Project data are valued assets that need to be managed over their entire life cycle beyond the immediate need. The data life cycle entails six primary stages (Figure 1-1):

1. **Plan:** A documented sequence of intended actions to identify and secure resources and gather, maintain, secure, and utilize data holdings.

2. **Acquire:** Acquisition involves collection or addition to existing data holdings. There are generally four methods of acquiring data: (a) collecting new data; (b) converting or transforming legacy data; (c) sharing or exchanging data (e.g., data from the Client or third party); and (d) purchasing data.

3. **Process:** Processing denotes actions or steps performed on data to verify, organize, transform, integrate, and extraction of data for subsequent use.

4. **Analyze:** Analysis involves actions and methods performed on data that help describe facts, detect patterns, develop explanations, and test hypotheses.

5. **Preserve:** Preservation involves actions and procedures to keep data for some period of time and/or to set data aside for future use and includes data archiving and/or data submission to a data repository.

6. **Publish/Share:** To prepare and issue, or disseminate, the final data products of the research or program activity.

Across the data life cycle stages, the data must be described, backed up and secured, and managed to ensure quality.

- **Document:** Throughout this life cycle process, documentation must be updated to reflect actions taken upon the data. This includes acquisition, processing, and analysis, which may touch upon any stage of the life cycle. Documentation also includes ancillary materials (e.g., field notes).

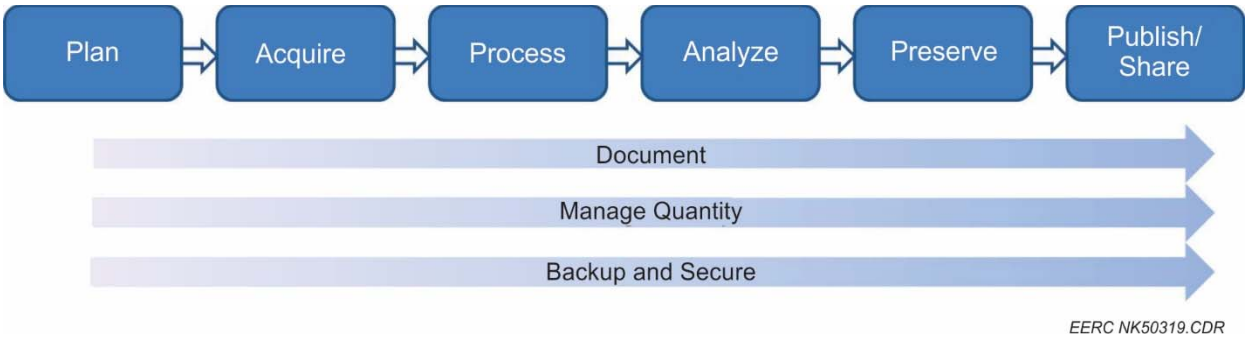


Figure 1-1. The six stages of the data life cycle (adapted from USGS [2014]).

- **Manage Quality:** Protocols and methods must be employed to ensure that data are properly collected, handled, processed, used, and maintained at all stages of the scientific data life cycle. This is commonly referred to as "QA/QC" (Quality Assurance/Quality Control). QA focuses on building in procedures to prevent defects, while QC focuses on testing for quality (e.g., detecting defects). QA makes sure you are doing the right things, the right way. QC makes sure the results of what you have done are what you expected.

- **Back Up and Secure:** Steps must be taken to protect data from accidental data loss, corruption, and unauthorized access. This includes routinely making additional copies of data files or databases that can be used to restore the original data or for recovery of earlier instances of the data.

This Data Management SOP is intended to facilitate good data stewardship, preserve data quality, and secure project data for the full data life cycle of the Bell Creek Study.

1.4 Organization of the Data Management SOP

The remained of this document is organized as follows:

- **Section 2:** Summarizes the Bell Creek Study file directory structure for the recently created Bell Creek Confidential Shared Folder on the "zfs" server.

- **Section 3:** Describes the file-naming conventions to be used by all Project Team members for new files saved to the project shared folder.

- **Section 4:** Outlines the QA/QC procedures that have been enacted to ensure long-term quality of the project data assets.

- **Section 5:** Summarizes the Bell Creek Confidential Shared Folder access controls.

- **Section 6:** Summarizes the Bell Creek Confidential Shared Folder backup policy.

- **Section 7:** Describes the Policy Adherence Initiative.

- **Section 8:** Discusses the Change Management Plan for adapting the current Data Management SOP to project changes.

- **Section 9:** Provides a list of references that were used in developing the Data Management SOP.

NOTE: These Appendices have not been included in this document because of the presence of business-sensitive data. However, they are referenced throughout the report:

- **Appendix A:** Provides top-level directories and their related subfolders for the Bell Creek Confidential Shared Folder on the "zfs" server.

- **Appendix B:** Provides detailed instructions for specific types of project data and the correct file path for saving raw (as-received) data to the Bell Creek Confidential Shared Folder.

- **Appendix C:** Provides specific processed data file directories for each of the 20 different top-level folders in the Bell Creek Confidential Shared Folder.

2.0 File Directory Structure

Good folder management leads to organized file management. The file directory structure of the Bell Creek Confidential Shared Folder is the initial starting point for saving project data and files. The top-level directories and associated subdirectories (subfolders) are the first tier of organization. Each subfolder is then further subdivided to better organize the information. The top-level directory of the Bell Creek Confidential Shared Folder is shown below in Figure 2-1.

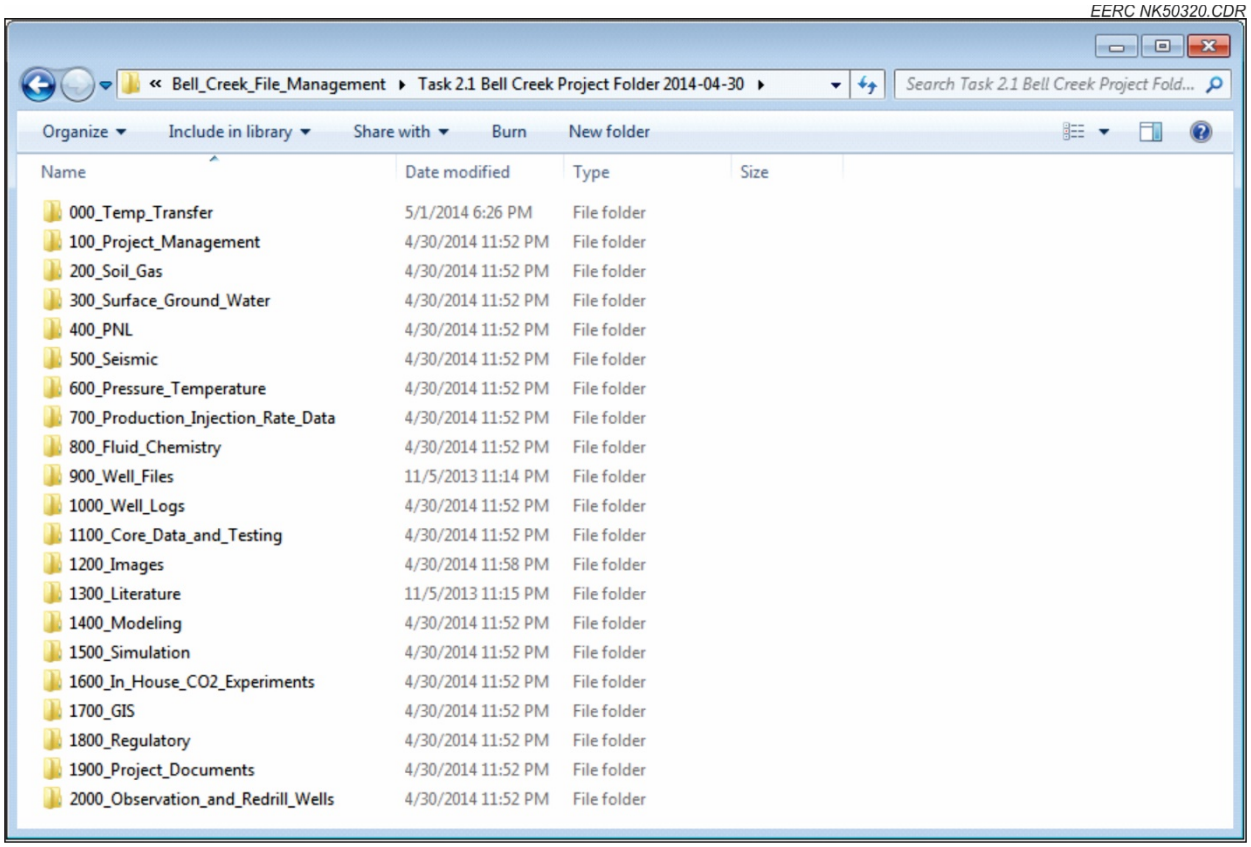


Figure 2-1. Bell Creek Confidential Shared Folder top-level folder directory.

The file directory structure outlined in this Data Management SOP is intended for use by all Project Team members and across all file types. Top-level directories and their related subfolders are shown in **Appendix A**.

2.1 Exceptions

2.1.1 Specialized Software

Specialized software used for geologic modeling (e.g., PETRA, Petrel, or TechLog), simulation modeling (e.g., Computer Modelling Group [CMG] Ltd.), and other processes such as seismic interpretation and geographic information system (GIS) generally require a specialized file directory structure that cannot be changed. In the event of a conflict between this Data Management SOP and a specialized software program, the file directory structure mandated by that specialized software program would supersede this Data Management SOP.

2.1.2 Photographs

A photograph may often comprise subject matter from multiple categories. For example, a photo of geologic core from the 05-06 OW well during drilling activities could be categorized as either "geologic core," "Drilling & Completion," or "05-06 OW." In order to simplify a potentially complex file structure and to facilitate a functional level of searchability, key words will be tagged in the metadata of each photograph rather than depending upon a rigid file directory. Photographs will be saved according to the date and major activity during which the photographs were acquired in order to provide a basic descriptor for file organization. A second exemption pertains specifically to laboratory core photos which are considered "core data" rather than "photographs" and are archived accordingly.

2.1.3 Temporary Transfer Folder

A temporary transfer folder is necessary to facilitate collaboration between various members of the Project Team and to provide a secure method of sharing confidential project data between authorized members. The temporary transfer folder will function similarly to the EERC x:\drive and be organized by the name of each individual Project Team member. All Project Team members will have read/write access to the Temporary Transfer Folder.

No original or key project data should ever be stored in the temporary transfer folder!

Only temporary copies of data that facilitate integration into ongoing work should be saved in the Temporary Transfer folder. For example, it may be helpful when writing a report to have copies of various data types readily accessible. Individuals could mine relevant data from the shared folder and place ***copies*** of relevant material within their Temporary Transfer folder for efficient reference. Once the work is completed, the copies should then be deleted. *The Temporary Transfer Folder may be purged at any time without notice to free network space. Data older than 1 month will be automatically purged.*

2.2 Folder Shortcuts

Duplicate files should not be created in the Bell Creek Confidential Shared Folder. Eliminating redundancy is essential for successful data management; if more than one file needs to be updated every time new information becomes available, then the Data Management SOP

will not achieve its stated purpose of efficiently managing project data quality over the entire life cycle.

Instead of copying a source file from its original folder into another folder, shortcuts will be used to point users from their working folder to the folder containing the source file. A shortcut is a link from one folder to another, where the user may simply click on this link and be directed to the designated folder.

The specific shortcuts within each subfolder are shown in **Appendix A**. Additional shortcuts can be created by the folder manager as needed.

2.3 Detailed File Path Standards

Detailed file path standards for each Bell Creek data type are described in Section 4.0 – Quality Assurance and Quality Control.

3.0 File-Naming Conventions

File-naming conventions will be used in conjunction with a consistent file directory structure (see Section 2.0 – File Directory Structure) to help EERC staff efficiently:

- Locate and retrieve specific files using readily identifiable file names.
- Save new files according to a simple set of rules.
- Differentiate among revisions for draft and interim work products (version control).

The file-naming conventions developed for the Bell Creek Study were derived from seven fundamental rules, which include:

- **Avoid long file names:** File names should not be lengthy and difficult to decipher. The maximum length for a file path is commonly defined as 260 characters. Therefore, extraordinarily long file names combined with extensive file paths are not good practice.

- **Avoid spaces in file names:** Spaces should not be used in file names. Some search tools do not work with spaces, and several software programs cannot utilize data files with spaces in the file name.

- **Use file names that are indicative of what the file contains:** It is generally not advised to use random numbers or other identifiers for file names. Rarely will all Project Team members recall the naming convention, which means that someone will have to open the file to figure out what it is. This is inefficient. Instead, file names should be readily identifiable and provide some information about what the file contains.

- **Use file names that are scannable with the human eye:** EERC staff will most likely try to find a particular file using the project folders and subfolders in Windows Explorer. Therefore, the file names should sort in some sensible order, and people should be able to easily recognize the file that they want.

- **Use the underscore (_) as an element delimiter:** Consistent with creating file names that are scannable with the human eye, the underscore provides a clear, visual means for field delimiting and avoids the use of spaces in file names.

- **Use file names that can be naturally ordered alphabetically and numerically:** Windows Explorer can sort by file type, name, and date. It is very useful to have the file names sort alphabetically and numerically so that the most current version is either first or last in a series of files with consistent names.

- **Use consistent file-naming conventions across technical staff:** Everyone on the Bell Creek Project Team will need to use the same file-naming convention.

3.1 File Types

There are many different file types that are saved on the Bell Creek Confidential Shared Folder. Common file types used by many of the EERC staff include the following:

- Microsoft Excel (.xls or .xlsx)
- Microsoft PowerPoint (.ppt or .pptx)
- Microsoft Word (.doc or .docx)
- Portable document format (PDF) (.pdf)
- Text file (.txt)
- Comma-separate-value (.csv)
- Image files (.jpg, .gif, .tif, .png, etc.)

Specialized software used for geologic modeling (e.g., PETRA, Petrel, or TechLog), simulation modeling (e.g., CMG), and other processes such as seismic interpretation and GIS generate specialized file types. The file-naming conventions outlined below are intended for use across all file types. However, some of the specialized software noted above may require additional file-naming conventions beyond the scope of this document.

3.2 File-Naming Format

Each file name should contain up to three elements, which include:

- **Descriptor:** The "Descriptor" element (see 3.2.1) is an abbreviated descriptor of the file that is easy to recognize.

- **Date:** The "Date" element (see 3.2.2) reflects a date that is relevant to the file.

- **Version:** The "Version" element (see 3.2.3) reflects the version of the file, which is particularly useful during document development (i.e., from draft → final document) and for data files that have undergone one or more changes between the original and the final version of the data file.

The file name components should be separated using the underscore (_). A hyphen (-) should be used to delimit words within an element, for example, a Descriptor element with multiple words.

Detailed instructions for using each element are provided below.

3.2.1 Descriptor Element

The Descriptor element should include keywords regarding the content of the file—within a reasonable length. The keywords should be descriptive so that other Project Team members can clearly understand the meaning. The length of the Descriptor element should be as short as possible but sufficient to avoid ambiguous search results. Abbreviations may be used to help

shorten the Descriptor element; however, the need to be easily recognizable supersedes the need to abbreviate.

The underscore (_) is reserved for delimiting elements (i.e., Descriptor_Date_Version). For Descriptor elements with multiple words or numbers, which will commonly arise, these words or numbers should be separated with a hyphen (-). For example, the Descriptor element "Well 0506 bottomhole pressure (BHP) second quarter" data would be written as "Well-0506-BHP-Q2."

For multiple files with similar file names (e.g., monitoring data collected routinely over time), the components of the Descriptor element should be ordered from general to specific detail of importance as much as possible. For example, for well BHP data that are collected quarterly, the Descriptor element would identify the well name, the measurement, and the quarter, followed by the remaining Date and Version elements:

- Well-0506-BHP-Q2_2014-05-01_V001.xls
- Well-0506-BHP-Q3_2014-08-15_V001.xls
- Well-0506-BHP-Q4_2014-10-01_V001.xls

Lastly, the Descriptor element should use only text or numbers, and special characters such as ~ ! @ # $ % ^ & * ( ) ` ; < > ? , [ ] { } ' " | must be avoided.

3.2.2 Date Element

The Date element will follow the International Organization for Standardization (ISO) 8601 standard for representation of dates and times and use the following code: YYYY-MM-DD, which is a four-digit year, two-digit month, and two-digit day separated by a hyphen (-) (e.g., 2014-01-01 for January 1, 2014).

The Date element may differ from the file date, which is simply the date that the file was created and saved to the project shared folder or local drive. Instead, the Date element should reflect a significant date relevant to the file, for example:

- Sample collection date
- Laboratory analysis date
- Modeling or simulation run date
- Date file was received from the Client or third-party subcontractor
- Other significant date relevant to the file generation or receipt by the EERC

This Date element standard dictates that dates be ordered from largest term to smallest term, which allows for file systems to sort the files appropriately.

3.2.3 Version Element

The Version element should start with "V" followed by 3 digits and should be placed as the last element in the file name. A minimum of 3 digits with leading zeros is required to ensure that search results are properly sorted. In other words, "Version 1" would be denoted as "V001"

not “V1”. If you use “V1,” then “Version 11” (“V11”) would sort before “Version 1” (“V1”), which is not correct. The 3-digit system will accommodate up to 999 revisions of the same file, which should be sufficient for the Bell Creek Project.

3.3 Importance of File Naming and Version Control

The file-naming conventions are the primary version control for the Bell Creek Confidential Shared Folder. For two or more files with the same Descriptor element, the file with the most recent Date and/or Version element will ***ALWAYS*** be considered by EERC staff to be the most current or correct version of the data. Therefore, correctly adhering to these file-naming conventions is imperative to ensure that the most recent Date and/or Version file is always the “correct” version.

3.4 File Naming for Document Development

Project document development (e.g., DOE deliverables, internal reports, value-added reports, conference/peer-reviewed publications, etc.) is a special case of file naming, as multiple draft versions are commonly generated between the first draft and the final work product. In general, the file-naming conventions are identical to those discussed above. However, several key exceptions exist, which are summarized below.

Each document will be saved in a subfolder within the top-level directory “1900_Project_Documents.” The subfolder will describe the name of the document, e.g., “D-66_Simulation_Rpt_Update_3.” Each document subfolder will contain the following additional subfolders to better-organize the individual pieces of the document:

- Draft_Elements
  - Text
  - Tables
  - Figures
  - Appendices
- Draft_to_Denbury
- Draft_to_DOE
- FINAL_APPROVED

The “Draft_Elements” subfolder is the first tier of separating draft document elements from final documents. All draft work products are to be saved into this folder. The subfolder “Text” is specific to the Microsoft Word® version of the document text. The subfolder “Tables” may be either Microsoft Excel® or Word® versions of tables that will be either inserted into the document or added to the end of the document. The subfolder “Figures” includes EERC figures with figure numbers (e.g., PDF files), Microsoft Word® figures, Microsoft Excel® figures, or any other source file types used to develop document figures. Lastly, the subfolder “Appendices” is to be used for document appendices, which may contain text, tables and figures.

The "Draft_to_Denbury" and "Draft_to_DOE" subfolders contain only the MS Word (or PDF) file that has gone through EERC workflow and was distributed to Denbury or DOE, respectively. Comments received from either Denbury or DOE would also be saved into these folders.

The "FINAL" subfolder is reserved for the final work product that has been approved by Denbury and DOE.

The document types may include the following:

- Abstract
- AGL report
- Conference paper
- Deliverable
- Fact sheet
- Internal report
- Milestone
- Presentation
- Publication
- Value-added

Each document type will be saved into the appropriate subfolder under 1900_Project_Documents.

File naming for tables, figures, and appendices will incorporate the prefix "Table," "Figure," or "Appendix" in the Descriptor element:

- Table: "Table-"
- Figure: "Figure-"
- Appendix: "Appendix-"

Example file names for the text, tables, figures, and appendices associated with a hypothetical Deliverable 66 Bell Creek Project Simulation Report (2014 Update 3) are listed below:

- D-66-Sim-Rpt-Update3_2014-08-31_V001.doc
- Table-Reservoir-Props_2014-08-31_V001.xls
- Table-Geochemical-Props_2014-08-31_V001.xls
- Table-Groundwater-Wells_V001.xls
- Figure-Integrated-Approach_2014-08-31_V001.pdf
- Figure-Site-Map_2014-08-31_V001.pdf
- Appendix-A-PVT-Model_D-66_2014-08-31_V001.doc
- Appendix-B-Res-Simulation_D-66_2014-08-31_V001.doc

3.5 Document Version Control

Document revisions can be one of the most complicated aspects of file management. It is common during the document development process (e.g., 30% first draft → 60% second draft → 90% → 95% pre-final draft → 100% final document) to lose track of revisions, especially when multiple authors are involved in the process. Folder management and file-naming conventions can help organize revisions to avoid such problems.

The file-naming convention described above can be used to order revisions of draft reports using combinations of Date and Version element changes. For example, if you are developing the Deliverable 66 Bell Creek Project Simulation Report (2014 Update 3) report that you started on May 1, 2014, the first draft would be named "D-66-Sim-Rpt-Update3_2014-05-01_V001.doc." Subsequent revisions would have a new Date element; for example, a revision to the first draft that is saved to the subfolder on May 15, 2014, would be renamed "D-66-Sim-Rpt-Update3_2014-05-15_V001.doc." File names adhering to this practice will sort chronologically, and the most recent draft will be at the bottom of the list, for example:

- D-66-Sim-Rpt-Update3_2014-05-01_V001.doc
- D-66-Sim-Rpt-Update3_2014-05-15_V001.doc
- D-66-Sim-Rpt-Update3_2014-06-01_V001.doc
- D-66-Sim-Rpt-Update3_2014-07-01_V001.doc

Oftentimes in developing a document, revisions are made within the same day. This scenario often occurs during the latter stages of document development near the due date. In this scenario, modifying the Date element is insufficient for differentiating versions. Instead, the file name should modify the Version element. File names adhering to this practice will continue to sort chronologically, and the most recent draft will be at the bottom of the list. *Only the lead author of the document, however, is warranted to make version changes, delete old versions, and maintain the current version.* For example, if three different versions were created on July 1, 2014, by the lead author, then the file names would become:

- D-66-Sim-Rpt-Update3_2014-07-01_V001.doc
- D-66-Sim-Rpt-Update3_2014-07-01_V002.doc
- D-66-Sim-Rpt-Update3_2014-07-01_V003.doc

Lastly, some reports have multiple individuals making revisions in parallel on the same date, as opposed to sequential edits on the same date as described above. This scenario often occurs during the latter stages of document development near the due date. Each author would be generating an identical file name if his/her edits were made on the same date in parallel, which would violate the "unique file name rule" (i.e., two files with the same name cannot be saved to the same subfolder). Good communication across the Project Team is essential for these late-stage revisions. It is the responsibility of the lead author of the document to coordinate revisions. If two or more individuals are making revisions in parallel and saving to the subfolder on the same date, then a fourth element would have to be added to the file name using the individual's three-letter initials. For example, if Charles Gorecki (CDG) and John Hamling (JAH) were both

editing the July 1, 2014, Version 3 D66 report and saving to the subfolder on the same day, then the file names would read:

- D-66-Sim-Rpt-Update3_2014-07-01_V003_CDG
- D-66-Sim-Rpt-Update3_2014-07-01_V003_JAH

The lead author of the document would then be responsible for merging edits from "CDG" and "JAH" into the next version number in the series (in this example, Version 4):

- D-66-Sim-Rpt-Update3_2014-07-01_V004

Adhering to the file-naming conventions listed above for document development will ensure that all draft work products are saved and that the most recent work product is always at the bottom of the file list. Continuing the D66 example from above, the files in the subfolder would sort as shown below in Figure 3-1. An individual working at the EERC would then be able to easily recognize the bottom file as the most current version.

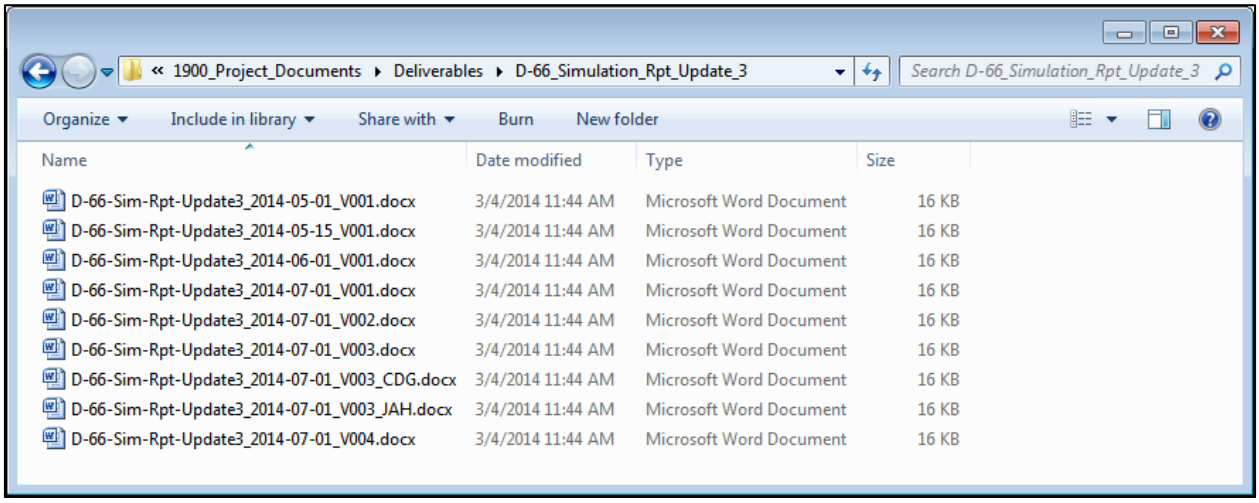


Figure 3-1. Example D66 document development highlighting the use of version control.

4.0 Quality Assurance and Quality Control

Quality Assurance and Quality Control (QA/QC) are closely related to file management and file-naming conventions but go beyond these practices to include additional steps to ensure the long-term quality of project data. QA refers specifically to the processes by which work products are developed. The goal of QA is to improve development processes so that errors do not arise while the work product is being developed. In contrast, QC focuses on identifying errors in the actual work products that are produced. QC is, therefore, a corrective tool. In this document, the term "QA/QC" will be used to describe the processes outlined in the Data Management SOP which both help to prevent errors from arising and help identify errors prior to work products being designated "final" by the EERC.

4.1 Quality Management and the Data Life Cycle

Best practices for managing data quality include several key points throughout the data life cycle stages where process and controls are implemented to ensure quality:

- **QA Before Data Collection:** This occurs during the Plan stage and includes the up-front project management, file directory, and field (or laboratory) procedures that are put in place prior to the Acquire stage.

- **QA/QC During Data Acquisition:** This occurs during the Acquire stage and includes the sampling methods, sample handling and custody, and analytical methods used to acquire new data. For data, acquisitions such as converting or transforming legacy data, sharing or exchanging data, or purchasing data, this QA/QC includes the specific software processes used and the agreements between the EERC and the Client or third parties.

- **QA/QC During Data Storage:** Data storage refers to saving data to the project shared folder. This QA/QC step occurs between the Acquire and Process stages and is intended to ensure that data are saved to the project shared folder according to the file directory structure and file-naming conventions outlined in the Data Management SOP.

- **QA/QC During Data Processing:** During the Process stage, data are prepared for the Analyze stage. Processing data may entail reorganizing, transforming, or extracting data in an appropriate output form for subsequent use. During this Process stage, QA/QC measures will take place to ensure that the future analyses of the data are using quality data for input.

The QA/QC measures described in this Data Management SOP are specific to QA/QC during data storage and data process stages. It is the responsibility of the individual Bell Creek Project task leaders to Plan to collect data (QA before data collection) and to Acquire data (QA/QC during data acquisition). This Data Management SOP does not address the specifics of these stages and only presents a high-level overview of QA/QC checklists to consider. Furthermore, many of these activities are already likely ones which you regularly perform, perhaps without even realizing it. The steps described herein are meant to formalize those practices.

4.2 QA Before Data Collection

As noted above, QA before data collection is the responsibility of the individual Bell Creek Project task leader and is beyond the scope of this Data Management SOP. However, at a minimum the following QA elements should be considered prior to acquiring data:

- A plan-of-action for data collection as well as key individual Project Team members that will be involved in the data acquisition.

- For specific analyses, the field and laboratory procedures that are used to acquire data could be included.

- If applicable, define the destination file directory for each type of data that will be acquired.

- Decide the format of how the data will be collected.
  - Will the data be collected by hand on paper?
  - Will the data be collected electronically via an instrument?
  - If the data are digital, then what format will be used?
  - Will samples be collected and submitted to a laboratory?

- Specify units of measurement.

- Identify any specialized training requirements (e.g., H_2S training certification).

- Assign responsibility to a single individual over QA/QC for each type of data.

4.3 QA/QC During Data Acquisition

As noted above, QA/QC during data acquisition is the responsibility of the individual Bell Creek Project task leader and is beyond the scope of this Data Management SOP. However, the following QA/QC elements should be considered during the Acquire stage:

- Field sampling and analysis plan (SAP) or other field document which clearly identifies:
  - Number of samples
  - How many sampling locations
  - Number of samples at each location
  - Number of composites (if any)
  - Number of QC samples (e.g., field replicates, etc.)

- Sampling methods
  - Standard operating procedures
  - Field calibrations procedures and frequency (e.g., daily)
  - Sample volumes and storage containers

- Sample handling and custody
  - Maximum hold times
  - Sample storage temperature (e.g., 4°C)
  - Chain-of-custody forms

- Laboratory analysis plan
  - Laboratory analytical methods
  - Performance criteria (e.g., method detection limits [MDL], practical quantitation limit [PQL], number of QC samples [lab duplicates], etc.)

- Any specialized instructions for specific types of data collection.

4.4 QA/QC During Data Storage

As noted in Section 4.2 above, the destination file directory for each type of data that will be acquired should be designated during the Plan stage prior to the Acquire stage. In addition, a single individual from the Project Team should be assigned responsibility over QA/QC for each type of data. These two steps help ensure that each data type is properly stored on the Bell Creek Confidential Shared Folder.

The purpose of the "QA/QC During Data Storage" process is to describe the uniform procedures that should be followed for field, laboratory, and other project data management from the point of data acquisition to storage on the project shared folder. The process entails mapping the flow of project data (project data stream) and identifying the appropriate file directory so that each data type can be saved on the project shared folder using the appropriate file directory and file-naming convention. This process is described below.

PLEASE NOTE: The QA/QC steps outlined in Section 4.4 are for *raw (i.e., unprocessed) data*. Raw data are the data as they were recorded in their native format, either in the field or in the laboratory, and do not reflect any postcollection data reduction or data processing. The QA/QC steps for processed data will be discussed in Section 4.5.

For all data types, the flow of information from the point of data acquisition to the point of data storage would follow the steps listed below (Figure 4-1):

- **Step 1 (Plan Stage) –** Define the destination file directory for the specific type of data that will be acquired. Assign responsibility to a single individual over QA/QC for the specific type of data that will be acquired. This individual will be referred to as the "Data Steward." A ***Data Steward*** is someone who manages one or more specific data files or data file types to ensure that they are accurate and are saved appropriately in accordance with the Data Management SOP.

- **Step 2 (Acquire Stage) –** Acquire data.

- **Step 3 (Check Gate 1: Documentation) –** Deliver field documentation (field books, field forms, etc.), any chain-of-custody files, and any hard copy or electronic data to

the Data Steward who was designated in Step 1. For laboratory data, the Data Steward will receive the chain-of-custody files from the laboratory and not the EERC field personnel. The Data Steward will cross-check the field documentation and chain-of-custody files (if available) against the scope of work for the data collection that was defined during the Plan stage. Specific checks include:

- Have all of the data been collected that were planned?
- Were there discrepancies between the Plan and the actual data that were collected (e.g., alternate sample locations, samples that could not be collected, etc.)?

- **Step 4 (Check Gate 2: Data QC) –** The Data Steward will perform a Tier 1 QC check to look for noticeable errors in the data. Specific checks include:
  - Are the data in the correct format?
  - Are the correct measurement units reflected in the data?
  - Are there visible entry errors?

- **Step 5 (Storage) –** The Data Steward will save the field documentation, chain-of-custody, and electronic data to the Bell Creek Confidential Shared Folder according to the Data Management SOP.

Appendix B provides detailed instructions for specific types of project data and the correct file path for saving raw (as-received) data to the project shared folder. To facilitate the use of this Data Management SOP as a reference document for use by the EERC staff, the specific file directories are listed for each of the 20 different top-level folders so that users may simply go to their specific data type to find the correct file directory. Many of the steps are identical across data types and are intuitive based on the mapping of project data streams and the file directory structure on the project shared drive. Additional notes for each data type are also provided.

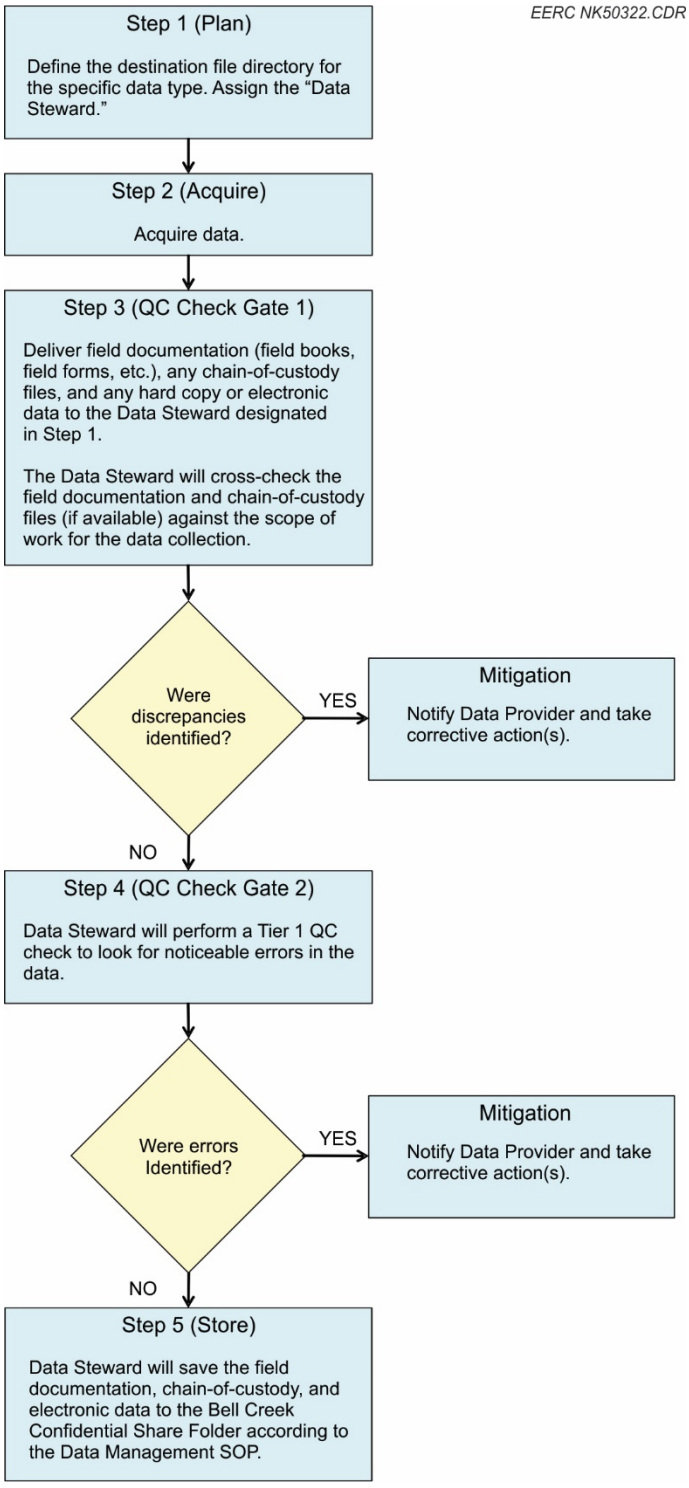


Figure 4-1. QA/QC procedures from the point of data acquisition to the point of data storage.

4.5 QA/QC During Data Processing

For most project data, the raw data will be modified from its native, "as-received" format into a different format for subsequent analysis. This step is referred to as the "Process" stage in the data life cycle. During the Process stage, two fundamental principles apply:

- First, the raw data or "source file" must not be altered from its original format nor may it be moved from its original storage location. The raw data may be viewed as the "recovery file"; i.e., if the processing creates an error in the data, then the user may ultimately go back to the raw data to start over and recover. Therefore, by maintaining the unaltered raw data in the project shared folder, the source data are not lost, and the major impact to the Bell Creek Project is lost time (as opposed to lost data).

- Second, methods of processing data must be thoroughly documented. File directory and file-naming conventions should be followed so that subsequent users can effectively identify the most current data and whether the data represent an interim work product (i.e., subject to change) or a final work product.

Section 4.4 described the procedures for saving raw data to the project shared folder. The following paragraphs describe the QA/QC procedures that are to be used for maintaining data quality during the Process stage. The procedures rely upon processed data subfolders (to separate raw data from processed data) and file-naming conventions (for version control).

The purpose of the "QA/QC During Data Processing" process is to describe the uniform procedures that should be followed for field, laboratory, and other project data management from the point of processing a raw data file (i.e., modifying a data file from its native, as-received format) to storage of the processed data file on the project shared folder. This process is described below. A general guideline to follow this process includes the following steps listed below:

- **Step 1 –** Open the raw data file from the project shared folder.

- **Step 2 –** Save the data file in the specified processed data subfolder using the file-naming conventions described in Section 2.0.

- **Step 3 –** Process the data file as-needed to prepare the data for analysis. Continue to follow the file-naming conventions for date and version control.

Please note that in some circumstances, individuals may need to work with the raw data file directly. In this case, best practice is to store a copy of the raw data file in the appropriate folder and work from this file, thus preserving the integrity of the original raw data.

4.5.1 Processed Data Subfolders

Raw data are saved into subfolders with names that imply raw data. Upon processing these raw data files, the user will save the new file into subfolders with names that imply processed

data. This subfolder management is the first tier in the QA/QC process to separate raw data from processed data.

Specific processed data file directories are listed in **Appendix C** for each of the 20 different top-level folders so that users may simply go to their specific data type to find the correct file directory for processed data. Additional notes for each data type are also provided.

4.5.2 File-Naming Conventions

File-naming conventions were described in detail in Section 3.0. The same file-naming conventions apply to processed data (i.e., Descriptor_YYYY-MM-DD_Version). As noted above, the processed data subfolders serve to separate raw data from processed data, while the file-naming conventions serve to designate version control. If a raw data file is processed only once, then subsequent file-naming conventions are irrelevant. However, if a raw data file will continue to be processed and revised over time, then the Date (YYYY-MM-DD) and Version (V001, V002, etc.) file-naming elements will serve to identify the current work product version.

5.0 Shared Folder Access Controls

Read and write access controls to specific project folders and subfolders within the Bell Creek Confidential Shared Folder have been established to protect critical folders and to limit access beyond specific Project Team members. The Project Manager (John Hamling) and Network Administrator (Josh Mason) have administrator-level permission on all folders. Each Project Team member has read-only access to all folders, meaning that they may open and view all folder contents. However, write access (i.e., the ability to save and rename files within a particular folder) has been restricted. Additional write access levels will be established for key data managers and champions corresponding to their area(s) of expertise and responsibility effective May 19, 2014.

Read-only access provides all Project Team members access to all project data. However, allowing write access to all Project Team members would hinder the ability to effectively control quality and ensure adherence to the Data Management SOP. The Folder Manager, as designated in Section 7.2, will work with management to determine which team members will have write-access to the current parent folders and subfolder structure. These permissions can be reviewed and modified in the future as team member responsibilities change.

6.0 Shared Folder Backup Policy

Steps must be taken to protect data from accidental data loss or file corruption. This process includes routinely making additional "backup" copies of data files that can be used to restore the original data or for recovery of earlier versions of the data. The backup policy for the new Bell Creek Project shared folder on the "zfs" server will incorporate a fixed-frequency automated backup that may be supplemented with a manual backup as-needed and as directed by the appropriate Project Team member(s).

6.1 Current Risk of File Loss

Efforts to quantify the rate of files that were saved to the current shared folders ("bellconf" and "pcorwater") have shown that the rate was highly variable. For example, between November 6, 2013, and March 21, 2104 (135 days), the number of files saved to the "bellconf" and "pcorwater" shared folders was 62,919, resulting in an average daily rate of files saved to the servers of about 466 files and a 95% credible interval of 12 to 1719 files. However, if we update that measurement with the number of files saved to these folders between March 21 and March 28, 2014 (7 days), then the average daily rate of files saved to the servers was only 20 files with a 95% credible interval of 17 to 24 files, which is significantly less (Figure 6-1).

The upper estimates for the current risk of file loss in any given 1-day and 7-day period are approximately 24 files and 166 files, respectively. However, the specific file types and relative value of each file cannot be quantified at this time.

Based on this variable rate of file accumulation, a weekly backup will be established as a balance between the risk of file loss and the burden to the servers for implementing the backup. However, this process will be evaluated over time, and if needed, the backup can occur daily (every evening).

6.2 Fixed-Frequency Automated Backup

The fixed-frequency automated backup will occur once a week (every 7 days) on Saturday. Backup details are described in Section 6.3 and 6.4 below.

6.3 Supplemental Manual Backup

The addition of large numbers of files, or large-sized files, likely occurs in discrete events that correlate with either (1) new data coming into the EERC (e.g., seismic data set, MVA event, etc.) or (2) major modeling and/or simulation efforts. Therefore, to avoid file loss during these events, a manual backup may be used to supplement the fixed-frequency automated backup.

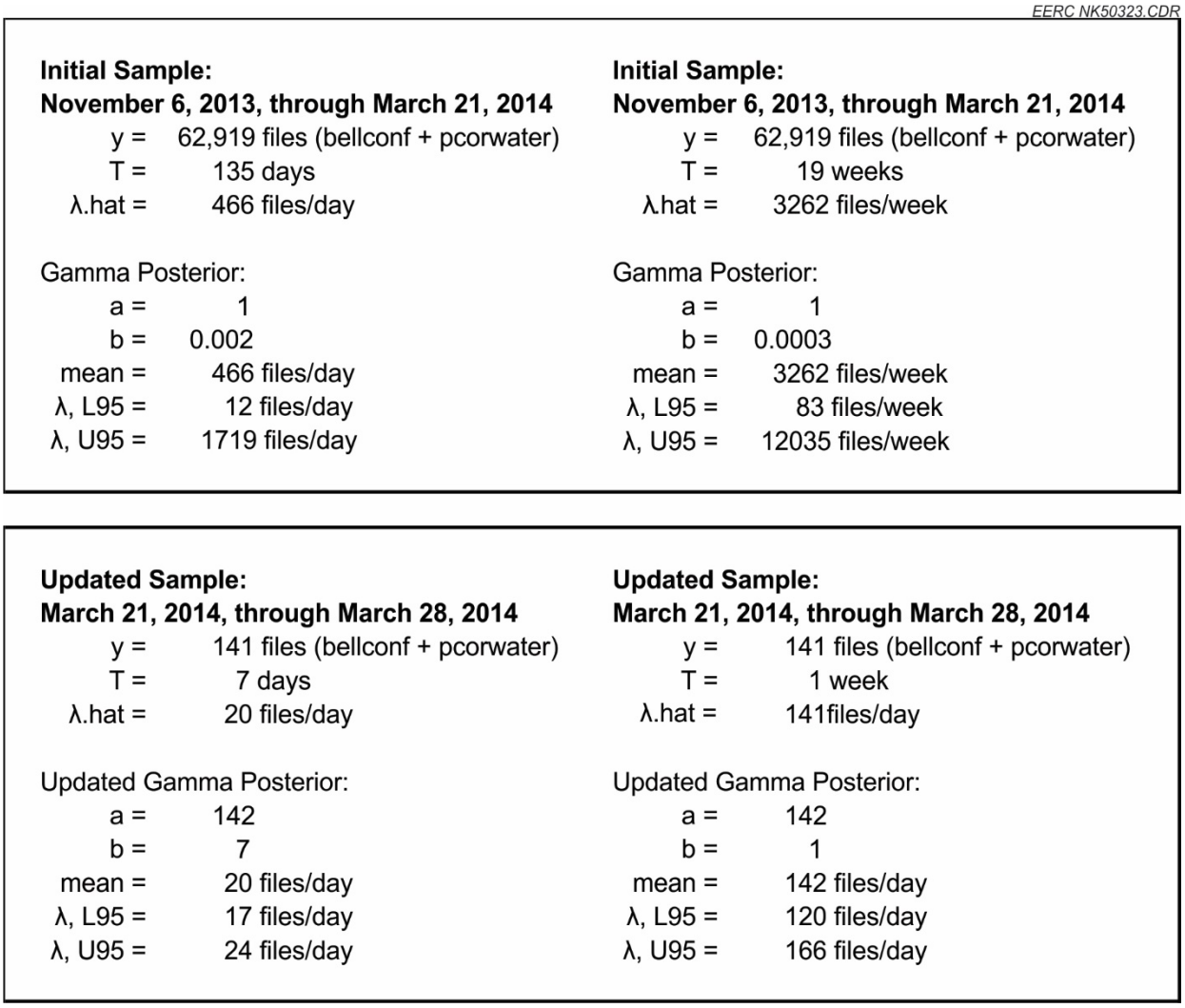


Figure 6-1. Bayesian statistical approach for quantifying the daily and weekly rate of files that were saved to the current shared folders ("bellconf" and "bconf_mva") between November 6, 2013, and March 21, 2014 (135 days – top panel) and updated with new information collected between March 21 and March 28, 2014 (7 days – bottom panel).

For example, if a large simulation were conducted on Monday and resulted in numerous additional files being added to the server, then the Project Team member responsible for these data may want to have the server backed up that evening to avoid the potential loss of these data between Monday and the Saturday (the regularly scheduled backup). To request a manual backup of the server, the Project Team member must submit a request via e-mail to a Network Administrator (Josh Mason) with a copy to the Project Manager (John Hamling) and Research Engineer responsible for the new shared folder (Nick Kalenze). The Network Administrator will document the request, implement the manual backup, and notify the individual who submitted the request that the backup was completed successfully.

Supplemental manual backups are to be used sparingly based on the best professional judgment of the individuals responsible for adding significant new data to the project shared folder. If supplemental manual backup requests become too frequent, then the project should reevaluate its fixed-frequency automated backup schedule to determine whether more frequent backup (e.g., daily) is warranted.

6.4 Backup Details

There will be one backup version of the Bell Creek Confidential Shared Folder. Subsequent backups will consist of only files that were modified since the previous backup. The fixed-frequency automated backup will occur once a week (every 7 days) on Saturday. **In the event of an unintended and accidental manipulation to file structure, data modification, data corruption, or data deletion, the user would have until Saturday (weekly backup) to recover the data. In the event that a potential data recovery is needed, the system administrator (Josh Mason) should be notified immediately so that a recovery procedure can be enacted and the automatic backup taken offline if necessary**.

The primary server currently has 2.99 TB of storage space and is located in the datacenter in EERC Building B. Backups are stored on an independent server that is located in the EERC Hydrogen Building to provide some protection in the event of a disaster. The backup server is running antivirus software to protect from virus damage.

Both the main server and backup server have redundant hard drives to avoid data loss.

Both the main server and backup server have redundant power in case one power supply fails. The backup server currently has a battery backup in the event of a complete power failure. The main server will soon be on battery as well.

In the event of data loss on the main server, the process for recovering data from the backup server requires that the Project Manager contact someone in the Computer Systems group to assist with data recovery.

7.0 Policy Adherence Initiative

7.1 Purpose

As noted in the Introduction, project data are valued assets that need to be managed over their entire life cycle beyond the immediate need. Ultimately, most if not all of the Bell Creek Project data will be delivered to DOE NETL and/or Denbury. Therefore, the organization and secure long-term storage of all project data are essential not only for the current efficiency of the project, but for the eventual transfer of data to the client(s). The Data Management SOP was a significant investment of resources by the EERC and represents both a technical and business decision. The development team made a strong effort to accommodate individual preferences during the development of the Data Management SOP. All Bell Creek Project Team members are expected to comply with the Data Management SOP.

This Policy Adherence Initiative was developed to help ensure that the Bell Creek Project Team members are complying with the Data Management SOP by:

(1) Saving raw and processed data into the appropriate file directory folders and subfolders.

(2) Using the established file-naming conventions.

(3) Incorporating the QA/QC procedures for managing data quality throughout the data life cycle.

A series of steps will be used to help monitor compliance with the Data Management SOP and to provide feedback to the Project Team members.

7.2 Folder Managers

Folder assignments were developed as part of the effort to combine all Bell Creek data onto a single shared folder on the "zfs" server. Existing folders located on one of three different shared folders: ("bellconf," "bconf_mva," or "denbury_imagery") were assigned to one or more Project Team members. These Project Team members are individuals with direct knowledge of the folder contents and the requisite project knowledge to properly refile the folder contents into the new file directory that was established in the Data Management SOP.

Table 7-1 lists specific top-level directories and subdirectories for the Bell Creek Confidential Shared Folder and the individuals to whom responsibility has been assigned for migrating data into the folder. These individuals are the "Folder Managers" for these specific folders. In addition to migrating data from the older shared folders to the new Bell Creek Confidential Shared Folder, the Folder Managers will also be responsible for helping to monitor compliance with the Data Management SOP for all files within their respective folders.

Project Folder/Subfolder Name	Folder Manager
100_Project_Management	John Hamling
200_Soil_Gas	Nick Kalenze
300_Surface_Ground_Water	Nick Kalenze
400_PNL	Neil Dotzenrod
500_Seismic	Shaughn Burnison
600_Pressure_Temperature	Ryan Klapperich
700_Production_Injection_Rate_Data	Kerryanne Leroux
800_Fluid_Chemistry	Gavin Liu
900_Well_Files	Kyle Glazewski
1000_Well_Logs	Jason Braunberger
1100_Core_Data_and_Testing	Loreal Heebink
1200_Images	Jib Wilson
1300_Literature	Ryan Klapperich
1400_Modeling	Jason Braunberger
1500_Simulation	Gavin Liu
1600_In_House_CO2_Experiments	Gavin Liu
1700_GIS	Kyle Glazewski
1800_Regulatory	Lisa Botnen
1900_Project_Documents	Nick Kalenze
2000_Observation_and_Redrill_Wells	Ryan Klapperich

Table 7-1. List of specific top-level directories and associated subdirectories for the Bell Creek Confidential Shared Folder and the Folder Managers to whom responsibility has been assigned.

7.3 Monitoring Schedule

Once the project data were successfully migrated to the Bell Creek Confidential Shared Folder, a monitoring plan was implemented to monitor the project top-level directories and subdirectories. The monitoring schedule is phased to aid in the process of adjusting and adapting to the new procedures. Therefore, higher-resolution (more frequent) monitoring occurs in the first 6 months, followed by a tapered frequency of monitoring to 12 months and beyond. Table 7-2 shows the phased monitoring schedule. The details of the monitoring plan are discussed in Section 7.4.

Monitoring Frequency	Implementation Schedule
Weekly	Month 1
Monthly	Month 2 through Month 6
Quarterly	Month 6 through Month 36

Table 7-2. Phased schedule for monitoring compliance with the Data Management SOP.

7.4 Monitoring Plan

The individuals in Table 7-1 will visually inspect their specific top-level directory and/or subdirectory contents following the monitoring schedule. The visual inspection will check to see that ***raw and processed data are saved into the appropriate file directory and subdirectories*** **and that the established file-naming conventions are being implemented for all new files.**

EXCEPTION: Files created prior to the migration to the new shared folder will maintain their original file name. However, all subsequently created files will adhere to the file-naming conventions outlined in the Data Management SOP.

A 36-month monitoring plan is entered into Microsoft Outlook and built into the resource planning for the Bell Creek Study. Individuals will receive a calendar notification on the last day of the week, last business day of the month, or last business day of the quarter reminding them to monitor their respective folder(s).

7.5 Folder Inspection Checklist

The Folder Manager should review files by file date, assessing each file that was created during the week, month, or quarter. The monitoring need not be extensive but should verify the following checklist of items:

- **Have new raw data files been added to the project folder/subfolder since the prior inspection?** If the answer is "no," then there is nothing to report. If the answer is "yes," then the following checklist should be reviewed.
  - Have raw data files been saved to the appropriate folder/subfolder as per Appendix B of the Data Management SOP?
  - Have raw data files been saved using the file-naming conventions as described in Section 3 of the Data Management SOP (i.e., Descriptor_Date_Version)?
  - Did the raw data go through the proper QA checks prior to saving on the shared folder as per Section 4 and Figure 4-1 of the Data Management SOP?
    - Were any discrepancies noted with the raw data file such that it was modified from its native (as-received) format?

- **Have files been processed or modified since the prior inspection?** If no file processing has occurred, then there is nothing to report. If the answer is "yes," then the following checklist should be reviewed.
  - Have processed data files been saved to the appropriate folder/subfolder as per Appendix C of the Data Management SOP?
  - Have processed data files been saved using the file-naming conventions as described in Section 3 of the Data Management SOP (i.e., Descriptor_Date_Version)?

An example checklist table that will be used by the Folder Manager to record the outcome of the folder inspection is provided in Table 7-3. An example of a completed checklist is provided in Table 7-4. Project Team members may develop their own record-keeping format, so long as it includes the core elements described above and captured in Table 7-3.

Discrepancies with the Data Management SOP should be noted on the checklist. *At this point, the Folder Manager should not immediately correct the error. While the temptation may be to fix the error, this places an unnecessary burden on the individuals with folder responsibilities and does not lead to the long-term improvement of the overall Project Team.* **The individual who saved or created the raw data or processed data file will ultimately make the correction** as part of the feedback program for continuous improvement in the Bell Creek Project data management.

Checklist Item		Outcome
Have new raw data files been added to the project folder/subfolder since the prior monitoring?		Yes / No
Raw Data Files		
Have raw data files been saved to the appropriate folder/subfolder as per Appendix B of the Data Management SOP?		Yes / No
File 1	File name and file directory path	Specific exception to the Data Management SOP (e.g., saved into the wrong folder/subfolder)
File 2	File name and file directory path	
File 3	File name and file directory path	
Etc.		
Have raw data files been saved using the file-naming conventions as described in Section 3 of the Data Management SOP (i.e., Descriptor_Date_Version)?		Yes / No
File 1	File name and file directory path	Specific exception to the Data Management SOP (e.g., incorrect file-naming convention)
File 2	File name and file directory path	
File 3	File name and file directory path	
Etc.		
Did the raw data go through the proper QA checks prior to saving on the shared folder as per Section 4 and Figure 4-1 of the Data Management SOP?		Yes / No
File 1	File name and file directory path	Specific exception to the Data Management SOP (e.g., no QA performed prior to saving file)
File 2	File name and file directory path	
File 3	File name and file directory path	
Etc.		
Processed Data Files		
Have raw data files been saved to the appropriate folder/subfolder as per Appendix C of the Data Management SOP?		Yes / No
File 1	File name and file directory path	Specific exception to the Data Management SOP (e.g., saved into the wrong folder/subfolder)
File 2	File name and file directory path	
File 3	File name and file directory path	
Etc.		
Have raw data files been saved using the file-naming conventions as described in Section 3 of the Data Management SOP (i.e., Descriptor_Date_Version)?		Yes / No
File 1	File name and file directory path	Specific exception to the Data Management SOP (e.g., incorrect file-naming convention)
File 2	File name and file directory path	
File 3	File name and file directory path	
Etc.		

Table 7-3. Example ***blank*** checklist for monitoring compliance with the Data Management SOP. This checklist would be completed by the Folder Manager in the event that one or more discrepancies were noted during the folder inspection.

Checklist Item		Outcome
Have new raw data files been added to the project folder/subfolder since the prior monitoring?		Yes
Raw Data Files		
Have raw data files been saved to the appropriate folder/subfolder as per Appendix B of the Data Management SOP?		No
File Name	ZFS:\Bell_Creek_Shared_Folder\200_Soil_Gas\Raw_Data\Isotech_Labs	Data should have been saved into the "Laboratory_GC" subfolder.
File Name	ZFS:\Bell_Creek_Shared_Folder\200_Soil_Gas\Raw_Data\Field_GC	Data should have been saved into the "Handheld_Meter" subfolder.
Have raw data files been saved using the file-naming conventions as described in Section 3 of the Data Management SOP (i.e., Descriptor_Date_Version)?		Yes
Did the raw data go through the proper QA checks prior to saving on the shared folder as per Section 4 and Figure 4-1 of the Data Management SOP?		Yes
Processed Data Files		
Have raw data files been saved to the appropriate folder/subfolder as per Appendix C of the Data Management SOP?		No
File Name	ZFS:\Bell_Creek_Shared_Folder\200_Soil_Gas\Processed_Data\Operational	Data are raw data and should not have been saved in the "Processed_Data" subfolder.
Have raw data files been saved using the file-naming conventions as described in Section 3 of the Data Management SOP (i.e., Descriptor_Date_Version)?		Yes

Table 7-4. Example ***completed*** checklist for monitoring compliance with the Data Management SOP for the 200_Soil_Gas folder.

7.6 Feedback to Project Team Members

Each Project Team member will have the opportunity to attend the training session that reviews the Data Management SOP prior to migrating the Bell Creek data to the Bell Creek Confidential Shared Folder. This training session will provide an opportunity for individuals to see the new processes that have been enacted and to ask questions. However, on occasion and particularly during the early phases of implementing the Data Management SOP, the Project Team members may revert to their former ways of saving or naming files that are inconsistent with the Data Management SOP. **The ultimate goal is for the processes outlined in the Data Management SOP to become second-nature to the entire Project Team, such that all future project data are saved and named according to the common conventions across the data life cycle.** Therefore, the monitoring plan and checklist will be combined with a feedback program to help the Project Team adapt to the new Data Management SOP processes.

The checklist in Table 7-3 will be completed weekly, monthly, or quarterly as follows the monitoring schedule. This checklist will then serve as the list of individuals to whom feedback must be given about one or more files that were saved in violation of the Bell Creek Data Management SOP. The Folder Manager will provide feedback to the individual responsible for the specific data file in violation. The individual who saved or created the raw data or processed data file will ultimately make the correction. Additional training may be requested by any individual or directed at the discretion of a Folder Manager by speaking with the Research Engineer (Nick Kalenze) responsible for administration of Bell Creek Data Management activities. Continued noncompliance with the Bell Creek Data Management SOP will be viewed as a violation of the Bell Creek Data Management Policy. Insubordination may result in disciplinary action up to and including termination. Section 8.0 provides a procedure for the modification of the Bell Creek Data Management SOP in order to improve functionality and/or to better facilitate compliance with the Bell Creek Data Management Policy.

8.0 Change Management Plan

The Bell Creek Project may change over time such that the Data Management SOP would need to adaptively change to meet the needs of the project. The Change Management Plan is a structured process for modifying the Data Management SOP.

8.1 Change in the File Directory Structure or Access Controls

To request a change in the file directory structure or access controls outlined in this Data Management SOP, the Project Team staff must first notify the Folder Manager with a description and explanation of how the desired change would benefit the project or SOP. If the Folder Manager deems the request in compliance with the Bell Creek Data Management Policy and agrees that the recommendation will improve functionality, they will, in turn, notify and recommend the change to the Research Engineer responsible for the Bell Creek Confidential Shared Folder (Nick Kalenze) and the Project Manager (John Hamling). If the file directory structure change request is approved, then the Project Manager will notify the Network Administrator (Josh Mason) to implement the change. The Network Administrator will document the request, implement the change to the Bell Creek Confidential Shared Folder, and notify the individual who submitted the request that the change was completed successfully.

The Data Management SOP will be changed accordingly, a new version will be posted to the Bell Creek Confidential Shared Folder, and all Project Team members will be notified of the new file directory structure or access controls.

8.2 Other Changes

To request a change to a broader process that is outlined in this Data Management SOP, for example where specific files are saved or file-naming conventions for a specific file type, the process is similar as the one outlined above but does not require the involvement of the Network Administrator.

To request a change to a broader process that is outlined in this Data Management SOP, the Project Team staff must first notify Folder Manager with a description and explanation of how the desired change will benefit the project or SOP. If the Folder Manager deems the request to be in compliance with the Bell Creek Data Management Policy and agrees that the recommendation will improve functionality, the Folder Manager will notify and recommend the change to the Research Engineer responsible for the Bell Creek Confidential Shared Folder (Nick Kalenze) and the Project Manager (John Hamling). If the change request is approved, then the Project Manager will notify the individual who submitted the request that the change was accepted.

The Data Management SOP will be changed accordingly, a new version will be posted to the Bell Creek Confidential Shared Folder, and all Project Team members will be notified of the new process.

8.3 Project Management Team Review

As previously discussed, data are considered a valued and protected resource to the EERC, the PCOR Partnership, and the Bell Creek Study. Adherence to the Data Management SOP is mandatory for all personnel who utilize Bell Creek data. It is the personal responsibility and prerogative of all personnel who utilize Bell Creek data to ensure compliance with the policy as outlined in the Data Management Standard Operating Procedure.

The Project Management Team recognizes that the Data Management SOP may need to adaptively change to meet the needs of the project. Therefore, a formal review is scheduled at 8 weeks, 6 months, and 1 year to assess the current Data Management SOP, evaluate feedback from Project Team members, and consider changes to the Data Management SOP.

In addition to the folder monitoring discussed in Section 7.3, an open-door policy is in place where Project Team members may provide feedback to the Project Manager (John Hamling), the Research Engineer responsible for the new shared folder (Nick Kalenze), or other members of the Project Management Team. Project Team members are encouraged to bring forth ideas for improvement and suggested solutions for the Data Management SOP, particularly if they feel that there are better, more functional, or more effective processes that can be enacted for specific file types or subfolders. Everyone on the Project Team should feel empowered and encouraged to both bring forth new ideas and concerns and to know that their ideas and concerns will be taken seriously.

The Project Management Team will formally review this feedback at 8 weeks, 6 months, and 1 year, which coincides with the folder-monitoring activities. This formal review will then determine whether or not changes need to be made to the Data Management SOP. If it is determined that changes need to be made, then a revised Data Management SOP will be generated, and the Project Team will be notified via e-mail that a new document is available.

If there is a major concern that arises prior to a scheduled review, then that concern should be brought to the attention of the Project Manager, and an immediate change to the Data Management SOP may be enacted.

A primary goal of the Data Management SOP is to make the system as functional as it can be for both the Project Team and the EERC. The Project Management Team review is aimed at ensuring that the long-term evolution of the Data Management SOP is consistent with that goal.

9.0 Sources

Chapman, A.D., 2005, Principles of data quality [PDF], version 1.0: Report for the Global Biodiversity Information Facility, Copenhagen.

NSF (National Science Foundation), 2014, Policy on dissemination and sharing of research results: www.nsf.gov/bfa/dias/policy/dmp.jsp (accessed March 5, 2014).

USEPA (U.S. Environmental Protection Agency), 2002, Guidance for quality assurance project plans: U.S. Environmental Protection Agency, Office of Environmental Information, EPA\240\R-02\009, Washington, D.C.

USGS (U.S. Geological Survey), 2014, Science data life cycle model for the USGS: Model Development Workspace, https://my.usgs.gov/confluence/display/cdi/Science+Data+Life+Cycle+Model+for+the+USGS%3A+Model+Development+Workspace (accessed March 5, 2014).