



Data Management Plan

Upper Columbia Basin Network

Version 1.0, December 2007

Natural Resource Report NPS/UCBN/NRR—2007/020



ON THE COVER

NPS Collage of data management photographs in the UCBN, courtesy of UCBN staff

Upper Columbia Basin Network Data Management Plan

Version 1.0, December 2007

Natural Resource Report NPS/UCBN/NRR—2007/020

Gordon H. Dicus
National Park Service, Upper Columbia Basin Network
University of Idaho, Department of Fish and Wildlife
Moscow, ID 83844-1136

Lisa K. Garrett
National Park Service, Upper Columbia Basin Network
University of Idaho, Department of Fish and Wildlife
Moscow, ID 83844-1136

December 2007

U.S. Department of the Interior
National Park Service,
Natural Resource Program Center
Fort Collins, Colorado

The Natural Resource Publication series addresses natural resource topics that are of interest and applicability to a broad readership in the National Park Service and to others in the management of natural resources, including the scientific community, the public and the NPS conservation and environmental constituencies. Manuscripts are peer-reviewed to ensure that the information is scientifically credible, technically accurate, appropriately written for the intended audience and is designed and published in a professional manner.

Natural Resource Reports are the designated medium for disseminating high priority, current natural resource management information with managerial application. The series targets a general, diverse audience and may contain NPS policy considerations or address sensitive issues of management applicability. Examples of the diverse array of reports published in this series include vital signs monitoring plans; "how to" resource management papers; proceedings of resource management workshops or conferences; annual reports of resource programs or divisions of the Natural Resource Program Center; resource action plans; fact sheets; and regularly-published newsletters.

Views and conclusions in this report are those of the authors and do not necessarily reflect policies of the National Park Service. Mention of trade names or commercial products does not constitute endorsement or recommendation for use by the National Park Service.

Printed copies of reports in these series may be produced in a limited quantity and they are only available as long as the supply lasts. This report is also available from the Upper Columbia Basin I&M Network website (<http://www.nature.nps.gov/im/units/UCBN>) on the internet, or by sending a request to the address on the back cover.

Please cite this publication as:

Dicus, G.H. and L.K. Garrett. 2007. Upper Columbia Basin Network Data Management Plan Version 1.0, Natural Resource Report NPS/UCBN/NRR—2007/020. National Park Service, Moscow, Idaho.

Change History

Original Version #	Date of Revision	Revised By	Changes	Justification	New Version #
1.0					

Notes on use of Change History log: Generally, version numbers will increase incrementally by whole numbers (e.g., version 2.0, 3.0, 4.0, etc.) as the document undergoes significant revisions. However, version numbers may increase by tenths (e.g., version 1.1, 1.2, etc) to reflect minor changes. A brief description of changes (e.g., Chapter number or sub-section) and a brief justification for the changes will be noted in this Change History log.

Note on Organization of this Document: Guidance documents and informational overview documents are bound in a separate accompanying document and included on a CD that accompanies this document.

Informational Overview Document
National-level Information Management Applications

Guidance Documents
File Naming Guidance
Directory Structure Guidance
Product Specifications Guidance
Metadata Guidance
Product Certification Form
Photo Management Guidance
Photo Database User Guidance

Contents

Page

Appendixes	vii
Figures	ix
Tables	xi
Executive Summary	xiii
Acknowledgements	xv
Chapter 1. Introduction	1
The Upper Columbia Basin Network Inventory and Monitoring Program	1
Data Management Goals and Objectives	2
Types of Data Included in this Plan	3
Sources of Natural Resources Data and Network Priorities	4
Scope and Organization of this Plan	5
Chapter 2. Roles and Responsibilities	7
Data Stewardship Roles	8
Documentation is Key	9
The Hub of Data Stewardship	9
Coordination	12
Chapter 3. Infrastructure and System Architecture	13
Computer Resources Infrastructure	13
National System Architecture	14
Network System Architecture and Centralized Directory Structure	15
Chapter 4. Project Work Flow and Data Life Cycle	19
Project Work Flow	19

Contents (continued)	Page
Data Life Cycle	23
Database Application Development	25
Integrating and Sharing Data Products	26
Chapter 5. Acquisition, Processing and Reporting	27
National Laws and Policies	27
Network Standards and Guidelines	27
Programmatic Data	28
Non-Programmatic Data	31
Analysis and Reporting	34
Changes to Procedures	34
Chapter 6. Quality Assurance and Quality Control (QA/QC)	37
NPS Mandate for Quality	37
Goals and Expectations	39
Roles and Responsibilities	40
General Operations	40
Data Collection	41
Data Entry	42
Data Verification	42
Data Validation	43
Review, Conformance and Communication	45
Chapter 7. Documentation	47
Laws, Policies and Mandates	47

Contents (continued)

Page

Network Standards	47
Content Standards	48
Format and Storage Standards	50
Metadata Tools	50
Work Flow Process	51
Chapter 8. Ownership, Sharing and Dissemination	53
Laws, Policies and Mandates	53
Network Standards and Guidelines	54
Data Requests	56
Feedback Mechanisms	57
Chapter 9. Records Management	59
National Policies	59
Network Objectives	59
General Guidelines	60
Chapter 10. Implementation	67
Education and Training	67
Measuring Success	68
Plan Revisions	68
Literature Cited	69
List of Acronyms and Abbreviations	71
Glossary of Terms Used in the Data Management Plan	73

Appendixes

Note: bound in a separate accompanying document and included on a CD that accompanies this document

Appendix A. National Information Management Applications

Appendix B. UCBN File Naming Guidance

Appendix C. UCBN Directory Structure Guidance

Appendix D. UCBN Product Specifications Guidance

Appendix E. UCBN Metadata Guidance

Appendix F. UCBN Product Certification Form

Appendix G. UCBN Photo Management Guidance

Appendix H. UCBN Photo Database User Guidance

Figures

Page

Figure 1.1. Map of Upper Columbia Basin Network parks.	2
Figure 1.2. Context of national I&M program guidelines, the UCBN Data Management Plan, Network-wide guidance and protocol-specific data management procedures.	5
Figure 2.1. Core data management responsibilities for the Network coordinator, data manager and project leader(s).	9
Figure 3.1. Main components of the UCBN infrastructure.	13
Figure 4.1. Generalized project work flow model for short- and long-term data collection efforts.	20
Figure 4.2. Diagram of the typical project data life cycle.	25
Figure 5.1. General flow of data acquisition and processing.	27
Figure 6.1. General flow of data and associated Quality Assurance/Quality Control procedures.	38
Figure 6.2. Common data management elements influencing the amount of QA/QC needed.	39
Figure 8.1. Data flow diagram for water quality data.	56
Figure 9.1. General workflow for storage and archiving of electronic and hardcopy products. .	60

Tables

Page

Table 1.1. Categories and examples of data products covered by this Data Management Plan. ...	4
Table 2.1. Categories of data stewardship involving all Network personnel and cooperators.	7
Table 2.2. Programmatic roles and associated data management responsibilities.	8
Table 3.1. General directory structure used on the UCBN Network Data Storage unit.	16
Table 4.1. An example of tasks per project stage and staff responsibilities.	22
Table 5.1. Satellite resolution, swath width, area, cost and history of various remote sensing imagery sources (sorted by cost).	30
Table 5.2. Summary of possible data sources for different types of information and repository where it should be maintained.	32
Table 8.1. Primary Repositories for UCBN products.	55

Executive Summary

A central mission of the National Park Service Inventory and Monitoring (I&M) Program is to provide timely, accurate, and usable scientific information about the status and trends of park resources to park managers. Fulfilling this mission requires an information management system that effectively produces, maintains, and distributes the products of scientific work completed in the parks. This Upper Columbia Basin Network (UCBN) Data Management Plan (DMP) documents an overarching strategy for ensuring that UCBN program data are documented, secure, and remain accessible and useful for decades into the future. In turn, the DMP refers to more specific guidance documents and standard operating procedures which convey the standards and steps necessary for individual monitoring protocols.

The UCBN data management mission is to ensure the quality, interpretability, security, longevity, and availability of program natural resource data. The procedures and work practices outlined by this DMP contribute to these four primary data management goals:

- Ensure the high quality and long-term availability of the ecological data and related analyses produced from the network's inventory and monitoring work
- Integrate data management activities with all aspects and at all stages of network business
- Specify data stewardship responsibilities for all personnel
- Work within and outside the network, as appropriate, to improve the quality and availability of legacy NPS datasets and data from outside sources

Effective data management practices are essential to achieving I&M Program goals and require adequate personnel time and funding to achieve the best results. While, some data collection procedures and data management practices are already in use and require minimal revisions, others may involve several iterations before deemed acceptable and functional. Some practices specified in the DMP may be new to staff and cooperators. With few exceptions, however, obligations come from law, Executive Orders, Director's Orders, or national I&M Program guidance. The DMP helps to put these requirements into context and provides operational guidance to achieve them.

This data management plan reflects a commitment to establishing and maintaining a robust system for data management to ensure the availability and usability of high-quality natural resource information. This plan supports the objectives of the I&M Program and provides guidance on the following network activities:

- acquisition and processing of data,
- assurance of data quality,
- documentation, analysis, summarization, and dissemination of data and information,
- maintenance of UCBN content within national-level data management systems,
- and maintenance, storage, and archiving of data products.

Acknowledgments

The Upper Columbia Basin Network Data Management Plan is a product of many efforts. The community of I&M and park data managers across all 32 networks fosters a collaborative environment sharing information freely; and the UCBN benefited greatly. Immense gratitude is extended for the dedication of the original Data Management Planning Workgroup, the continued efforts of “Tier 2” and “Tier 3” data managers and the leadership of Margaret Beer, Peter Budde, Steve Fancy, Lisa Nelson and Joe Gregson. Leona Svancara did an excellent job of developing the first drafts of this Data Management Plan. In addition, Leona pioneered the UCBN data management infrastructure and accomplished enormous strides in organizing all of the UCBN’s initial data assets. Chris Collins at Nez Perce National Historical Park provided IT insight and guidance.

Chapter 1. Introduction

Collecting natural resource data is the first step toward gaining a better understanding of ecosystems within national parks. These “raw” data are used to analyze, synthesize and model aspects of ecosystem form and function. In turn, the results and interpretations are used to make decisions concerning park resources. Thus, *data* collected and maintained by the Upper Columbia Basin Network (UCBN) becomes *information* through analysis, synthesis and modeling. The central mission of the National Park Service (NPS) Inventory and Monitoring (I&M) Program is to provide this scientific information on the status and trends of park resources in a timely and useful fashion. A robust data management system is critical for such a long-term monitoring program where the lifespan of a dataset will encompass the careers of several natural resource managers.

This chapter summarizes the data management goals and objectives of the UCBN data management system. This Data Management Plan (DMP) documents the overarching strategy for ensuring that UCBN program data are documented, secure, and remain accessible and useful for future generations. The DMP, in turn, refers to other guidance documents and standard operating procedures which convey the specific standards and steps for achieving the network’s data management goals.

The Upper Columbia Basin Network Inventory and Monitoring Program

The I&M Program represents a long-term commitment by the NPS to assess and document the status and trends of park ecological resources. In 1998, the National Parks Omnibus Management Act established a framework for the I&M Program, which fully integrates natural resource monitoring and other scientific activities into the management processes of the National Park system. The I&M Program’s long-term goals are to:

- Inventory the natural resources and park ecosystems under National Park Service stewardship to determine their nature and status.
- Monitor park ecosystems to better understand their dynamic nature and condition and to provide reference points for comparisons with other, altered environments.
- Establish natural resource inventory and monitoring as a standard practice throughout the National Park system that transcends traditional program, activity, and funding boundaries.
- Integrate natural resource inventory and monitoring information into National Park Service planning, management, and decision making.
- Share National Park Service accomplishments and information with other natural resource organizations and form partnerships for attaining common goals and objectives.

To accomplish these goals, the national I&M Program grouped parks with significant natural resources into 32 networks based on geographic similarities, common natural resources and resource protection challenges. The network approach facilitates collaboration, information sharing and economies of scale as well as provides a “minimum infrastructure” to initiate natural resource monitoring.

As one of these 32 networks, the UCBN includes nine parks located in four states (Figure 1.1). Oversight to the UCBN is provided by a Board of Directors (see Garrett et al. 2007, Chapter 8).

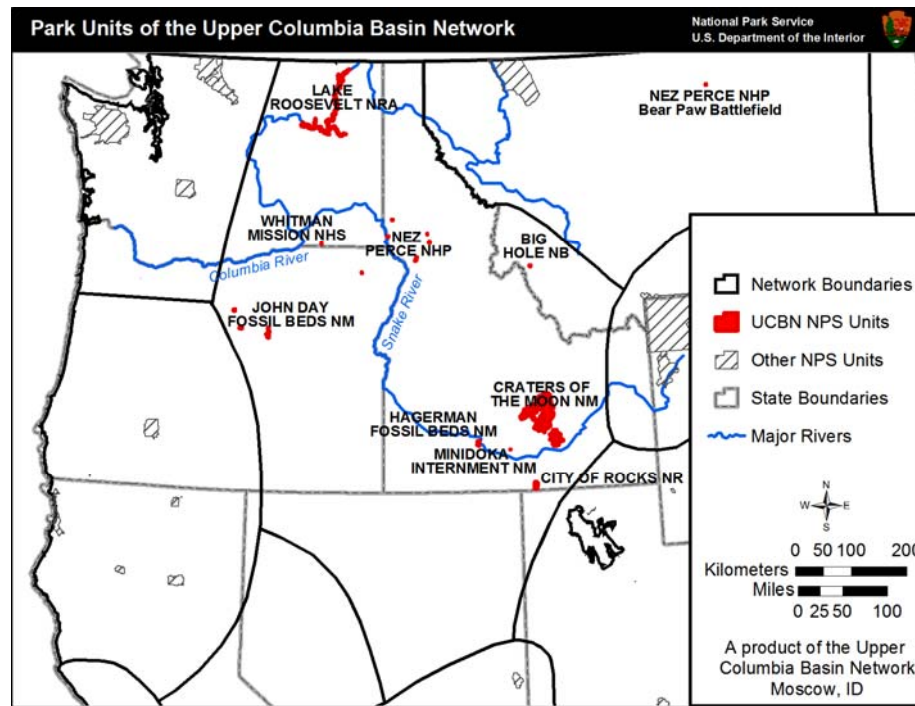


Figure 1.1. Map of Upper Columbia Basin Network parks.

Data Management Goals and Objectives

The I&M Program's data management mission is to ensure the quality, interpretability, security, longevity, and availability of program natural resource data. This Data Management Plan outlines the procedures and work practices that support effective data management. The data management goals and associated objectives of the UCBN are as follows:

- **Goal 1** - Ensure the high quality and long-term availability of the ecological data and related analyses produced from the network's inventory and monitoring work.
 - a. Objective - Outline the procedures and work practices that support effective data management.
 - b. Objective - Establish an organizational plan for UCBN program data and information so that they are retrievable by staff, cooperators, and the public.
 - c. Objective - Establish standards for data, data distribution, and data archiving to ensure the long-term integrity of data, associated metadata, and all supporting information.
- **Goal 2** - Integrate data management activities with all aspects and at all stages of network business.
 - a. Objective - Encourage effective data management practices as an integral part of project management so all data are available and usable for Network and park management decisions, research, and education, now and into the future.
 - b. Objective - Establish quality control and quality assurance standards.

- Goal 3 - Specify data stewardship responsibilities for all personnel.
 - a. Objective - Guide current and future staff of the UCBN to ensure that sound data management practices are followed.
 - b. Objective - Establish data management roles and responsibilities for UCBN staff.
- Goal 4 - Work within and outside the network, as appropriate, to improve the quality and availability of legacy NPS datasets and data from outside sources.
 - a. Objective – As time and resources permit, migrate high-priority legacy datasets into modern formats and improve the quality and documentation of these datasets or other data originating from outside the Inventory and Monitoring Program.
 - b. Objective – Work with partner agencies and institutions to promote the sharing and development of data, software applications, and analyses.

Types of Data Included in this Plan

A distinction is usually made between the term ‘data,’ which refers to assemblages of raw or un-interpreted facts, records or observations, and ‘information,’ which is created from data through the process of analysis, synthesis, modeling, or other types of interpretation. For the sake of discussion in this plan, data will be used to signify both data and information.

This plan encompasses a range of products that are coordinated or managed by UCBN. These products fall into four general categories: data, documentation, reports, and administrative records (Table 1.1). Documentation, in the form of protocols, data dictionaries, database user guides, and standard operating procedures, provides the long-term value of data by setting the context of how and why the data were collected, analyzed, and reported.

Table 1.1. Categories and examples of data products covered by this Data Management Plan.

Category	Description	Examples
Data		
Raw Data	Data obtained from the environment and that has not been subjected to any quality assurance or control beyond those applied during field work.	- Field data sheets - Specimens - Remotely sensed data - Data gathered electronically on field computers - GPS rover files - Photographic imagery
Verified and Validated Data	Data that has been evaluated for completeness, correctness, and conformance/compliance of a specific data set against the standard operating procedure (verified), as well as reviewed for specific analytic quality (validated).	- Relational databases - Tabular data files - Laboratory results - GIS layers - Maps
Analyzed Data	Data that have been subjected to analytical routines after field collection and verification. This includes statistical operations conducted on the data for the purposes of arriving at a measure of the given ecological parameter or a compilation of analyzed data from different sources or time periods to derive new information.	Summarized reports, data and maps from statistical or query operations
Documentation	Documentation provides the information required to understand the context of the data.	- Data collection protocols - Data processing/analysis protocols - Record of protocol changes - Data dictionary - FGDC metadata - Database design documents - QA/QC reports - Catalogs
Reports	Reports provide a means of presenting and publishing the methods and the results of analysis in the context of which it was intended.	- Annual progress reports - Final reports - Trend analysis reports - Publications
Administrative Records	Administrative records supplement the context of a project and should be considered part of the project deliverables.	- Contracts and agreements - Study and work plans - Research permit - Critical administrative correspondence

Sources of Natural Resource Data and Network Priorities

Potential sources of important data and information about the condition of natural resources include:

- I&M Inventories
- I&M Vital Signs Monitoring
- Natural resource projects, monitoring, and research done by park staff and cooperators
- Monitoring or research done by other agencies or on adjacent lands

The priorities for network data management efforts are:

1. Produce and curate high-quality, well-documented data originating from the Inventory and Monitoring Program.
2. Assist with data management for current projects, legacy data and data originating outside the Inventory and Monitoring Program that complement program objectives.
3. Help ensure good data management practices for park-based natural resource projects that are just beginning to be developed and implemented.

Scope and Organization of this Plan

The UCBN DMP is a guide to ensure continuity and documentation of data management methods and procedures over time. This plan reflects a commitment to establishing and maintaining a robust system for data management to ensure the availability and usability of high-quality natural resource data. The DMP describes how the Network will:

- Support I&M Program objectives
- Acquire and process data
- Assure data quality
- Document, analyze, summarize and disseminate data and information
- Maintain high-quality UCBN data within national database systems
- Maintain, store and archive information.

The DMP is designed as a stand-alone document that complements the UCBN Vital Signs Monitoring Plan (Garrett et al. 2007), but may also be used by parks or other programs. It is directed by a wealth of overarching, national-level guidance and is supported by more detailed standard operating procedures (SOPs) and guidance documents (Figure 1.2). These documents describe Network practices as a whole, whereas project-specific details (e.g., methods, frequency and timing of data acquisition) are addressed in detailed, protocol-specific SOPs.

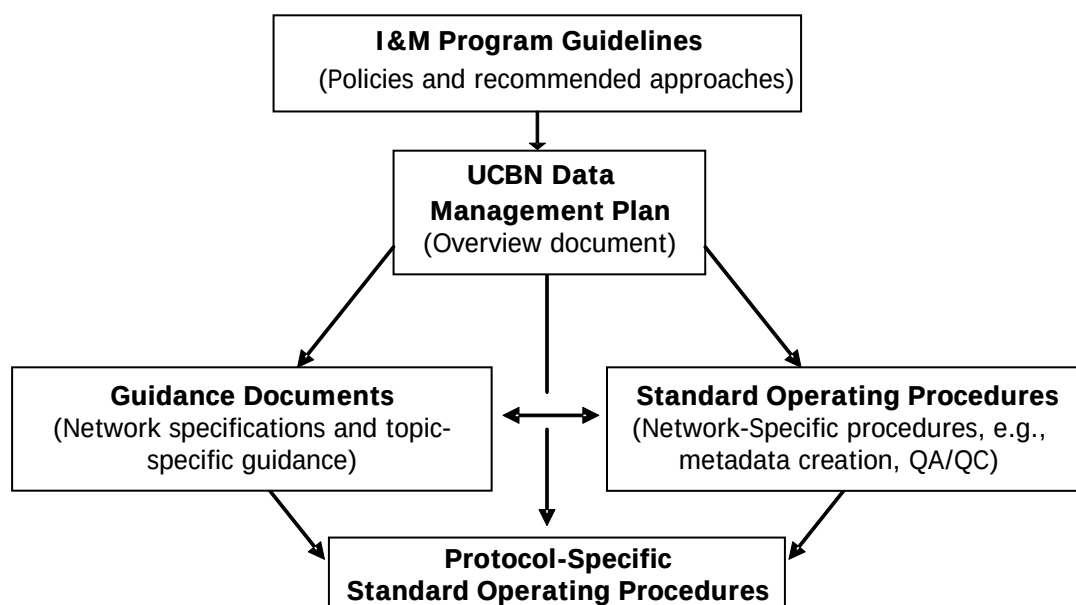


Figure 1.2. Context of national I&M program guidelines, the UCBN Data Management Plan, Network-wide guidance and protocol-specific data management procedures.

The intended audience for this plan includes Network staff and any cooperating NPS staff or agency. To increase the readability and usability among this wide audience, the plan was developed with a basic overview (Chapters 1-4) describing the roles and responsibilities, processes, workflow, infrastructure and framework for data management. It is followed by discussion on the operations of data management (Chapters 5-9), including sections on quality and documentation, and presents legal mandates when applicable. A final chapter addresses education, measuring success, and procedures for plan updating.

Chapter 2. Roles and Responsibilities

Data management is about people and organizations as much as it is about technology, database theory and application. Nearly every person in an organization manages information at some level. Good data stewardship could not possibly be accomplished by “data managers” alone – it is truly a collaborative endeavor that involves many people fulfilling a broad range of roles and responsibilities. In the context of this plan, a *role* is a function or position (e.g., Project Leader) while a *responsibility* is a duty or obligation (e.g., review data records).

Managing data for multiple monitoring protocols and inventories involves timeframes of several years and includes changes in information technology, turnover in staff, new scientific insights and shifting priorities. This requires every individual involved in Network business to understand data stewardship roles and perform responsibilities in the production, analysis, management and/or end use of the data (Table 2.1). In fact, the national I&M Program expects each network to invest at least one third of available resources (time, money, and expertise) in data management, analysis and reporting.

As coordinator of these activities, the fundamental duty of the Network data manager is to understand program and project requirements, create and maintain data management infrastructure and standards, and communicate and collaborate with all responsible individuals. In addition to the overall data management roles and responsibilities presented here, each monitoring protocol and inventory study plan contains specific instructions for assignments and tasks.

Table 2.1. Categories of data stewardship involving all Network personnel and cooperators.

Stewardship Category	Related Activities	Position
<u>Note:</u> Each position is listed in only one category according to primary responsibilities. However, most positions contribute in each category.		
Production	Collecting data from any original or derived source; including recording locations, images, measurements and observations in the field, digitizing source maps, data entry, converting existing data sources, image processing and preparing and delivering informative products (e.g., tables, maps, charts and reports)	Project Crew Leader Project Crew Member Data/GIS Technician
Analysis	Using data to predict and quantify ecosystem elements, structure and function as part of efforts to understand these components, address monitoring objectives and inform park and ecosystem management.	Project Leader Statistician/Biometrician
Management	Preparing and executing policies, procedures and activities to keep data and information resources organized, available, useful, compliant and safe.	Network Data Manager National I&M Data Manager
End Use	Obtaining and applying available information to develop knowledge that contributes to understanding and managing park resources.	Network Coordinator Park Manager Superintendent

Data Stewardship Roles

Increasing demand for more detailed, higher quality data and information about natural resources and ecosystem functions requires a group of people working together to steward data and information assets. The current and expected capacity, diversity and rate of change in information technology make managing any large amount of data a greater task than any person can expect to do alone. Knowledgeable individuals from scientific, administrative and technological disciplines must work in concert to ensure data are collected using appropriate methods and resulting data, reports, maps, models and other derived products are well managed. Stewardship responsibilities apply to all personnel who handle, view, or manage data (Table 2.2). Learning, understanding and acting on the responsibilities for all data stewardship roles can help prevent critical aspects of from being overlooked or under-funded.

Table 2.2. Programmatic roles and associated data management responsibilities.

Role	Data Management Responsibilities
Network Coordinator	Ensure programmatic requirements are met. Responsible for final submission and review of all products and deliverables.
Network Data Manager	Ensure I&M data are organized, useful, compliant, safe and available. Work with IT Specialist to provide hardware, software and network support. Provide GIS support including long-term storage of data, updated software and technical assistance.
Project Leader	Oversee and direct operations, including data management needs. Maintain communication with project staff, data manager and resource specialist(s) regarding data management. Ensure useful data are collected and managed. Work with statistician to analyze data and document procedures. Prepare project objectives, sample design, analysis, documentation, synthesis and reports. Responsible for final submission of all products and deliverables, including documentation.
Project Crew Leader	Supervise crew to ensure data collection, entry, and management obligations are met, including data verification and documentation. Communicate regularly with data manager and project leader.
Project Crew Member / Technician	Collect, record and verify measurements based on project objectives and protocols. Perform data entry and processing. Document methods, procedures and anomalies.
Statistician / Biometrician	Consult on analyses and procedures. Work with Project Leaders to analyze data and complete documentation.
Resource Specialist (Park)	Understand project objectives, data and management relevance. Make decisions about validity, sensitivity and availability of data.
Research Coordinator (Park)	Facilitate data acquisition by external researchers and communicate NPS requirements to permit holders. Notify Network Coordinator of natural resource research requests.
Curator (Park or Region)	Manage collection, documentation, and preservation of specimens.
GIS Manager (Region)	Provide GIS support, updated software and technical assistance.
IT Specialist (Network or Region)	Work with Network data manager to provide IT support for hardware, software and local area network.
I&M Data Manager (National)	Provide Servicewide database availability and support. Provide information management coordination among networks.
End Users (public, managers interpreters and researchers)	Inform the scope and direction of science information needs and activities. Responsible for appropriate use and application of data.

Documentation is Key

Among all information management responsibilities shared by project and Network personnel, the careful documentation of data, source(s) and collection methodology is paramount. This careful and thorough documentation establishes the basis for appropriate use of the data in resulting analysis and products, both in the short and long-term. Network monitoring protocols contain key elements of data documentation and responsibilities (see also Chapter 7 for important guidance on documentation and metadata).

The Hub of Data Stewardship

The Network coordinator, data manager and project leader(s) comprise the central data management team for UCBN I&M projects. Each is responsible for certain aspects of project data and all share responsibility for some overlapping tasks (Figure 2.1). Because of the collaborative nature of project data management, good communication among these personnel is essential to meeting program goals. Communication is promoted by providing:

- Documentation
- Shared working files
- Centralized data management guidance and information on the UCBN website
- User needs assessments and surveys
- Newsletters and/or periodic email updates
- Work groups discussions and meetings
- Presentations
- Training

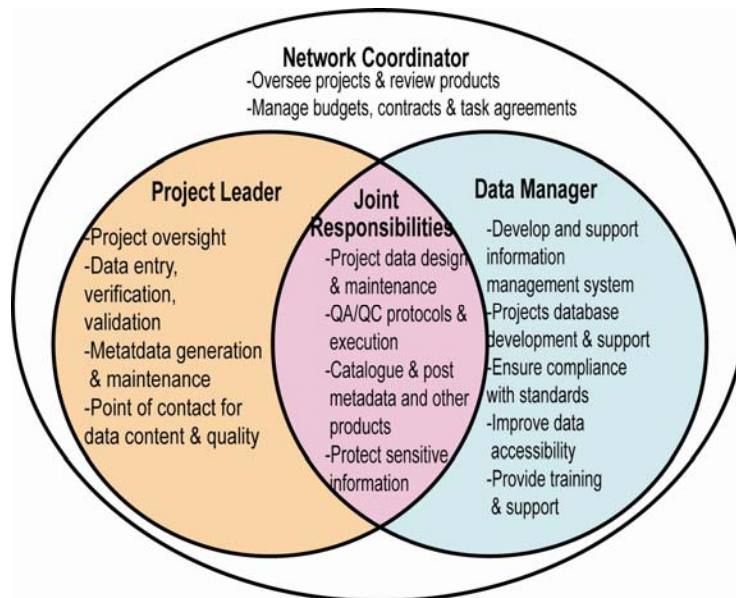


Figure 2.1. Core data management responsibilities for the Network coordinator, data manager and project leader(s).

Network Coordinator

The Network Coordinator supervises project leaders and has the ultimate responsibility for data entry, validation, verification, summarization/analysis and reporting. In the case of sensitive

information, the Network Board of Directors and Science Advisory Committee give final approval for products to be integrated, distributed to the public or parks, or protected.

Project Leader

Project leaders oversee and supervise all phases of an I&M project from initiation to product delivery. Assuming “ownership” of the data, the project leader’s active involvement in data management determines the quality and usefulness of project data. The project leader provides oversight, directs on-the-ground data collection efforts and provides a cohesive link between data collection, analysis, synthesis, interpretation and reporting. Project leaders must act as stewards for all project data and work with Network personnel to ensure proper handling of project data. Project leaders are responsible for designating an alternate leader who is capable of maintaining project operations in his or her absence, to ensure project continuity and data integrity. Specifically, a project leader is responsible for:

- Project documentation that describes the ‘who,’ ‘what,’ ‘where,’ ‘when,’ ‘why,’ and ‘how’ of a project
- Documentation and implementation of SOPs for field data collection and data handling
- Quality assurance and quality control measures, which include the supervision and certification of all field operations, staff training, equipment calibration, species identification, data collection, data entry, verification and validation
- Maintenance of concise explanatory documentation of all deviations from standard procedures
- Detailed post field-collection documentation
- Maintenance of hard copies of data forms and archiving of original data forms
- Scheduling of regular project milestones such as data collection periods, data-processing target dates and reporting deadlines
- Regular summary reports, periodic trend analysis of data, resulting reports and their public availability
- Identifying sensitive information that requires special consideration prior to distribution
- Acting as the main point of contact concerning data content/quality.

The project leader may also work closely with the data manager and/or a data technician to:

- Develop quality assurance and quality control procedures specific to project operations
- Identify training needs for staff related to data management procedures
- Coordinate changes to data entry procedures (field data forms and database interface)
- Document and maintain master data (including metadata generation and maintenance)
- Manage the archival process to ensure regular archival of project documentation, original field data, databases, reports and summaries and other products from the project
- Define the process of how project data will be transformed from raw data into meaningful information
- Create data summary procedures to automate and standardize this transformation process
- Identify and prioritize legacy data for conversion to a useable format
- Increase the interpretability and accessibility of existing natural resource information
- Catalog project data and reports in nationally maintained I&M Program databases such as NPSpecies and NatureBib.

Network Data Manager

The data manager oversees the development, implementation and maintenance of information infrastructure and standards for the Network. The data manager is responsible for ensuring the compatibility of spatial and non-spatial project data with program standards as well as the long-term integrity and availability of that data. Working with project leaders and other program staff, the data manager determines the GIS data and analysis needs for a project, designs databases, applications and products and facilitates dissemination of project data and information products. Creating relationships between spatial and non-spatial data through appropriately designed databases and GIS applications facilitates the analysis of both data types and the sharing of this information with others. The data manager exercises data stewardship to ensure data are archived, documented according to national standards and compatible with other program data. The data manager ensures the maintenance of mechanisms for making data discoverable and available. Both the data manager and the project leader ensure that the information conveyed via these mechanisms is up-to-date and accurate.

General data management duties for the data manager are:

- Overall coordination of data management activities in the Network.
- Develop and maintain storage and dissemination system for all program products.
- Work to improve the acquisition, accessibility and transparency of digital data.
- Ensure the information system is populated and up-to-date with relevant program output.
- Develop and maintain logs recording changes to the data structure and to data certification procedures.
- Act as point-of-contact for access to program products.
- Ensure data security (archiving operations, etc.).

Data managers will also work closely with the project leader to:

- Develop and maintain the infrastructure for metadata creation, project documentation and project data management.
- Create and maintain project databases following best practices and current standards.
- Provide training in data management tailored to the needs of project personnel.
- Establish and implement procedures to protect sensitive data according to project needs.
- Integrate tabular data with geospatial data in a manner that meets project objectives.
- Define the scope of the project data and create a data structure that meets project needs.
- Become familiar with how the data are collected, handled, and used.
- Review quality control aspects of project protocols and SOP documentation.
- Identify elements that can be built into the database structure to facilitate quality control, such as required fields, range limits, pick-lists and conditional validation rules.
- Create a user interface that streamlines the process of data entry, review, validation and summarization and is consistent with the capabilities of project staff.
- Develop automated database procedures to improve efficiency of the data summarization and reporting process.
- Ensure project documentation is complete, complies with metadata requirements and enhances the interpretability and longevity of the project data.
- Ensure regular archival of project materials.
- Inform project staff of changes and advances in data management practices.

Coordination

The Natural Resource Challenge states that collaboration among the NPS, other public agencies, universities and non-governmental organizations is necessary to effectively acquire, apply and promulgate the scientific knowledge gained in national parks. The I&M program encourages coordination among participants at all levels to help ensure data collected in parks are properly developed, maintained, and available for management decision-making, research, and education.

Data management is an adaptive process that facilitates science and fosters partnerships among data managers and scientists within the Network and across the nation. Network data managers work with national I&M program data management staff and regional resource information management personnel to maintain a high-level of involvement in Servicewide and regional databases and data management policy. They regularly coordinate with each other and national program staff via annual meetings, conference calls, workgroups, listservs, web sites, and informal communication. Data managers continue to demonstrate effective cooperation in the sharing of ideas and technology and collaborating on development of data management standards and documentation. This communication and cooperation promotes practical levels of compatibility among protocols and data among networks and organizations.

Data managers also work locally with Network personnel, park staff, and cooperators to promote and develop workable standards and procedures that result in the compatibility and availability of data. Key contacts include park specialists and the project leaders for each monitoring or inventory project. Involvement and input from park scientists and resource management staff is also essential. Consistent and productive communication among these individuals leads to common understanding and better synchronization of data management activities. This communication may take the form of personal visits, phone calls, email, joint meetings and training sessions, and participation in meetings and work activities of the UCBN Science Advisory Committee and Board of Directors.

Chapter 3. Infrastructure and System Architecture

Modern data management infrastructure and system architecture represent the foundation upon which the UCBN data management system is built. Infrastructure refers to the system of computers and servers functionally or directly linked through computer networking services. System architecture refers to the applications, database systems, repositories, and software tools that make up an data management enterprise framework.

The UCBN monitoring program relies heavily on park, regional, and national information technology (IT) personnel and resources to maintain its infrastructure and architecture. Ongoing maintenance of computer resources includes, but is not limited to, hardware replacement, software installation and support, security updates, virus protection, telecommunications networking, and backup processes. As technology relating to computer hardware resources advances, the Network's infrastructure will evolve to meet program objectives. Communication with park and regional IT specialists is essential to ensure adequate resources and service continuity. Rather than focusing on a detailed description of the current computer resources, this chapter describes infrastructure in more general terms and focuses on the systems architecture that is central to I&M data management.

With a small staff and an office located off-site from a NPS park unit, the UCBN receives minimal IT support, necessitating a simple and efficient local infrastructure. The current infrastructure presented here conforms to NPS standards and policies, as will all future infrastructure developments. Further, the UCBN conforms to all national I&M Program standards and mandates in the interest of program integration and information sharing.

Computer Resources Infrastructure

An important element of an data management system is a reliable, secure network of computers and servers. The UCBN digital infrastructure consists of a network-based local area network (LAN) linking UCBN workstations and a Network Data Storage unit, and a regional wide area network (WAN) that includes regional servers as well as access to national servers (Figure 3.1). One UCBN staff member works from a community college location which is outside of the LAN.

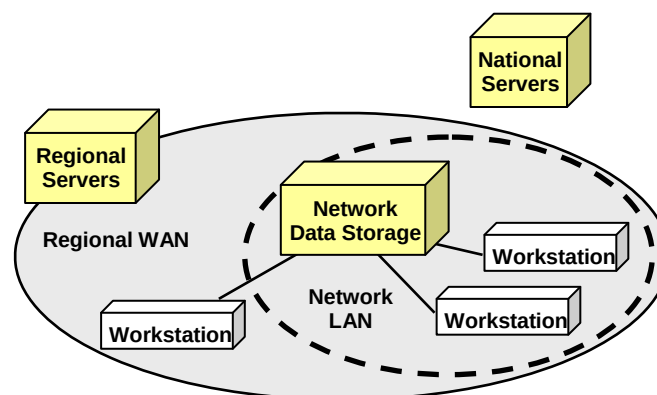


Figure 3.1. Main components of the UCBN infrastructure.

Each of these infrastructure components hosts different parts of the natural resource information system and is maintained and secured by various park, Network, regional and national IT specialists and staff. Network-level IT duties are currently administered by the Network data manager with assistance from the Nez Perce NHP IT specialist. The data manager works with park and regional IT specialists as much as possible to ensure required standards are met and the best data security measures are in place.

National Servers

- Master applications – integrated client-server versions of NatureBib and NPSpecies
- Centralized repositories – NPS Data Store, Protocol Clearinghouse
- Public internet sites – portals to NatureBib, NPSpecies, NPSFocus, websites for monitoring networks

Network Data Storage

- Master project databases – compiled data for monitoring projects and other multi-year efforts that have been certified for data quality
- Common lookup tables – e.g., park name, personnel, species
- Digital library – Network repository for finalized versions of project deliverables (e.g., reports, methods documentation, data files, metadata, etc.), archived final data
- Local applications – desktop versions of national applications such as NPSpecies
- GIS files – base spatial data, imagery and finalized project-specific themes

Network Workstations

- Working files – draft/working databases, geospatial data, reports, administrative records
- Development – project databases and geodatabases in development
- GIS files – personal GIS projects, draft project-specific themes, remote sensing analysis and working data.
- Local applications – desktop versions of national applications such as NPSpecies

National System Architecture

The need for multi-discipline natural resource data management cuts across NPS divisional boundaries and requires national-level strategies for success. Broadly separated into two categories, I&M information needs are:

- *Detailed data needed for on-site resource management and protection.* The information used to guide natural resource management decisions must be specific and useful to management staff at parks and central offices.
- *Summary information needed to describe the resources and their condition.* The information used by NPS and DOI managers and central office personnel to answer requests from Congress and for budget, program and project planning occasionally needs to be aggregated across the National Park Service.

The NPS Natural Resource Program Center (NRPC) and the national I&M Program actively develop and implement a national-level, program-wide data management framework (see the National-level Data Management Applications overview in the appendices). NRPC and I&M staff integrate desktop database applications with internet based databases to serve both local and national-level information requirements. NRPC staff members work with regional and support

office staff to develop extensible desktop GIS systems that integrate closely with the database systems. Centralized data archiving and distribution capabilities at the NRPC provide for long term data security and storage. NRPC sponsors training courses on data management, I&M techniques, and remote sensing to assist I&M data managers with developing and effectively using natural resource data.

At present, the NRPC and national I&M Program are working to streamline the user interface for a variety of their database applications. Their new Integrated Resource Management Application (IRMA) incorporates a Service Oriented Architecture and includes, in its initial phase, the NPSpecies, NatureBib, and NPS Data Store applications. Other on-going developments at the national-level are an internet mapping service (using ArcIMS) to display areas where Vital Signs Monitoring by the I&M Program is occurring, and a centralized data server to host Network datasets within standardized relational database management system formats (such as SQL Server).

Network System Architecture and Centralized Directory Structure

UCBN offices are located in downtown Moscow, Idaho, with the exception of one employee who maintains an office at Central Oregon Community College in Bend, Oregon. A local area network (LAN) for workstations in the UCBN Moscow office is managed through a broadband router. This LAN includes the Network Data Storage unit. Access to the NPS wide area network (WAN), through which all UCBN workstations access regional and national servers, is accomplished by virtual private network (VPN) connection.

Working files are typically maintained on individual workstations, with users ensuring file backups by regularly copying files to an external hard drive and/or to the Network Data Storage unit. Final project files are stored on the Network Data Storage unit, within a standardized file directory structure (see Table 3.1), as well as posted to the NPS Data Store and the UCBN website, as applicable. All files on the Network Data Storage unit are regularly backed up using an automated backup routine and a rotating set of external hard drives, with one current hard drive stored off site. When working with cooperators and contractors, a project directory may be “checked out” from the Network Data Storage unit with the understanding that a similar structure will be delivered as part of the project final products, and that all files will follow UCBN guidance regarding file naming and documentation standards.

Digital versions of all deliverables, as specified in cooperative agreements, permits, study plans, or other administrative records, will be electronically stored on the UCBN Network Data Storage unit by the following general methods:

1. Final products, validated and certified, stored within appropriate subfolders in the Project’s Close folder on the Network Data Storage unit. Final products will include reports, data products, GIS products, data analysis products, and individual file metadata as applicable.
2. Project archive on Network Data Storage unit, with all certified, integral project files “packaged” together (i.e., bundled or zipped together as a stand alone, easily distributed entity). Types of data to be included are listed in Chapter 1, Table 1.1.
3. Final products and updated datasets posted for access by a broad audience (e.g., NPS Data Store), and integrated into national databases (e.g., NPSpecies, NatureBib, NPSTORET)

and national servers (e.g., master Network datasets in relational database management system formats, such as SQL Server).

These files will be centrally stored on the UCBN Network Data Storage unit using a generalized directory structure (Table 3.1). Files will be reviewed for quality and standards prior to storage on the Network Data Storage unit, and only finalized files and certified data will be posted to national databases, online repositories, and servers.

Table 3.1. General directory structure used on the UCBN Network Data Storage unit. *Exact directory structure is subject to change.*

Location (within Network Data Storage unit share folder)	Description	Custodian
UCBN_Root\ \Admin_and_Meetings \Board_of_Directors \Equipment \IM_Meetings_WASO \PacWest_Region \Sci_Adv_Committee \Presentations_Outreach \Workspace_Temp	Storage of files relating to Network administration, presentations and outreach, and temporary workspace file needs.	All UCBN staff
UCBN_Root\ \Data_Management \Applications_WASO \Data_Mng_Plan \Monitoring_Plan \Web	Storage of files relating to data management planning and general applications, Monitoring Plan documents, and the local copy of all UCBN website files and directory folders.	UCBN Data Manager
UCBN_Root\ \GIS_Network_Parks \Admin_GIS \GIS_Data \<individual parks> \Network \GIS_Metadata_Tools \Network_Maps_pdf	Storage of GIS data and metadata files (best accessed through ArcGIS software), GIS administrative files, metadata applications, and general network map products.	UCBN Data Manager
UCBN_PhotoDB\ \ImageDB \Images \Photos_TEMP-Hold_for-ImageDB \ThumbsPlus_Reference-Material	Storage of UCBN Photo Management application backend database, all image files, and reference material relating to photo management software (ThumbsPlus)	UCBN Data Manager
UCBN_NPSonly\ \Admin	Storage of potentially sensitive administrative files	UCBN Coordinator UCBN Data Manager

Table 3.1 (continued). General directory structure used on the UCBN Network Data Storage unit.

Location (within Network Data Storage unit share folder)	Description	Custodian
UCBN_Root\ \Inventories_Projects \VitalSign_Protocols <individual VSign protocol> \1_Initiate \Agreements_Contracts \Meetings_Correspondence \Proposals \2_Plan \Conceptual_Models \Data_Mng_Models \Meetings_Correspondence \Monitoirmg_Objectives \Protocol_Develop_Summary \Protocol_SOP_Drafts \Research_Permits \Study_Plans \3_Implement \Data_Analysis \Data_Management \Data_Dictionary \Database_Working \Download_Files \GIS_Data_Working \Meetings_Correspondence \Photos \Final \Originals \Working \4_Close \Archived_Data_Products \Certified_Data_GIS_Metadata \Final_Reports \Annual_Reports \Investigator_Annual_Reports \Protocol_SOP_Final \Other_Final_Deliverables \Presentations	Storage of files relating to Vital Sign monitoring protocols and to other projects such as inventories. For all Vital Sign protocols and major projects, UCBN will apply a standardized directory structure based on four project phases – initiate, plan, implement, and close.	All UCBN staff
NPS Data Store \GIS_Data \Bibliography_PDFs \Biodiversity_Data	Copies of a project's final products will also reside on the WASO Data Stores	WASO

Chapter 4. Project Work Flow and Data Life Cycle

Natural resource projects follow a general work flow characterized by how data are generated, processed, finalized, and made available. By describing the progressive stages of a project and life cycle of the resulting data, overall objectives and specific steps of the data management process can be more easily communicated. In addition, this awareness helps manage staffing resources needed to produce, maintain, and deliver quality data and information.

From a data management standpoint, there are two types of projects: short-term and long-term. Short-term projects usually last from one to three years and include individual research projects, inventories, or pilot work done in preparation for long-term monitoring. Long-term projects include the Network's vital sign monitoring and other multi-year research and monitoring projects performed by other agencies and cooperators. A primary difference between short- and long-term projects is an increased need to adhere to standards for the latter to ensure internal consistency over time. While the need to follow standards is still present for short-term projects, cost of compliance may outweigh the benefits due to the scope, budget, and level of I&M influence over the project details. Nevertheless, both short-term and long-term projects share many work flow characteristics and both generate data products that must be managed and made available.

Project Work Flow

Projects can be divided into five primary stages: planning and approval, design and testing, implementation, product integration and evaluation, and closure (Figure 4.1, Table 4.1). Each stage is characterized by a set of activities carried out by specific staff. Primary responsibility for these activities rests with different individuals according to the different project phases (see also Chapter 2). Tracking of project status and delivery of products is the responsibility of all UCBN staff, and must be accomplished through regular communication and updates.

Planning and Approval

During planning and approval, preliminary decisions are made regarding project scope and objectives as well as addressing funding sources, permits, and compliance. Primary responsibility rests with the Network Coordinator and project leader(s). Although this phase lacks specific data management activities, it is important that data managers remain informed of projects in this phase. This is especially true as timelines for deliverables are finalized. All contracts, agreements and permits should include standard language that describes the formats, specifications, and timelines for project deliverables.

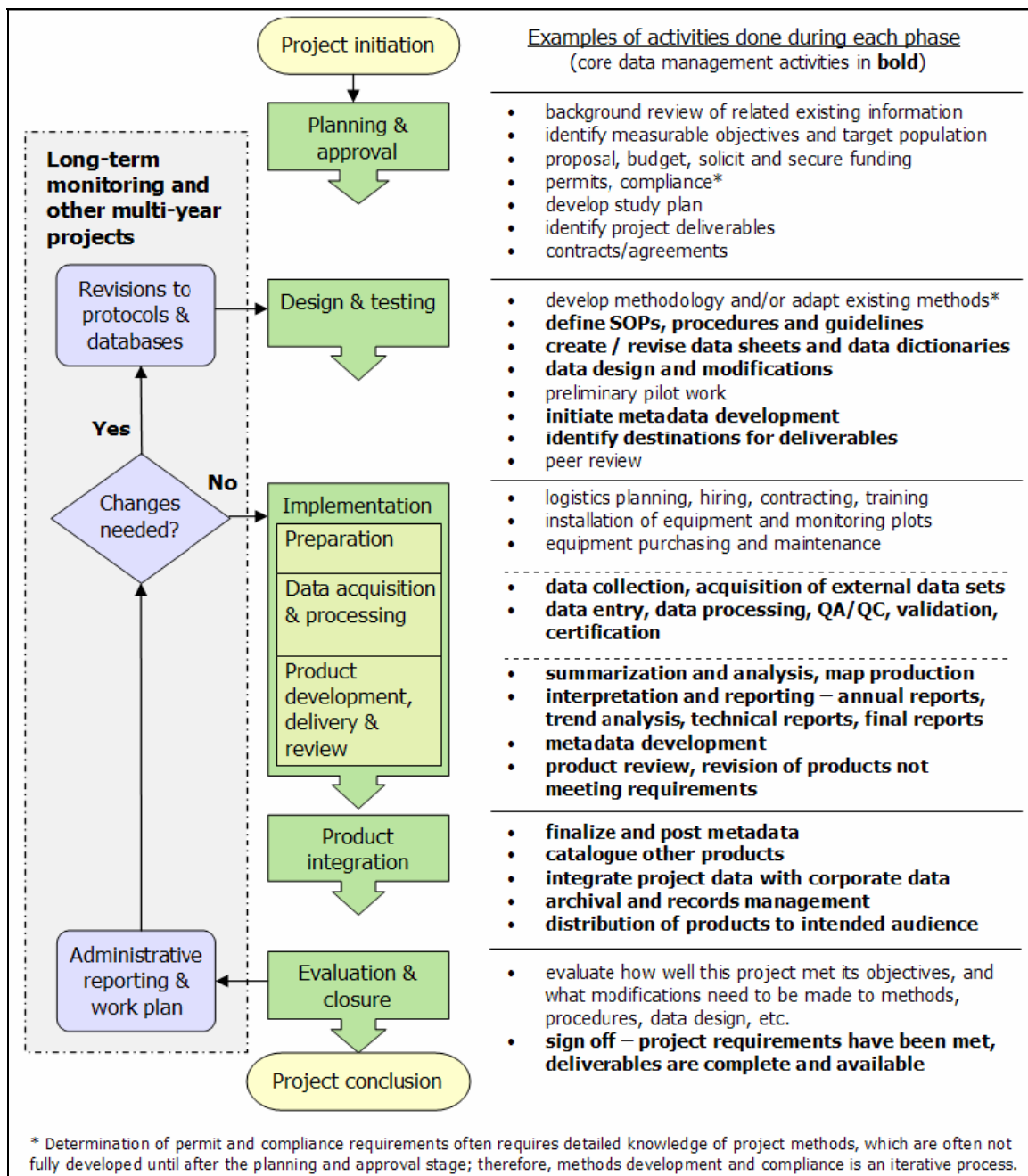


Figure 4.1. Generalized project work flow model for short- and long-term data collection efforts. *Bold face activities indicate core data management elements.*

Design and Testing

During design and testing, a project leader begins formalizing and testing sampling methods and analysis procedures. Details are formalized regarding how data will be acquired, processed, analyzed, reported, and made available to others. The project leader is responsible for developing and testing project methodology, or for modifying existing methods to meet project objectives. It is critical that the project leader and the data manager work together throughout this phase. The dialog between these two will help build and reinforce good data management throughout the

project – especially during the crucial stages of data acquisition, processing, and retrieval. By beginning collaborative development as soon after project approval as possible, data integrity and quality can most easily be assured. An important part of this collaboration is development of the data design and data dictionary, where the specifics of database implementation and collected parameters are defined in detail. Devoting adequate attention to this aspect of the project is possibly the single most important part of assuring the quality, integrity, and usability of the resulting data. Although additional changes are likely once implementation has begun, it is important to minimize these post-hoc changes through careful design work. Once the project methods, data design and data dictionary have been developed and documented, a database can be constructed to meet project requirements.

Implementation

In the implementation phase, data are acquired, processed, error-checked and documented, and products such as reports, maps, and GIS themes, are developed and delivered. Elapsed time between collection, data entry, and analysis is minimized if all crew members are closely working together during this phase.

The project leader oversees all aspects of implementation – from logistics planning, training and equipment procurement to data acquisition, report preparation and final delivery. Throughout this phase, data management staff function primarily as facilitators – providing training and support for database applications, GIS, GPS and other data processing applications; facilitation of data summarization, validation and analysis; and assistance with the technical aspects of documentation and product development. The specific roles of data management staff during this phase will depend primarily on the technical capabilities of the project staff. As much as possible, these roles are worked out in advance of implementation.

Toward the end of this phase, project staff members develop and finalize deliverables identified in the project planning documents. In general, all administrative records, raw and derived data products, metadata, reports and other documentation are delivered to appropriate park and Network staff as specified. Project deliverables should be developed and delivered according to product specifications stipulated in protocols.

Product Integration

Data products and other deliverables are integrated into national and Network databases, metadata records are finalized and posted in clearinghouses and products are distributed or otherwise made available to their intended audience. Product integration includes creating records for project documents in NatureBib, posting metadata records completed and submitted by project leaders, and updating NPSpecies to reflect any new species occurrence information derived from the project. This will allow the project data to be searchable and available to others via Servicewide search engines and will provide for the long-term storage of these products. Certain projects may have additional integration needs, such as working jointly with other agencies for a common database. Ensuring this step of the project is done is a joint responsibility between the project manager and the data manager.

Evaluation and Closure

For long-term monitoring and other cyclic projects, this phase occurs at the end of each field season and leads to an annual review of the project. For short-term projects, this phase represents project completion. After products are catalogued and made available (see Chapter 9), the Network coordinator, project leader, and data manager assess how well the project met its objectives and determine improvements to various aspects of methodology, implementation, and reporting. For monitoring protocols, careful documentation of all changes is required. Changes to protocol narratives, methods and SOPs are maintained in version tables associated with each document. Major revisions may require additional peer review.

Table 4.1. An example of tasks per project stage and staff responsibilities. *PL = Project leader; NC = Network Coordinator; DM = Data Manager. * = Primary role; - = Secondary role.*

Project Stage	Task	Task Description	PL	NC	DM
Planning & Approval	1.1	Notify data manager of new project in planning stage	*	*	
	1.2	Develop study plan and budget	*	*	-
	1.3	Identify project deliverables and determine data sensitivity and ownership	*		-
	1.4	Develop language for contracts/agreements and apply for permits	*	*	-
Design & Testing	2.1	Notify data manager of changing status	*		
	2.2	Clarify roles and responsibilities	*	*	*
	2.3	Scoping for data systems design	*		*
	2.4	Initiate metadata record	*		-
	2.5	Create data dictionary and data forms	*	-	*
	2.6	Define quality assurance measures	*	-	-
	2.7	Database and application development	-	-	*
	2.8	Sample site selection and documentation	*	-	-
	2.9	Provide training as needed and time permits			*
Implementation	3.1	Notify data manager of changing status	*		
	3.2	Training	*		
	3.3	Data acquisition, data entry, upload data logger files	*		-
	3.4	Upload GPS and other data logger data into database	*		-
	3.5	Quality assurance: verification, validation/processing, data certification	*	-	-
	3.6	Update metadata and documentation	*		
	3.7	Deliver data and metadata to DM	*		
Product Development & Delivery	4.1	Summarization, analysis, map production	*	-	-
	4.2	Interpret results, reporting, peer review	*	-	-
	4.3	Assemble products and deliver to DM	*		
	4.4	Check for completeness, proper format, metadata and certification			*
	4.5	Rename files, store in digital library			*
	4.6	Upload certified data into master project database			*
	4.7	Finalize and parse metadata record	-		*
Product Integration	5.1	Publish products and integrate with national applications			*
	5.2	Electronic permanent storage			*
	5.3	Hardcopy archival and distribution	*		-
Evaluation & Administrative Reporting	6.1	Verify that all products are accounted for	*	*	*
	6.2	Administrative evaluation/reporting, decision on continuation/revisions	-	*	-
	6.3	Communicate decisions and data management needs to DM		*	-

Data Life Cycle

During various phases of a project, data take on different forms and are maintained in different places as they are acquired, processed, documented and archived. This data life cycle is characterized by a sequence of events that can be modeled to facilitate communication. These events involve interactions with the following objects:

- *Raw data* – Analog data recorded by hand on hard-copy forms and digital files from handheld computers, GPS receivers, telemetry data loggers, etc.
- *Working database* – A project-specific database for entering and processing data for the current season (or other logical period of time). This might be the only database for short-term projects where there is no need to distinguish working data for the current season from the full set of validated project data.
- *Certified data and metadata* – Completed data and documentation for short-term projects, or one season of completed data for long-term monitoring projects. Certification is a confirmation by the project leader that the data have passed all quality assurance requirements and are complete and ready for distribution. Metadata records include detailed information needed for proper use and interpretation (Chapter 7).
- *Master database* – Project-specific database for storing the full project data, used for viewing, summarizing and analysis. Only used to store data that have passed all quality assurance steps.
- *Reports and data products* – Information that is derived from certified project data.
- *Edit log* – A means of tracking changes to certified data.
- *National databases and repositories* – Applications and repositories maintained at the national level, primarily for the purpose of integration among NPS units and for sharing information with cooperators and the public.
- *Local archives and digital library* – Local storage of data copies, metadata and other products generated by projects. Archives are for hard-copy items and off-line storage media, whereas digital products are stored on the UCBN Network Data Storage unit.

Although the data life cycle may vary depending on specific project needs and objectives, it proceeds as follows (Figure 4.2):

1. *Acquire data* – For data recorded by hand in the field, data forms are reviewed regularly (at least daily) for completeness and validity to capture errors as close to their origin as possible. Electronically recorded data are downloaded daily.
2. *Archive raw data* – Copies of all raw data files are archived intact. Digital files are copied to the digital library section for the project; hard copy forms are either scanned and placed in the digital library or are copied and placed in the archives.
3. *Data entry / import* – Analog data are entered manually and digital data files are uploaded to the working database.
4. *Verification, processing and validation* – Verify accurate transcription of raw data; process data to remove missing values and other data flaws; and validate data using database queries to capture missing data, out-of-range values and logical errors.

5. *Documentation and certification* – Develop or update project metadata and certify the data.

6. *Archive versioned data* – Copies of the certified data and metadata are placed in the digital library. This can be accomplished by storing a compressed copy of the working database or by exporting data to a more software-independent format (e.g., ASCII text; see Chapter 9).

7. *Upload data* – Certified data are uploaded from the working database to the master project database. May not apply to short-term projects where there is no need to distinguish working data for the current season from the full set of certified project data.

8. *Reporting and analysis* – Certified data are used to generate data products, analyses and reports, including semi-automated annual summary reports for monitoring projects. Depending on project needs, data might be exported for analysis or summarized within the database.

9. *Store products* – Reports and other data products are stored in appropriate directories following file naming standards – either in the digital library, on off-line media, or in the document archives (see Chapter 9).

10. *Post products and update national databases* – To make data available to others, reports, data and other products are posted to national repositories and databases (e.g., NPS Data Store, NPSpecies, NPSTORET, NatureBib). Data products containing protected information on rare, threatened or endangered species, or other natural resources of management concern may not be posted (see Chapter 8).

11. *Distribute data and information* – Data, metadata, reports and products can be shared and distributed in a variety of ways – especially via the web-based national databases and repositories, by FTP or mailing in response to specific requests, or by providing direct access to project records. In all cases, distribution will follow legal requirements under the Freedom of Information Act and limitations established to protect information about sensitive resources (see Chapter 8).

12. *Track changes* – All subsequent changes to certified data are documented in an edit log, which accompanies project data and metadata upon distribution. Significant edits will trigger reposting of the data and products to national databases and repositories.

While this sequence of events occurs in an iterative fashion for long-term monitoring projects, it is followed only once for short-term projects. For projects spanning multiple years, decision points include whether or not a separate working database is desirable, and the extent to which product development and delivery is repeated annually.

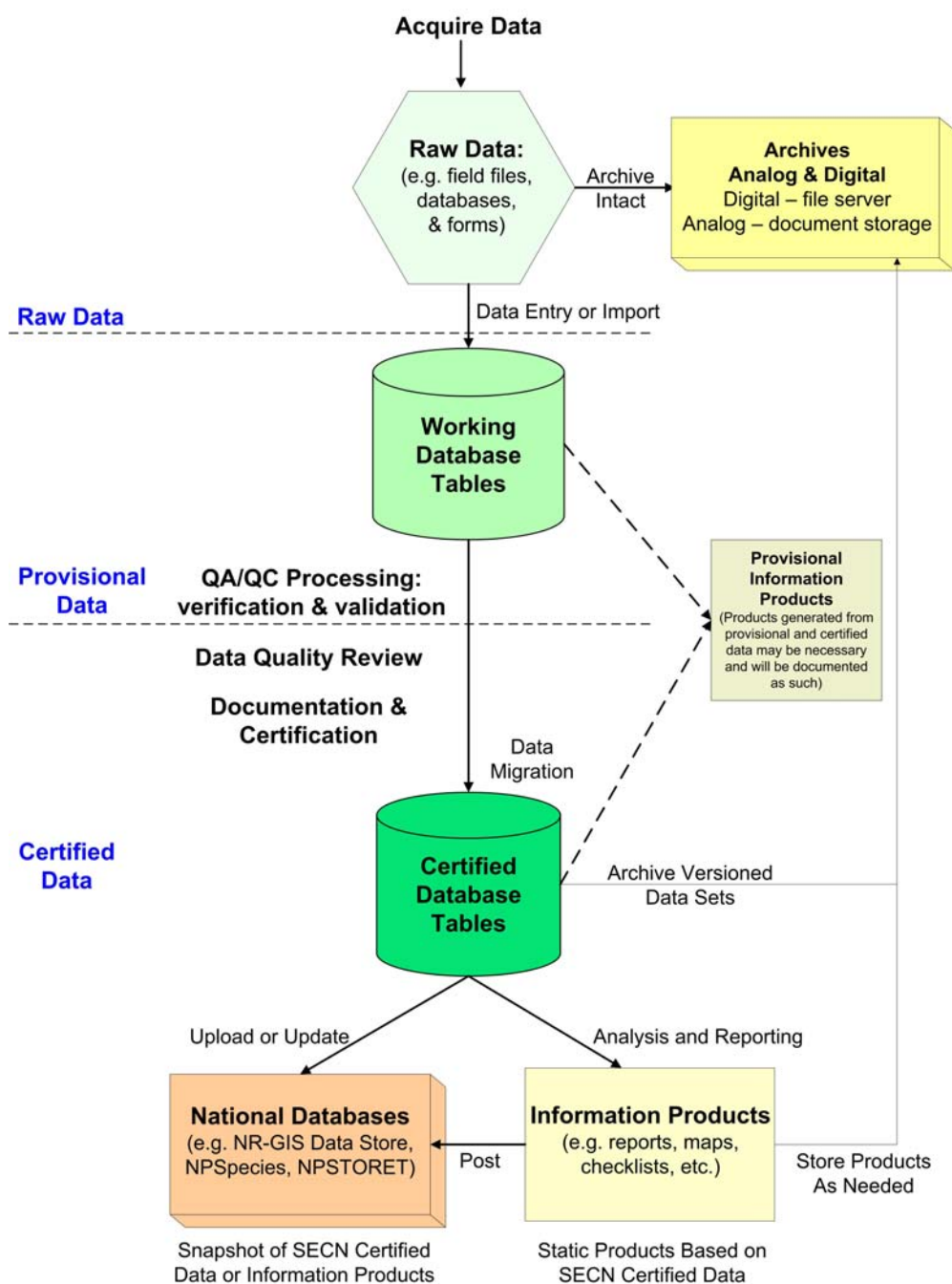


Figure 4.2. Diagram of the typical project data life cycle.

Database Application Development

Rather than developing a single, integrated database system, UCBN strategy relies upon modular, stand-alone project databases that share design standards and links to centralized data tables. Individual project databases are developed, maintained and archived separately. There are numerous advantages to this strategy:

- Data are modular, allowing greater flexibility in accommodating the needs of each project area. Individual project databases and protocols can be developed at different rates

without significant cost to data integration. In addition, one project database can be modified without affecting the functionality of other project databases.

- By working up from modular data, large initial investments in a centralized database and the concomitant difficulties of integrating among project areas with very different – and often unforeseen – structural requirements is avoided.

To support these stand-alone databases, UCBN has created an Integrated Data Dictionary containing master look-up tables for data fields that are shared across multiple protocol- and project-specific databases. This Integrated Data Dictionary includes park codes, watershed and hydrologic unit codes, protocol and project contacts, and a species module.

Protocol- and project-specific UCBN database applications employ a “frontend” or interface through which users interact with data, and a “backend” consisting of the actual data tables contained in a separate MS Access database. It is the backend database file that will be validated and certified each year, and these certified annual datasets will be posted to national repositories, will provide updates to national databases such as NPSpecies and NPSTORET, and will be appended to master database versions. More details on UCBN database development strategies, data modeling, and database naming conventions can be found in supporting guidance documents (see Database_Naming_Conventions and Database_Standards-and-Models).

UCBN will employ two master database strategies. Local master databases will be maintained in MS Access format, which will allow data exploration and analysis through MS Access frontend database applications. In addition, master databases will also be maintained in a SQL Server format, allowing master databases to be maintained and updated on national servers, and allowing for greater flexibility in future development of internet-based data access utilities.

Integrating and Sharing Data Products

Once finalized, project data and data products are secured in long-term storage and made available to others. This requires use of a range of data systems such as development workspaces, product repositories, clearinghouses and web applications, each with a different purpose and function (see Chapters 5 and 9).

Chapter 5. Acquisition, Processing and Reporting

Large, multi-scale natural resources programs increasingly rely on data and information gathered from multiple sources. The UCBN I&M program is no exception. This chapter describes the general steps involved with acquiring, processing and reporting data to meet standards established by the National and Network I&M Programs. Additional specifics can be found in the UCBN Monitoring Plan (Garrett et al. 2007) and individual project protocols.

National Laws and Policies

Current laws and policies encourage programs to make use of existing geospatial data before investing in new acquisitions. Specifically, Executive Order 12906, Section 3 (d) states that each agency must adopt internal procedures to ensure the agency accesses the National Geospatial Data Clearinghouse before it expends Federal funds to collect or produce new geospatial data.

Network Standards and Guidelines

General standards and guidelines for the acquisition and processing of programmatic and non-programmatic data (as defined in Chapter 1) are provided below and illustrated in Figure 5.1. Quality assurance and control procedures, essential for ensuring the production of quality products, are discussed in greater detail in Chapter 6.

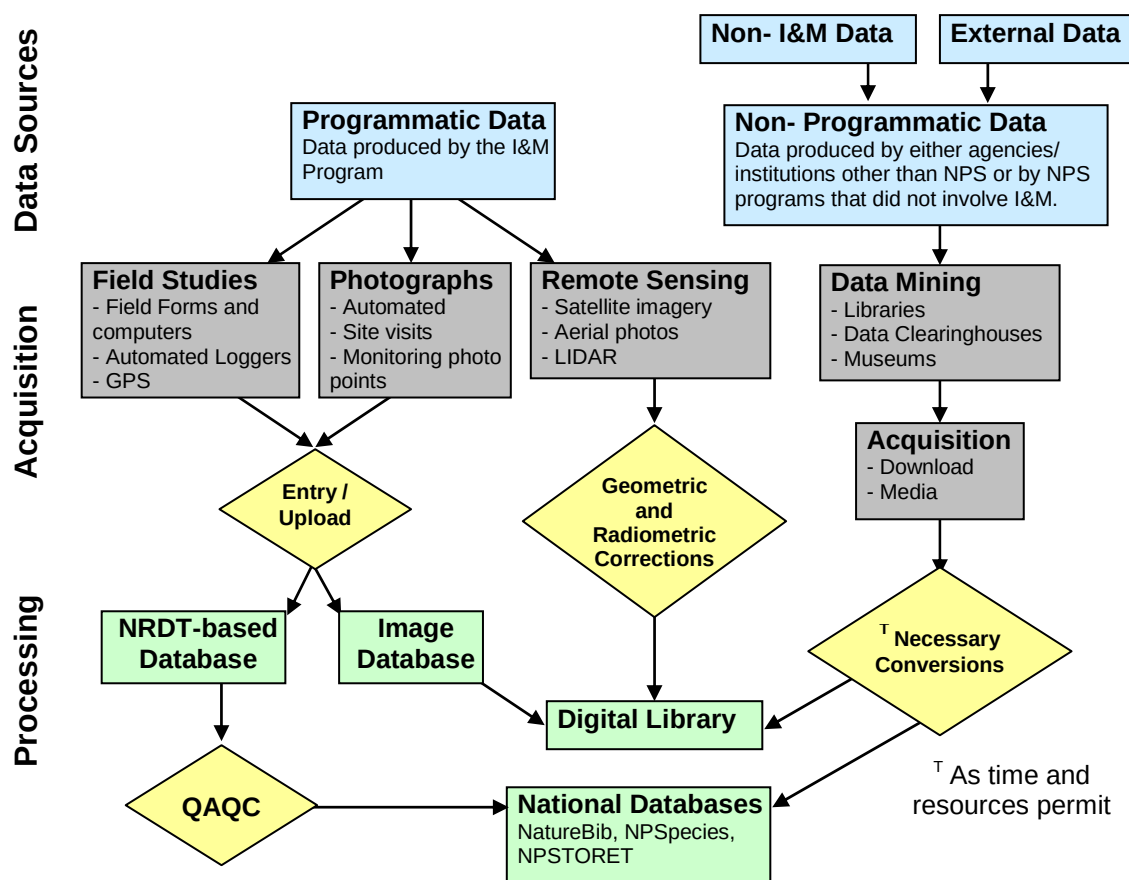


Figure 5.1. General flow of data acquisition and processing.

Programmatic Data

Project leaders and data managers are responsible for ensuring data collection, data entry, verification, validation, storage and archiving are consistent with Network standards. In addition to general SOPs that define Network-wide requirements, protocol-specific SOPs may be developed to detail procedures and/or methodologies.

Data Collection

Data collection initiated by the I&M Program usually takes place as part of either natural resources inventories or vital signs monitoring projects but can also include collaborative efforts between the UCBN and other regional programs, parks or institutions. Data for these projects are typically collected by I&M personnel, park staff or by cooperators/contractors.

Field Studies

The data manager and project leader work closely to develop guidelines and methodologies connected with data collection, storage, and maintenance. This may range from detailing the proper usage of data entry forms or databases to outlining calibration procedures for automated data loggers. Individual protocols will contain project specific SOPs while the Network webpage will contain general SOPs and guidelines.

Listed below are some tools available for field data collection. Refer to the protocol specific SOPs for details on how the following tools may apply to individual projects.

Field Forms – the most common method of recording field data. Inexpensive but more opportunities for errors during the collection/data entry process. Requires neat, legible handwriting and rigorous QA/QC.

Field Computers – increase data collection and data entry efficiency. Data can be downloaded from field computers to office desktops thereby eliminating the data entry step. Fewer chances for error as QA/QC checks can be built into the database. This method for data collection could be inefficient if lengthy notes or comments need to be recorded in the field.

- **Personal Data Assistants (PDAs)** – the small size and relative low cost of these devices make them attractive options for collecting field data. Good for small field projects but not powerful enough for large, data intensive field projects. PDAs can be weatherized fairly easily and inexpensively. Most run either Windows CE or Palm operating systems which may require additional processing/programming to transfer/create the database structure in the field units.
- **Tablet PCs** – same properties as most laptops and provide the user with the convenience of a touch screen interface. They are bulkier, more expensive and harder to weatherize than PDAs, but more powerful as well. Good for data intensive field projects. Because these units run Windows XP (Tablet Edition), project databases can be directly transferred from desktop to field units without additional programming.

Automated Data Loggers – mainly used to collect ambient information such as weather data or water quality information. Must be properly calibrated and maintained so field crews must receive proper training and have access to written SOPs outlining these procedures.

- Permanently deployed devices – often cost prohibitive. Data from these devices must be retrieved and batteries changed on a regular basis defined in the protocol.
- Portable hand-held devices – deployed for sampling only during site visits. Generally less expensive than units that are permanently deployed in the field.

GPS Units – Two types of units are used in the Network to collect or navigate to locations.

- Recreation Grade Units – include both self-contained units (e.g., Garmin Map76CSx) and companion units for PDAs (e.g., Holux). Good for collecting general position information but not recommended for obtaining high accuracy (sub-meter) locations.
- Survey Grade Units – such as some Trimble GPS receivers. Good for collecting highly accurate (sub-meter) location information but often cost prohibitive.

Photographs

Photographs provide an excellent visual record of field visits and are useful for capturing point records of long-term study sites. They also serve well for automated data collection by remotely recording information using web cams or trip cameras. Photographs taken as part of a Network monitoring effort will follow guidelines in the UCBN Photo Database User Guide.

Remotely Sensed Data

Remote sensing technologies can be a powerful tool for characterizing and analyzing landscape data, as well as readily capturing data within areas of low accessibility. Considerations for selecting remote sensing imagery include accuracy and resolution, frequency of measurement, costs, licensing for public use and ortho-rectification standards.

Each remote sensing product is unique. It is imperative users fully understand the product being used and that products are well documented. Any protocols involving remote sensing will involve consultation with a professional remote-sensing specialist and will consider the trade-off between accuracy and costs among different imagery sources (Table 5.1).

The UCBN GIS library, stored on the UCBN Network Data Storage unit, currently includes base cartographic layers, digital orthophoto quadrangles, and satellite imagery for most UCBN parks. As individual monitoring protocols identify additional remote-sensing imagery or GIS needs, the Network will pursue cost-efficient options for acquisition in suitable formats.

Table 5.1. Satellite resolution, swath width, area, cost and history of various remote-sensing imagery sources (sorted by cost).

Sensor	Multi-spectral Resolution (m)	Panchromatic Resolution (m)	Swath (km)	Area (sq km)	Cost/sq km (\$)*	History
Quickbird ^T	2.4	0.7	16	272	29.73	2001-present
Ikonos ^T	4	1.0	11	121	27.03	1999-present
Spot5 ^T	10	2.5 and 5	60	3599	0.77	2002-present
ETM+	30	15	185	34221	0.03	1999-2003
TM	30	NA	185	34221	0.03	1982-1999
ASTER	15	NA	60	3599	0.02	1999-present
MSS (ESTS)	79	NA	185	34221	NA	1972-1997

* cost uses the multi-spectral unit specifications, ^T requires tasking

Data Processing

Programmatic projects will have an associated MS Access database based on the Natural Resource Database Template (NRDT). The database template is highly flexible and can be customized for each project to meet specific needs and requirements. The databases will also incorporate mechanisms such as pick lists and validation rules for quality assurance purposes. The general process is:

- Field crews enter all data into the database provided.
- Field crews periodically forward project data files to the project leader and/or data manager (refer to individual protocols for specific requirements).
- All data undergo QA/QC procedures (see Chapter 6 for more specifics relating to data verification and validation).
- The data manager maintains the master copy of the database and updates it with certified data files received from the project leader.

Two primary national databases (NPSpecies and NatureBib) are used to store, manage and disseminate scientific information on natural resources and the biodiversity of organisms in NPS units. Both databases are available in online (Oracle) and desktop (MS Access) versions and are dynamically linked through common citations. Specific guidelines for use are available online at <http://science.nature.nps.gov/im/apps/index.cfm>. In general:

- *References* – All bibliographic information on park natural resources are entered into NatureBib. All reports which document species presence in a park unit are entered into both NPSpecies and NatureBib databases.
- *Observations* - All species observation data collected are entered into the NPSpecies database.
- *Vouchers* - Information relating to voucher specimens collected during a programmatic field study must be entered into NPSpecies. Once catalog and accession numbers are obtained, voucher information must also be entered into ANCS+. The parks where the vouchers were collected technically own the specimens and have the right to decide where specimens will be stored.

Remote sensing data such as satellite imagery or aerial photography will require varying levels of processing depending on how they are received. These may include geospatial processing or spectral processing. Ideally, all spatial data will be received in a georeferenced format and may

only require geographic transformations to meet UCBN standards (UTM NAD83). Individual monitoring protocols will outline additional spatial and spectral processing necessary to adequately answer specific questions. The general process is:

- Images are geometrically registered using nearest-neighbor resampling methods and co-registered to UTM NAD83. Registration accuracy is assessed.
- Images are radiometrically corrected and converted to exo-atmospheric reflectance.
- Atmospheric corrections are made, if necessary.
- All iterations (raw, intermediate corrections and final) are maintained until project completion when raw and final products are archived.

Non-Programmatic Data

A large percentage of data relevant to managing resources in UCBN parks are collected by entities outside of the Network (parks, universities, other agencies, other NPS programs). The data collected and products produced by such efforts provide a great deal of information about park natural resources and are therefore relevant to the I&M Program mission. These data can be classified as follows:

Current or ongoing – These data are pre-determined for acquisition by the Network and follow very specific acquisition and processing steps (i.e., protocol-dependent). They can be from park-based or external NPS sources and are used either to accompany Network-based monitoring data for data analyses and comparison purposes, or as baseline data obligatory for the completion of a vital sign monitoring protocol (e.g., weather data needed for analyses and comparison purposes).

Legacy - Those data found and compiled through the data mining process. These include vertebrate and vascular plant species data, other important natural resource inventory data, specimen or voucher data, bibliographic data and existing monitoring data.

Data Discovery

Data discovery, or data mining, is the process of locating and identifying useful data. Non-programmatic data can be obtained from academia, private organizations and non-profit groups, as well as local, state, and federal government agencies. Within the NPS, park, regional and national programs are all potential sources of information.

Parks in the UCBN may use base funds or receive funding through NRPP (Natural Resources Protection and Preservation) programs to support park-level projects. Typically, these include:

- *Park-based biological inventories* - Network parks often conduct their own park-based inventory projects, the data from which can be used to supplement Network-level inventories conducted by the I&M Program.
- *Park-based monitoring projects* - Parks also engage in park-level monitoring projects (such as vegetation or water quality) which produce valuable information when developing Network-level monitoring protocols.
- *Park and multi-park based projects* - other studies or projects conducted at the park or regional level that do not fall into one of the previous two categories (e.g., restoration projects).

NPS regional and national programs support all parks within the Pacific West Region and are good resources for information. These include:

- *Air* – National-scale programs collect data, maintain databases, assure data quality and perform the trend analyses relevant to UCBN air quality issues.
- *EPMT* - Exotic Plant Management Teams (EPMT) collect and maintain data regarding the presence of exotic species in many parks and develop and document the methods used to treat these species. Data are stored in the Alien Plant Control and Monitoring Database (APCAM) and maintained by the EPMT Coordinator.
- *Fire Program* – Data on fire occurrence within the UCBN are maintained both at individual park and regional levels. National databases such as FEAT, Fire-Pro, SACS and Fire Program Analysis have been and will be used to maintain information regarding fire incidence and the resources dedicated to fire management.
- *GIS* – The UCBN is supported by regional GIS specialists to help ensure regional GIS data are available and accurate. Much of this data is also available through the Spatial Data Clearinghouse.
- *Wildlife Management* - The regional Threatened and Endangered (T&E) Species Specialist provides management support and coordinates reporting of T&E species.

Additional data sources are listed in Table 5.2. Many of these sources are accessible via the internet, but some require visiting local research or academic institutions, museums or local parks to conduct the searches. Information collected during the data discovery process is maintained either electronically or in hard copy format depending on how it was collected, and is documented as fully as possible (see Chapter 7).

Table 5.2. Summary of possible data sources for different types of information and repository where it should be maintained.

Type of Data	Possible Source	Repository
Bibliographic / Literature	• National NPS Databases (e.g., NatureBib) • Online literature databases (e.g., First Search) • Library catalogs (e.g., academic or research institutions)	• NatureBib • Reference cabinets for hard copy materials • Digital archive for electronic materials
Geographic Data	• Park archives through ANCS+ • Regional centralized GIS data • Federal and State Geographic Data clearinghouses • Local, state and federal government offices • Regional and Park GIS specialists	• NPS Data Store • Digital archive
Biologic / Natural Resource Data	• NPSpecies • Voucher collections (museums, parks, universities) • Network parks	• NPSpecies • Digital archive

Data discovery is an integral part of project development, but data discovery efforts should not be limited solely to project development needs. This should be an ongoing process requiring regular data searches and visits to Network parks to ensure the Network maintains as much relevant material pertaining to the parks as possible. Encouraging data sharing among Network parks will assist in this process and may alleviate the need for regular searches of park records.

Data Collection

Protocols for vital signs monitored through data harvesting of outside sources will describe in detail how data are to be acquired and processed. Many of these SOPs are still in development, but as the Network monitoring plan moves from draft to final version, outside data sources will be identified and guidelines for acquiring these data will be put in place.

The process of harvesting data will be standardized as much as possible. Considerations for harvesting are as follows:

- Does the data source organization know of the I&M Program and its data needs?
- Does a Memorandum of Understanding or Cooperative Agreement need to be in place?
- What is the contingency if this data source is no longer available?
- Can downloads or requests for data be consistently exported/imported and scheduled?
- How will the downloaded data be stored and integrated into the I&M Program?
- How are errors addressed?
- Is the data source organization interested in integrating I&M Program data, where appropriate?
- Is the documentation adequate?

Harvested data may not exactly match UCBN data needs or standards. For instance, boundaries may not coincide with park boundaries or specific parameters may not be collected. Project leaders are encouraged to work with the data source to incorporate Network requirements as much as possible.

Agencies or organizations that compile these data often have the expertise to conduct proper quality control procedures and the capability to function as a repository and clearinghouse for validated data. When not kept in-house, data can be either downloaded from online databases or requested for on CD or other media. In some cases, portions of external databases may be incorporated into UCBN databases and thereby made more accessible to staff.

Data Processing

Much of the information identified during the data discovery process is likely to be *legacy data*, or data collected prior to the inception of the UCBN I&M Program. As time and resources permit, legacy information should be converted to file formats compatible with current software standards. Similarly, the Network intends to scan hardcopy references and materials, save them as .pdf files and create a digital library as time permits.

Although the I&M Program has begun to share many of its file and database standards with other NPS programs and outside agencies, much of the data collected from external sources must undergo some degree of processing to meet Network standards. Basic processing steps include:

- Enter all biodiversity data into NPSSpecies (especially important for park-based biological inventories) and enter all associated references into NatureBib.
- Enter all natural resource reports and publications related to UCBN parks into NatureBib. Hard copies should be stored in the appropriate park collections and electronic copies archived in the proper directory on the UCBN Network Data Storage unit.

- Ensure all GIS data are in the proper projection and accompanied by FGDC-compliant metadata. This is especially applicable to data collected in support of the land use/land cover change vital sign.

Certain data will require more than the basic processing steps described above. The level of data processing necessary will depend on the status of the existing data and on the desired output. In such cases, specific protocols will provide necessary data processing requirements.

It is important that park, regional and Network staff work closely together to ensure information is maintained in a manner that promotes data sharing. Accordingly, the data manager will:

- Work closely with park and regional staff to ensure high quality data are available
- Provide training to staff interested in learning NPSpecies and/or NatureBib
- Offer to enter information collected or maintained by the parks into NPSpecies or NatureBib
- Provide assistance with processing voucher specimens collected in Network parks
- Develop databases based on the NRDT that meet the needs of park resource managers

Analysis and Reporting

Providing meaningful results from data summary and analysis is a cornerstone of the I&M Program and characterizes the program's data management mission to provide useful information for managers and scientists. The UCBN Monitoring Plan (Garrett et al. 2007, Chapter 7) contains the background and overall approach to data analysis and reporting for the Network. The associated data management must provide valid data in formats that support scheduled and ad hoc display, query, analysis, summary, and reporting. Routine and scheduled data summary, analysis, and reporting requirements and procedures should be identified in project protocols.

The development of products will be guided by project objectives, protocols, and data management SOPs. Products defined in advance, such as routine summaries and output formats can be automated through the use of queries and reports stored in a project database. Queries can be used to view raw or summarized data and output either type as custom-formatted reports or files that can be imported by other analytical tools (e.g., statistical software). Other queries can be built to facilitate data exploration and unscheduled analysis. Automation of data summaries and annual reports facilitates the ability to manage multiple projects. However, specific needs should be determined by individual projects and project leaders need to work with data managers to create and test these queries and reports.

Most likely, data will be exported from project databases for most statistical analyses beyond means, standard deviations and other descriptive statistics. UCBN will use third party statistical software (e.g., R, SAS) for frequency distribution plots, tests for normality, analysis of variance and/or other applicable statistical tests. Other external software (e.g., DISTANCE, PRESENCE) may also require specialized data exports.

Changes to Procedures

Changes to established data collection procedures are discouraged unless there are acceptable, valid reasons for altering the methodologies. Ideally, all problems should be identified during the

design and testing stages of the project and changes implemented prior to the collection of any field data. Protocols should attempt to identify any foreseeable issues that might occur as well as contingencies to address them. Inevitably, unforeseen problems may occur which require procedure or protocol revision after data collection has begun. Significant changes to protocols must be approved by the Network Coordinator. Proposed changes should be discussed with the project leader and data manager. The key official must evaluate the proposed changes and determine if additional peer review is required before accepting them. All changes must be carefully documented within the SOPs and any associated databases.

Altering data collection, analysis, or reporting procedures may also occur as a result of the comprehensive reviews that all monitoring protocols undergo. During the review, data are evaluated to determine if the current protocol accomplished its goal. If it is concluded that the protocol in its present form has not achieved the desired results, changes could be recommended.

Chapter 6. Quality Assurance and Quality Control (QA/QC)

The success of I&M networks is dependent on the quality of the data collected, managed and disseminated. Analyses performed to detect ecological trends or patterns require data that are recorded properly and have acceptable precision and minimal bias. Otherwise, detection of subtle changes in ecosystem patterns and processes could be limited, incorrect interpretations and conclusions drawn, and credibility of the I&M Program greatly compromised.

Quality assurance (QA) can be defined as an integrated system of management activities involving planning, implementation, documentation, assessment, reporting and quality improvement to ensure a process, item, or service is of the type and quality needed and expected by the consumer (Palmer 2003). Quality control (QC) is a system of technical activities to measure the attributes and performance of a process, item, or service relative to defined standards (Palmer 2003). While QA procedures maintain quality throughout all stages of data development, QC procedures monitor or evaluate the resulting data products.

To ensure the UCBN produces and maintains data of the highest possible quality, procedures have been established to identify and minimize errors at each project stage associated with the data life cycle (Figure 6.1). Specific QA/QC procedures will depend upon the individual vital signs being monitored and are specified in the protocols and SOPs. Some general concepts, however, apply to all Network projects. The general guidelines presented in this plan were primarily adapted from the Draft Data Management Protocol (Tessler and Gregson 1997) and ideas contained in Michener and Brunt (2000).

NPS Mandate for Quality

Director's Order #11B, "Ensuring Quality of Information Disseminated by the National Park Service," was issued to promote data quality (National Park Service 2002). It defines 'quality' as incorporating three key components: objectivity, utility, and integrity.

Objectivity consists of: 1) presentation: whether disseminated information is being presented in an accurate, clear, complete and unbiased manner within a proper context; and 2) substance: ensuring accurate, usable and reliable information.

Utility refers to the usefulness of the information to its intended users.

Integrity refers to the security of information; e.g., protection from unauthorized access or revision to ensure information is not compromised through corruption or falsification.

Director's Order #11B also specifies that information must be based on reliable data sources that are accurate, timely and representative of the most current information available. These standards apply not only to NPS-generated information, but also to information provided by other parties to the NPS if the NPS disseminates or relies upon this information.

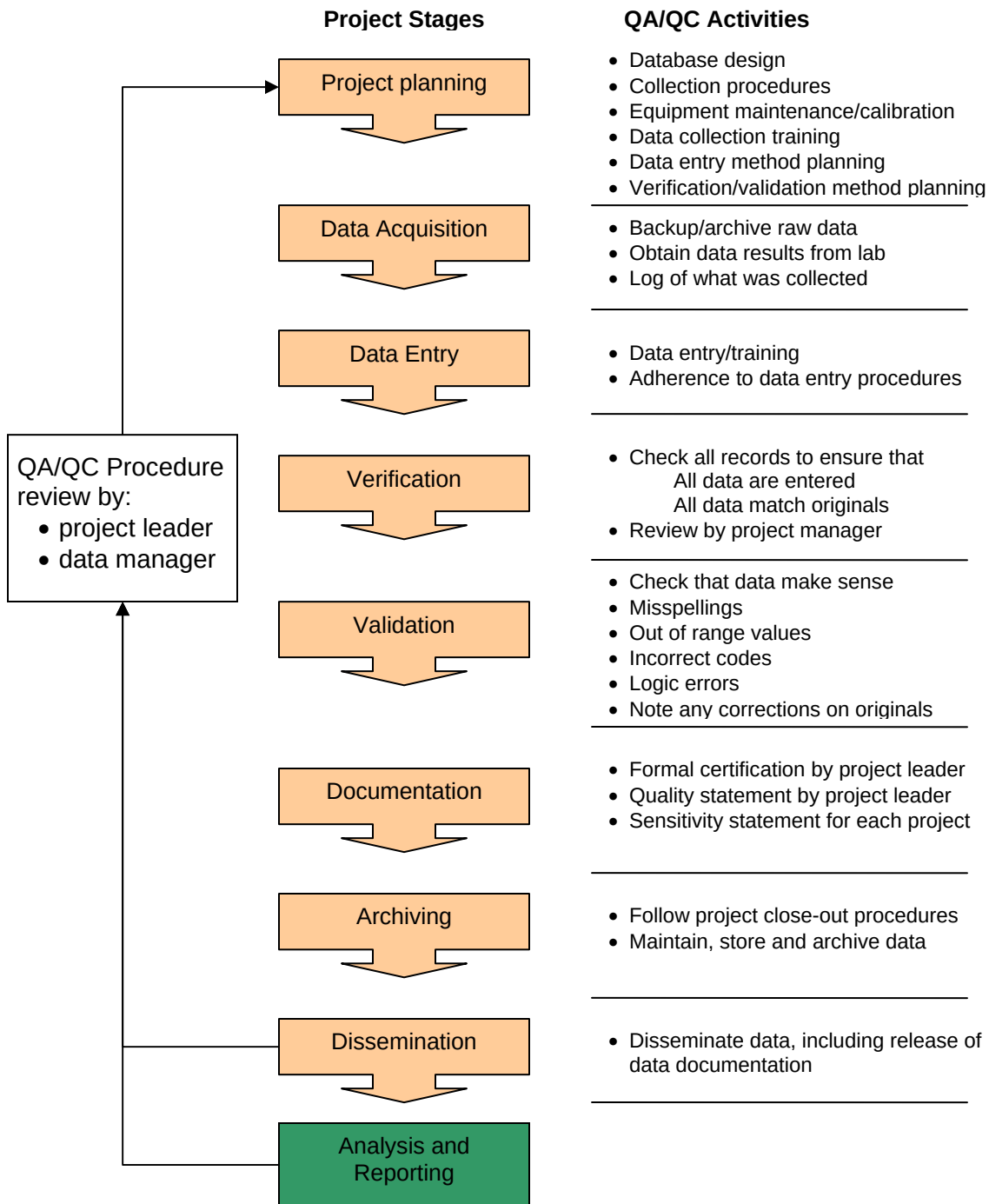


Figure 6.1. General flow of data and associated Quality Assurance Quality Control procedures. Quality control with regards to data analysis is specific to each project and addressed in appropriate standard operating procedures.

Goals and Expectations

Although data containing no errors would be ideal, the cost of attaining 95%-100% accuracy may outweigh the benefit. Therefore, at least two factors are considered when setting data quality expectations:

- Frequency of incorrect data fields or records
- Significance of error within a data field.

Errors are more likely to be detected when data are clearly documented and what constitutes a ‘significant’ error within *that* data is understood. The significance of an error can vary both among data sets and within a single data set. For example, a two-digit number with a misplaced decimal point (e.g., 99 vs. 9.9) may be a significant error while a six-digit number with an incorrect decimal value (e.g., 9999.99 vs. 9999.98), may not. However, one incorrect digit in a six-digit species number could indicate a different species.

QA/QC mechanisms are designed to prevent data contamination, which occurs when a process or event introduces either of two fundamental types of errors into data:

- Errors of commission include those caused by data entry or transcription errors, or malfunctioning equipment. They are common, fairly easy to identify and can be effectively reduced up front with appropriate QA mechanisms built into the data acquisition process, as well as QC procedures applied after the data have been acquired.
- Errors of omission often include insufficient documentation of legitimate data values, which could affect the interpretation of those values. These errors may be harder to detect and correct, but many should be revealed by rigorous QC procedures.

Selected QA/QC procedures relative to the amount of planning and quality control necessary to have confidence in the data are illustrated in Figure 6.2. The most effective mechanism for ensuring a project produces high-quality data is to establish and adhere to procedures that direct project staff through accurate data collection, entry and validation.

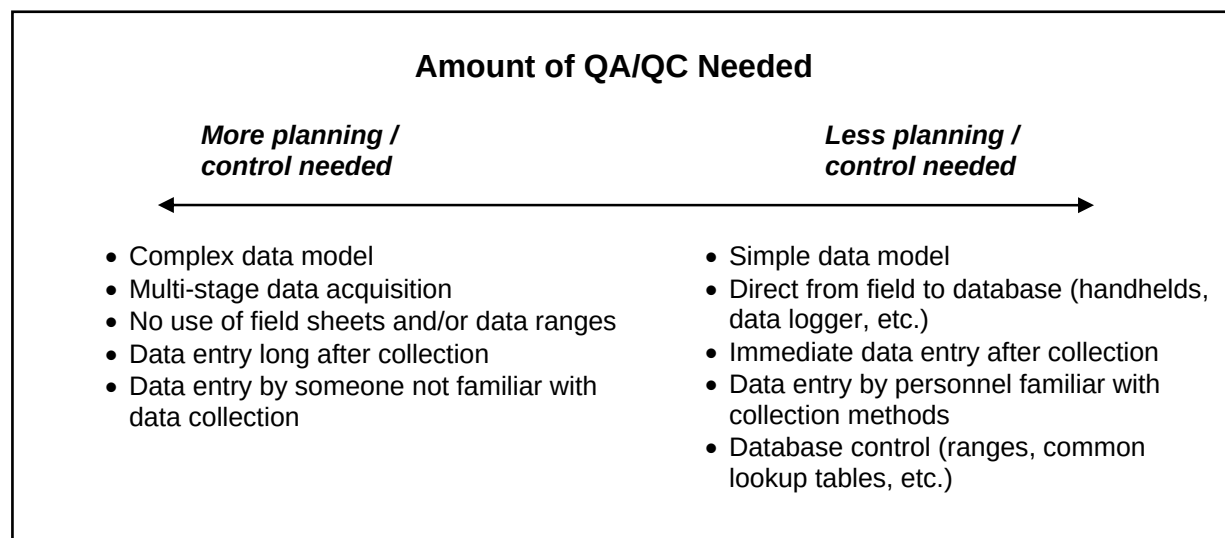


Figure 6.2. Common data management elements influencing the amount of QA/QC needed.

Roles and Responsibilities

Producing and maintaining high quality information is the responsibility of everyone involved with handling project data. It is essential that each member of the team have a stake in data quality and is responsible for the quality of the results generated from his or her tasks. While Chapter 2 specifies data management roles and responsibilities, selected QA/QC duties are emphasized here.

Project leaders need to:

- Be aware of quality protocols and convey their importance to technicians and field crews
- Ensure compliance with the protocols
- Plan for and ensure proper execution of data verification and validation
- Review and certify all final reports and data products.

Project technicians must:

- Follow established protocols for data collection, data entry, and verification
- Inform the project leader or data manager of quality-related problems or difficulties.

The data manager is responsible for:

- Developing Network-wide SOPs to ensure data quality
- Making project leaders, technicians and others involved aware of the established procedures and enforcing adherence to them
- Evaluating the quality of all data against NPS standards before dissemination
- Performing periodic data audits and QC checks to monitor and improve operations.

General Operations

Version Control and File Naming Standards

Version control is the process of managing copies of changing files over the course of a project and requires adherence to file naming standards. Because of the scope and quantity of files being consolidated into one place, it is critical that computer files be given unique, identifying names.

The following conventions apply to all UCBN file names:

- No spaces or special characters within the name (hyphens between words are acceptable)
- Use underscores (“_”) as delimiters
- Include dates for version control in yyyyymmdd format (e.g., 20060211 for Feb. 11, 2006).

File names for final products will begin with “UCBN” or the park code and contain a brief but clear explanation of the file content followed by an eight digit date. Files in development may also require an indication of the QA/QC status (i.e., draft, raw, final, etc.). Each of these file name elements is preceded by an underscore. For example, a draft version of the UCBN DMP dated Oct. 23, 2006, would be indicated as “UCBN_Data-Mng-Plan_draft_20061023.doc”.

Version control practices allow for full recovery of a file as it existed prior to any change, either alteration to the structure or content of the files, being made. Before making any major changes to a file, a copy with the appropriate file name is saved. As indicated above, files are stored with eight-digit dates which serve as version control. In the case of multiple versions on the same day,

a version number (e.g., “v1”, “v2”, etc.) can be appended to the file name. This allows changes to be tracked over time and facilitates collaboration between multiple personnel working on common files. With proper controls and communications, versioning ensures only the most current information is used.

These general version control guidelines apply to files maintained on the UCBN Network Data Storage unit. They do not apply to enterprise-level server applications such as the I&M Program’s Servicewide Oracle databases; updates to these master databases are accomplished either through online data entry/edit or through file submission to national staff.

Project Planning and Data Design

The methods a given project chooses to use and the array of information to gather are integral to data quality. The UCBN holds as policy that techniques and procedures chosen for a project will maximize data quality. Quality assurance will be achieved by stipulating that:

- Common lookup tables are created for parameter values recorded by multiple projects (e.g., weather variables, standard equipment, field personnel, etc).
- Each protocol has SOPs that address core data management practices (e.g., use and maintenance of equipment, training, data backup, entry, verification, validation, etc.).
- Each project uses field computers (PDA or tablet) for data collection if possible. If not, at a minimum, standardized field sheets are used.
- Database data entry forms, if necessary, resemble the field sheets.
- Automated error checking features are included in database applications.
- Database application design maximizes the use of auto-fill, auto-correct, range limits, pick lists and other constraints specific to projects.
- Database applications include a means to track the date a record is created or modified, the person creating/modifying the record and errors reported after dissemination.
- Database maintenance logs are maintained for each UCBN database and housed in association with database files.

Data Collection

Chapter 5 addressed general data acquisition and initial handling as well as changes to data collection protocols. Attention to detail during the data collection phase, however, is crucial to overall data quality. Unlike a typographical error that occurs during data entry, an incorrect observation in the field is not easily corrected.

The UCBN adopts the following guidelines regarding data collection that affect data quality:

- Field crews will receive proper training.
- Any project using field equipment will include a calibration and maintenance SOP that will also specify establishment of an equipment maintenance log.
- Wherever possible and appropriate, data loggers or field-based computers will be used to collect data. When this is not possible, data will be recorded on formatted, project-specific data sheets that reflect the overall design of the project and are designed to minimize the amount of writing necessary to effectively record observations.
- When field sheets are necessary, the format will reflect the database entry interface to help ensure all information is recorded and subsequent data entry errors are minimized.

- Completed field forms will be proofed for errors each day in the field. Any edits will be done by drawing a single line through the information to be changed and adding any replacement information in clear writing next to the original entry.
- Separate SOPs will describe the use of data loggers or field-based computers and will include direction for daily review and back-up.

Data Entry

Data entry is the process whereby raw data are transferred from paper field forms into an electronic data format. When data are gathered or stored digitally in the field (e.g., on a data logger), data entry consists of downloading to an office computer where data can be further manipulated. The goal of data entry is to transcribe field observations into a computer database with 100% accuracy, although errors are unavoidable. Subsequent data verification is conducted to ensure raw data matches entered data. Following verification, data validation may result in changes *to the entered data*. Data entry is a separate operation from data validation and care must be taken to not impose validation (beyond that automatically imposed by programming rules in a database) during data entry.

The UCBN adopts the following guidelines regarding data entry:

- To the extent possible, data entry will be automated. This may simply entail downloading data from field-based computers but may include the application of new technology to allow for machine-driven data entry (e.g., voice-recognition software).
- Data will be entered as soon as reasonably possible after collection.
- Data entry will be completed by someone familiar with data collection. The project leader (with assistance from the data manager if needed) must ensure that data entry staff is familiar with the database software, database structure and any standard codes used by the Network. At a minimum, data entry technicians should know how to open a data entry form, create a new record and exit the database properly. They must also learn how to correct mistakes made while typing.
- If feasible, data will be entered by two qualified persons; one person will read the observations and the other will enter the data.
- Data will be entered into pre-designed database forms that resemble field sheets and maximize error control.

Data Verification

Data quality is appraised by applying verification and validation procedures as part of the quality control process. These procedures are more successful when preceded by effective quality assurance practices (i.e., planning). *Data verification* checks that the digitized data match the source data, while *data validation* checks that the data make sense. Although data entry and data verification can be handled by personnel who are less familiar with the data, validation requires in-depth knowledge about the data.

The UCBN adopts the following guidelines regarding data verification:

- Project leaders are responsible for specifying in the project protocol one or more of the data verification methods available and ensuring proper execution. At the discretion of the project leader, additional verification methods may be applied.
- Data verification is carried out by staff thoroughly familiar with data collection and entry.

- All records (100%) will be verified against original source data.
- A subset of randomly selected records (10%) will be reviewed after initial verification by the project leader. If errors are found, the entire data set will be verified again.
- A record of the verification process for data, including number of iterations and results, will be prepared by the project leader as part of formal metadata generation (Chapter 7).
- Spatial data collected as part of the project will be viewed in a GIS and visually inspected for accuracy (e.g., points outside park boundaries, upland locations occurring in water).

Each of the following methods has a direct correlation between effectiveness and effort. The methods that eliminate the most errors can be very time consuming while the simplest and cheapest methods will not be as efficient at detecting errors.

- *Visual review at data entry.* The data entry technician verifies each record after input and immediately corrects any errors. This method is the least complicated since it requires no additional personnel or software. Its reliability depends entirely upon the person keying data and thus, is probably the least reliable data verification method.
- *Visual review after data entry.* Upon completion of data entry, all records are printed and compared with the original values from the hard copy. Errors are clearly marked and corrected in the database as soon after data entry as possible. Reliability increases if someone other than the person keying the data performs the review. Alternatively, two technicians (one reading from the original data and one checking the entered data) can perform this review.
- *Duplicate data entry.* The data entry person completes all data entry as normal. Random records are selected (every n th record) and entered into an empty replica of the permanent database, preferably by someone other than the person keying the permanent data. A database query is then used to automatically compare the duplicate records from the two data sets and report on any mismatches. Disparities are manually reviewed and corrected if necessary. This method adds the overhead of retyping the selected records, as well as the creation of a comparison query. However, it becomes increasingly successful as the value of n decreases. Professional data entry services frequently use this method.

Data verification can be improved by calculating summary statistics and identifying duplicate or omitted records. For example, the number of known constant elements, such as the number of sampling sites, plots per site, or dates per sample can be evaluated. The more checks devised to test the completeness of the data, the greater confidence level in the verified data.

Data Validation

Validation is the process of reviewing computerized data for range and logic errors and may accompany data verification *only* if the operator has comprehensive knowledge of the data and subject. More often, validation is a separate operation carried out *after* verification by a project specialist who can identify generic and specific errors in particular data types. It is essential that all data are validated as truthful and do not misrepresent the circumstances and limitations of collection. Failure to follow SOPs for data entry, validation, and verification will render data suspect.

General step-by-step instructions are not possible for data validation because each data set has unique measurement ranges, sampling precision and accuracy. Nevertheless, validation is a critically important step in data certification and a required component of any UCBN protocol. Invalid data commonly consist of misspelled species names or site codes, wrong dates, or out-of-range errors in parameters with well defined limits (e.g., pH). More interesting and often puzzling errors are detected as unreasonable metrics (e.g., stream temperature of 70°C) or impossible associations (e.g., a tree 2 feet in diameter and only 3 feet high). These types of erroneous data are called *logic errors* because they produce illogical (and incorrect) results. The discovery of logic errors has direct, positive consequences for data quality and provides important feedback to the methods and data forms used in the field. Histograms, line plots and basic statistics can reveal possible logic and range errors.

The UCBN adopts the following guidelines regarding data validation:

- Each project protocol will address a process for data validation that includes at least one of the available methods.
- Corrections or deletions as a result of data validation require notations in the original paper field records about how and why the data were changed, with the editor's initials.
- Modifications of the field data will be clear and concise while preserving the original data entries or notes (i.e., no erasing).
- Validation efforts will also include a check for the completeness of data since field sheets or other sources could easily be overlooked.
- Use of automated routines and/or data summary and visualization (e.g., histograms, line plots and basic statistics) will be maximized to identify possible logic and range errors.
- Use of database programming will be maximized to control data entry. This will be achieved via the use of lookup tables and/or field-type design (e.g., yes/no field-types).

Specific procedures for data validation depend on the vital sign being monitored and are included in the project protocols. However, the following general methods can be used as guidelines:

- *Data entry application programming.* Certain components of data validation are built into data entry forms. This method is essentially part of the database design and is discussed earlier in this chapter. Not all fields, however, have appropriate ranges known in advance. Caution must be exercised when using lookup tables to constrain variable values. Values occurring outside the range set by a lookup table (established during database design) may not always be invalid. As part of data validation procedures, the project leader is responsible for correct use of lookup tables or other automated value range control.
- *Outlier Detection.* According to Edwards (2000), “the term outlier is not (and should not be) formally defined. An outlier is simply an unusually extreme value for a variable, given the statistical model in use.” Any data will undoubtedly contain some extreme values, so the meaning of ‘unusually extreme’ is subjective. The challenge in detecting outliers is in deciding how unusual a value must be before it can (with confidence) be considered ‘unusually’ extreme.

Data QA procedures should not try to *eliminate* outliers. Extreme values naturally occur in many ecological phenomena; eliminating these values simply because they are extreme is equivalent to pretending the phenomenon is ‘well-behaved’ when it is not. Eliminating

data contamination is perhaps a better way to explain this quality assurance goal. When an outlier is detected (via GIS, database, graphic and statistical tools for ad-hoc queries and displays), the possibility of contamination will be evaluated and noted.

- *Other exploratory data analyses.* Palmer and Landis (2002) suggest calculations for assessments of precision, bias, representativeness, completeness and comparability may be applicable and, for certain types of measurements, evaluation of detection limits may also be warranted. Normal probability plots, Grubb's test and simple and multiple linear regression techniques may also be used (Edwards 2000).

Review, Conformance, and Communication

The NPS requires QA/QC review and approval prior to communicating or disseminating data and information. Documentation of the QA/QC standards used in producing the information and that substantiate the information quality must be formally certified and distributed with the related data and information. Mechanisms must also be in place for receiving and addressing comments or complaints pertaining to data quality (see also Chapter 8).

As part of the close-out and evaluation stage of each UCBN project, QA/QC procedures will be reviewed by the project leader and recommendations for change will be included in the annual report. Similarly, the UCBN data manager will review and revise the QA/QC procedures included in the DMP and/or SOPs as needed.

The data manager will conduct periodic data audits of random UCBN monitoring projects to ensure compliance with the DMP and protocol QA/QC procedures. Correction of any deficiencies found during this process will be facilitated and tracked by the data manager. These quality checks promote a cyclic process of continuous feedback and improvement of both the data and quality planning process. Audits may include verification of the following:

- Data collection and reporting requirements are being met
- Data collection and reporting procedures are being followed
- Verification and validation procedures are being followed
- Data file structures and maintenance are clear, accurate and according to plan
- Revision control of program documents and field sheets are adequate
- Calibration and maintenance procedures are being followed
- Seasonal and temporary staff have been trained in data management practice
- Metadata collection and construction for the program proceeds in a timely manner
- Data are being archived and catalogued appropriately for long term storage

The final step in the QA/QC process is preparation of summary documentation that assesses overall data quality. The statement of data quality is composed by the project leader and incorporated into formal metadata. Metadata will also provide information on the specific QA/QC procedures applied and the results of review. Typically, data quality information will be conveyed as part of FGDC-compliant metadata (see Chapter 7) and will be available via the UCBN website and the NPS Data Store (see Chapter 8).

Chapter 7. Documentation

Documenting data is a time-consuming task. Thorough documentation however, is essential for preserving the integrity and longevity of data and the products of its analysis. In fact, data longevity is roughly proportional to the comprehensiveness of their documentation (Michener 2000). Therefore, documentation is an essential component of sound data management.

Metadata is defined as information about the content, context, quality, structure, accessibility, and other characteristics of data. In addition to ensuring data longevity, metadata increase the possibility of data sharing and reuse for multiple purposes. While the importance of metadata is widely accepted within the data management community, creating and maintaining comprehensive metadata is neither a simple nor quick process. It requires an up-front time investment for planning and organization and an ongoing investment to keep it current.

Laws, Policies and Mandates

Executive Order 12906, signed by President Clinton in 1994, mandates federal agencies to “...document all new geospatial data it collects or produces, either directly or indirectly...” using the Federal Geographic Data Committee (FGDC) *Content Standard for Digital Geospatial Metadata* (CSDGM). In addition, Executive Order 12906 directs agencies to plan for legacy data documentation and provide metadata and data to the public.

The FGDC *Biological Data Profile* contains all the elements of the CSDGM plus additional elements for describing biological data. Metadata created in compliance with the Biological Data Profile can be added to the National Biological Information Infrastructure (NBII, www.nbi.gov/datainfo/metadata) Clearinghouse. Although not a requirement, completion of the Biological Data Profile for appropriate data is recommended by NBII.

The NPS *Metadata Profile* (<http://science.nature.nps.gov/nrdata/docs/npsprofile.cfm>) extends the original seven sections of the CSDGM with information specific to the NPS. The NPS GIS Committee requires all new spatial data be described according to FGDC standards and the NPS Metadata Profile. The NPS Metadata Profile is also utilized by the NPS Data Store for managing GIS and other data for both internal and external use.

Network Standards

In regards to data documentation, the UCBN data management strategy is to:

- Document all significant spatial and tabular data to the required FGDC and NPS standards.
- Maintain and leverage the investment made by NPS staff and cooperators in producing quality data.
- Aid the discovery of relevant data by NPS staff, collaborators and the public.
- Preserve data and its documentation indefinitely.

From project development through final delivery of data products, the data manager and project leaders place a high priority on documenting the aims, quality and meaning of data and allow for the time this will take when scheduling. Metadata adhere to the FGDC and NPS standards and their development are guided by the practices described in the UCBN Metadata Guidance

document included in the appendixes. Project leaders will start metadata at project initiation and set the tone for record-keeping throughout the course of the project. Project leaders are responsible for training crews to record decisions made in the field that affect data quality or meaning.

Content Standards

Network metadata content standards will adhere to those set by FGDC and NPS policy as described above. Any contracts entered into by the UCBN will stipulate the submission of FGDC- and NPS-compliant metadata. The data manager or project leader will assist with metadata acquisition by providing tools, format protocols and file transfer services.

The CSDGM

The CSDGM consists of seven sections (Sections 1-7), all of which contain elements required for spatial data. The NPS Metadata Profile adds another section (Section 0). The Biological and ESRI Profiles are added as elements to Sections 1-7. A summary of the sections is provided below.

Section 0, NPS Information: Documents metadata purpose, park unit and data steward.

Section 1, Identification Information: Includes who produced the data, when and why it was produced and where it is from. Constraints on access and use (e.g., sensitive data) are also recorded in this section. For biological data, this section also includes the geographic extent, bounding altitudes, taxonomy and analytical tools used.

Section 2, Data Quality: Documents the accuracy of attributes and geographic positions and the procedures used to ascertain accuracy. This section also documents the completeness and lineage of the data. Lineage includes data sources and processing steps and for biological data, the methodology used to derive the data.

Section 3, Spatial Data Organization: Includes methods of spatial reference. Mandatory for spatial data and generally created automatically.

Section 4, Spatial Reference Information: Documents coordinate system definitions. Mandatory for spatial data and generally created automatically.

Section 5, Entity and Attribute Information: Includes attribute names, definitions, codes and their meanings and other information essential to a basic understanding of the data.

Section 6, Distribution Information: Describes the methods and contacts used for obtaining the data. Also documents information critical for using biological data formatted in ASCII.

Section 7, Metadata Reference Information: Includes who created the metadata, when it was written, the profile used and the frequency of its update.

Metadata *fully compliant* with FGDC and NPS standards have entries in Section 0 and all element fields in Sections 1-7 considered ‘mandatory’ or ‘mandatory if applicable’. A mandatory

element is one which must be populated for all data. A mandatory if applicable element is one which must be populated if the data exhibit the characteristic being documented by the metadata element. For example, the element defining the vertical coordinate system would be mandatory if the data contains elevation values.

Metadata *minimally compliant* with FGDC and NPS standards have entries in Section 0 and all ‘mandatory’ and ‘mandatory if applicable’ element fields in Sections 1, 6, 7 and Section 2 for biological data. These include the fields used by the NPS Data Store. Data documented to this extent can be distributed via the Data Store's online upload utility.

Requirements by Data Type

Different types of data and information require different kinds and levels of documentation. Standards for documentation of UCBN data are described below.

Spatial Data: Include, at a minimum, all of Section 0 and the required elements of Sections 1-7.

Non-Spatial Data: Include, at a minimum, all elements of Section 0, the required elements of Sections 1, 6, 7 and Section 2 for biological data. The minimum requirements for non-spatial data therefore meet the requirements for minimum compliance with FGDC and NPS standards.

Relational Databases: Document according to the Non-Spatial Data standards above. Complete documentation will also include entity relationship diagrams, business rules and programming code. The NPS Metadata Profile currently does not support this additional documentation, so it will be stored separately from the formal metadata but in the database folder. Relational databases will also use internal documentation such as table and field descriptions and will include a table to track modifications.

Legacy Data: Documented to the extent possible according to the standards outlined above. Metadata that accompany legacy spatial and non-spatial data are suitable for upload to the NPS Data Store if they include entries sufficient for minimum compliance with FGDC and NPS standards. Priority for documentation will be:

- Data needed for current project development
- Data used frequently by park staff
- Historic data archived for possible future use

Data from Outside the Network: Data generated and/or managed outside of UCBN programs but used in analysis with UCBN data or distributed in any manner by the NPS require the same level of documentation produced for UCBN-generated data, including but not limited to, data produced under contract with the NPS. Metadata will be requested from the originating entity by the data manager or project leader.

Generally, external data will not be posted by the UCBN without accompanying metadata. This includes any metadata downloaded with the data, plus additional information regarding date of download and any alterations made to the data by NPS staff. Network staff will occasionally post ‘value added’ external data (e.g., a digital elevation model clipped to park boundaries and

converted to the standard projection) and associated metadata will reflect the source data as well as processing prior to upload.

Sensitive Data: Metadata documentation provides one means of labeling sensitive data to ensure their protection and integrity over time. Sensitive data will be documented according to the standards outlined above and, instead of being uploaded to the NPS Data Store, will be archived in secure locations at the Park and/or Network.

Format and Storage Standards

To enable uploading to the NPS Data Store and further public distribution by NPS servers, metadata must be formatted in a manner that will allow parsing with the USGS *MetaParser*. Proper formatting can be most easily accomplished using the metadata tools described below.

Metadata and other documentation will be stored with the associated data files (i.e., in the same folder), with formal metadata files saved in extendable markup language (XML). Metadata may be exported in text, HTML or other formats as needed but the original XML file will be maintained with the data as the master copy.

Hypertext links within the metadata document will point to accompanying data stored on the Network Data Storage unit. In general, a single metadata document will apply to both raw and certified versions of the data.

Metadata Tools

While a multitude of metadata generation and editing tools are available, UCBN implements only two, ESRI ArcCatalog and the NPS Metadata Tools and Editor (see also the UCBN Metadata Guidance in the appendices).

ESRI ArcCatalog is a multifunction application for managing geospatial data and for editing FGDC compliant metadata. ArcCatalog uses stylesheets to display XML metadata in a format for easy viewing and editing. Because ArcCatalog metadata are linked to the data they describe, certain inherent properties of the data, such as bounds, coordinate system, feature count and attribute names, can be automatically populated and maintained in the metadata.

NPS Metadata Tools and Editor is a custom software application for authoring and editing NPS metadata. It extends the basic functionality of ArcCatalog for managing geospatial metadata and provides a stand-alone tool for creating and manipulating non-spatial metadata outside of ArcCatalog. The Metadata Editor formats metadata according to the NPS Metadata Profile using stylesheets. A variety of stylesheets are offered, each tailored to a specific type of metadata and displaying only the elements or sections pertinent to that type. NPS-specific stylesheets are based on the NPS Metadata Profile extension and contain elements from the FGDC, ESRI and Biological metadata standards and profiles plus NPS-specific elements. Output format is XML and suitable for upload to the NPS Data Store. Metadata Tools also includes utilities for searching, cataloging, parsing (with USGS *MetaParser*) and spell-checking metadata records.

The “*Metadata in Plain Language*” questionnaire (see UCBN Metadata Guidance) is a user-friendly information sheet that can be used to elicit metadata needed to complete all FGDC

required elements for non-spatial data. It currently does not include questions specific to the Biological Data Profile nor the NPS Profile and therefore, does not alone, meet the content standards specified by the NPS Metadata Profile. Additional information must be obtained to comply with these specifications.

Work Flow Process

The Network will establish guidelines for metadata generation and maintenance (UCBN Metadata Guidance, included in appendices). In general, metadata development begins with project design and planning (see Chapter 4). Metadata documents should be created with the tools recommended above to ensure standardization and the ability to upload to the NPS Data Store. Other tools may be used as long as the metadata produced meet the appropriate requirements for content, format and organization. Initially, some sections may be populated by UCBN staff with information from formal written project protocols. As the project progresses, the document will be completed by the project leader with information such as field season dates, protocol changes, taxonomy information, sensitivity and project certification. The data manager or project leader ensures metadata are uploaded to the NPS Data Store and/or other repositories.

Long-term monitoring projects present a different set of metadata questions and requirements that may extend beyond the scope of established FGDC standards. Essential documentation such as algorithms, output files, or spatial analyses may reside in different systems and formats and could potentially be overlooked when distributing or applying the data. Depending on the project, documentation may need to include details on data models or algorithms used, procedures for data synthesis and associated input and output files. Data use and data request histories and secondary research or publications resulting from long-term monitoring projects, may also need to be tracked.

Chapter 8. Ownership, Sharing and Dissemination

As stated in NPS Director's Order #66, "There are two basic pathways by which NPS shares information: by normal, proactive and often collaborative information dissemination activities and as directed by FOIA [Freedom of Information Act] after determining that the information does not qualify as protected (NPS Draft RM 66B 2004)."

For the purposes of releasing data, the term *sharing* means releasing to an individual or entity upon request or as part of a work agreement that benefits the parks. The term *dissemination* includes releases to the public via publications or other standard routes of dissemination such as the NPS Data Store. The NPS defines conditions for the ownership and sharing of data, natural resource collections and results based on funding by the US government. All cooperative and interagency agreements, as well as contracts, should include clear provisions for data ownership and sharing as defined:

- All data and materials collected or generated using NPS personnel and funds become the property of the park from which they were collected.
- Any important findings from research and educational activities should be promptly submitted for publication. Authorship must accurately reflect the contributions of those involved.
- Investigators must share collections, data, results and supporting materials with other researchers whenever possible. In exceptional cases, where collections or data are sensitive or fragile, access may be limited.

Laws, Policies and Mandates

Information gathered and products created with UCBN support are considered property of the NPS (OMB, Circular A-110, Section 36). Section 36 also states the Federal Government has the right to obtain, reproduce, publish or otherwise use the data first produced under an award and to authorize others to receive, reproduce, publish, or otherwise use such data for Federal purposes. Two major pieces of legislation set the tone for information dissemination and protection by federal agencies. They are the Freedom of Information Act (FOIA) and the National Parks Omnibus Management Act of 1998 (NPOMA). Each provides for wide and proactive sharing of information and each recognizes instances where such sharing may be harmful.

FOIA stipulates that the US, including the NPS, must provide access to data and information of interest to the public, regardless of whether or not the federal government created the records. Agencies are directed to post commonly requested items in public reading rooms and/or on public websites. Nine exemptions provide for withholding information. Of these, the third exemption allows withholding information protected under other statutes, which include:

- The Archaeological Resources Protection Act, which requires information on the location and nature of archaeological resources to be protected.
- The National Historic Preservation Act, which directs that information regarding historic resources should be withheld if disclosure would cause significant invasion of privacy or risk harm to historic resources.
- The Federal Cave Resources Protection Act, which prohibits release of information regarding the nature and location of caves.

- Executive Order 13007, Protection and Accommodation of Access to Indian Sacred Sites, which allows for maintaining the confidentiality of Indian Sacred Sites.
- The NPOMA, 16 USC Section 5937, which provides protection of sensitive natural and cultural resource information.

When any of these regulations are applicable, public access to data can be restricted. While NPOMA clearly requires a higher level of information sharing with the research community and other interested publics, as well as use in management decisions, the exemptions in Section 207 allow for protection of information about resources not previously specified in other legislation.

Network Standards and Guidelines

The UCBN shares data and information internally through Network channels, externally through online databases and repositories and by special request as appropriate once quality control procedures have been fully implemented. UCBN data management aims to ensure that:

- Non-sensitive data are easily discoverable and obtainable
- Data not yet subjected to full quality control are not released, unless necessary in response to a FOIA request
- Distributed data are accompanied by fully compliant metadata (see Chapter 7).
- Sensitive data are identified and protected from unauthorized access.
- Clear guidelines for the ownership and sharing of data and information are established.
- A complete record of data distribution/dissemination is maintained

UCBN staff will notify NPS resource managers prior to making data available to the public. This will allow each park resource manager the opportunity to make a request to further restrict access to the data. The UCBN Network Coordinator will review each request and work with the park resource manager to determine how long the data will remain restricted.

Collaborative Agreements

To ensure that proper ownership, format and development of Network products are maintained, all cooperative or interagency work must be conducted as part of a signed cooperative or interagency agreement. Every cooperative or interagency agreement or contract involving the Network must include OMB Circular A-110 cited under the *Reports and Deliverables* Section of all agreements and contracts. The following shows appropriate language to use when citing Circular A-110:

“As the performing organization of this agreement, *institution or organization name* shall follow the procedures and policies set forth in OMB Circular A-110.”

Every cooperative or interagency agreement or contract must include a clearly defined list of deliverables and products. These include, but are not limited to, field notebooks, photographs (hardcopy and digital), specimens, raw data and reports. Details on formatting and media types required for final submission must be included and must follow the most recent UCBN Product Deliverables and Specifications guidance document. The following statement regarding product specifications should be included in the *Reports and Deliverables* section of all agreements and contracts:

“All reports and deliverables must follow the most recent version of the Upper Columbia Basin Network's Product Specifications”.

Researchers should also provide a schedule of deliverables that includes sufficient time for NPS to review draft deliverables before scheduled final submissions.

Mechanisms for Distribution

According to FOIA (specifically the 1996 amendments), all information routinely requested must be made available to the public via reading rooms and/or the Internet. A number of Internet-based databases and repositories have been developed as part of the NPS I&M Program to store and disseminate a variety of natural resource information including:

- The UCBN public website
- National applications (NPSpecies, NatureBib, NPSTORET)
- The NPS Data Store

The Network plans to use the internet as a principal means of distributing I&M data including annual reports and detailed project reports. Links and a brief description will be provided to public versions of the national and Network databases and repositories, if available. The applications and repositories used to distribute products developed by the program are indicated in Table 8.1.

Table 8.1. Primary Repositories for UCBN products.

Repository	Item
UCBN Network Data Storage unit, UCBN website	Project data, metadata and other products • Raw and certified data • Metadata, protocols, SOPs • Reports and administrative records • Digital photographs, derived products
UCBN Project Databases	Comprehensive data for multi-year projects
Park Collections and/or National Archives	Administrative records, voucher specimens*, raw data forms, hard copy reports
National Databases - NPSTORET, NPSpecies, NatureBib	Compiled information on water quality, species lists and taxonomic documentation, park resource bibliographies
NPS Data Store	Copies of digital data (non-sensitive) and metadata

* Biological specimens can also be retained at other facilities (e.g., University of Washington Burke Museum, University of Idaho Herbarium) with an appropriate agreement.

National databases such as NPSpecies and NPSTORET have local desktop databases which can be updated with data collected during the course of a project. Desktop databases are then uploaded and synchronized with the national databases on a regular basis. Synchronization with the master version of NPSpecies occurs annually, or more frequently depending on the timing and amount of updates. NatureBib is updated as new documents become available.

Data collected by water quality projects are directly entered into NPSTORET and managed according to guidelines from the NPS Water Resources Division (WRD). UCBN water-related

data will be formatted according to WRD standards and delivered at least annually for upload to the EPA STORET database (Figure 8.1).

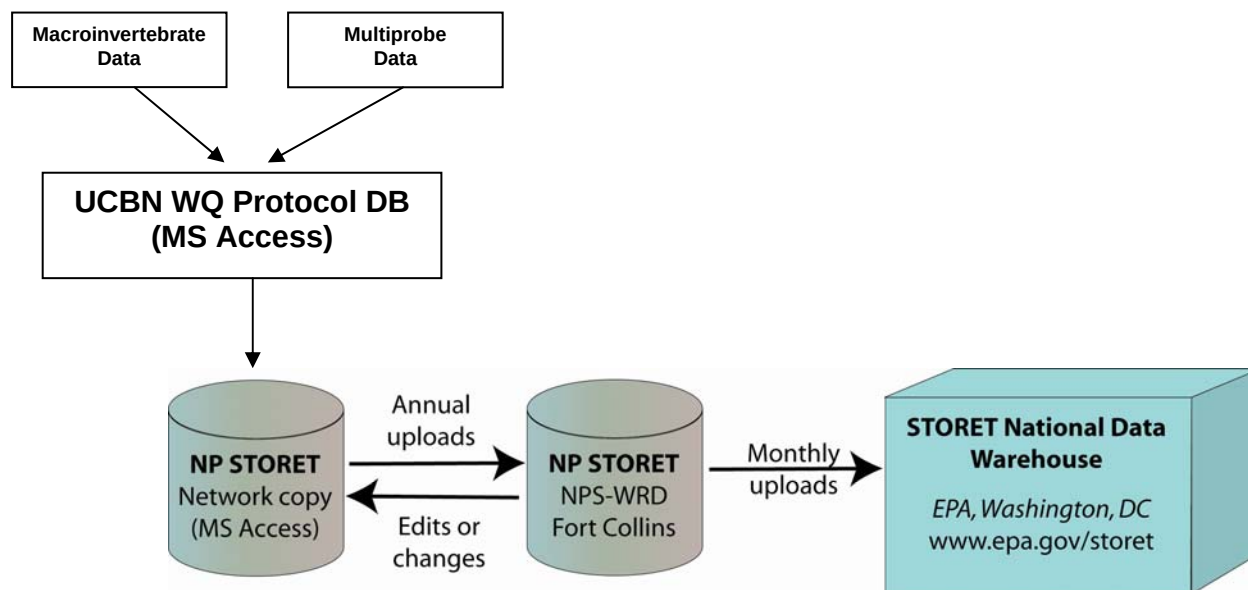


Figure 8.1. Data flow diagram for water quality data.

Data Requests

The UCBN will develop an acquisition policy, including a data sharing agreement, that will be made available to all Network website users who wish to acquire program data and information. This policy will include such things as:

- A mandatory questionnaire which will allow Network staff to maintain a distribution log specifying recipient name and contact information, intended use of data, export file format, delivery date and method and data content description noting range by date and geography of data delivered. It must be completed and emailed or mailed to the data manager before data can be acquired.
- A statement directing the appropriate use and any access restrictions.
- A request that acknowledgement be given to the NPS I&M Program in all resulting reports and publications.

Requests under the Freedom of Information Act

Records requested pursuant to FOIA will be forwarded to the requester via the Park FOIA Officer in a timely manner. The UCBN will maintain records subject to FOIA in a manner so as to make them rapidly available and leave no doubt that the recent versions of accurate, non-sensitive information are provided to requesters. Network staffs are responsible for reading, understanding and following applicable laws, policy and procedural guides.

If information requested under FOIA has not already undergone sensitivity classification, it will be thoroughly but rapidly assessed for presence of sensitive information and, if such information is found, the data manager will work with appropriate Park and Regional FOIA Officers to formulate a response to the requester, which may include complete denial, complete fulfillment, or partial fulfillment once sensitive information have been cleared from requested records.

Access Restrictions on Sensitive Data

Data released outside of the federal government without a confidentiality agreement is considered a release of information under FOIA and triggers application of the “release to one, release to all” principle. There are many venues for release of information that constitute release to the public including:

- Internet posting
- Interpretive programs, media and activities
- News releases
- Resource management storing and sharing activities
- NEPA compliance documents.

Only a small subset of information collected through the I&M Program is likely to be considered protected. Nevertheless, all data and associated information from UCBN activities, including spatial data, should be assessed to determine sensitivity and any protected information should be carefully managed to prevent its release. Inadvertent access may be deemed a release depending on the circumstances. Network staff will classify sensitive data on a case-by-case, project-by-project basis and will work closely with park staff and project leaders to ensure potentially sensitive park resources are identified, information about these resources is tracked throughout the project and potentially sensitive information is removed from documents and products that will be released outside the Network.

Feedback Mechanisms

Overall comments and questions concerning UCBN I&M Program are welcome at any time and may be submitted via e-mail or telephone to the Network Coordinator or Data Manager. The UCBN website provides an opportunity for NPS staff, cooperators and the public to provide feedback on data and information gathered as part of Network programs.

The following procedure describes the UCBN process to receive, verify and correct data errors identified by public users:

1. Users send a notification email through the UCBN website to the Network Coordinator or Data Manager.
2. An acknowledgment of receipt is sent to the user.
3. Network staffs determine whether the data in question are correct or incorrect. If correct, the user is informed that no corrections are to be made and the information stands. If incorrect, Network staffs make the appropriate corrections and notify the user and original data collectors.
4. The Network website is refreshed with the corrected information

Chapter 9. Records Management

Electronic and hard-copy products resulting from UCBN projects and activities are crucial pieces of information that require management to ensure their long-term maintenance. An essential part of any archive (digital or non-digital) is the accompanying explanatory materials (Olson and McCord 1998). Thus, effective long-term maintenance of these products depends on thoughtful and appropriate data documentation (see Chapter 7). While this chapter is not intended to provide a full description of archiving procedures, it is expected to provide guidance in making the transition from completed products delivered by the project leader to a secure and long-term storage facility managed by park and/or regional repositories.

National Policies

Direction for managing many I&M products is provided in NPS Director's Order 19: Records Management (2001) and its appendix, NPS Records Disposition Schedule (NPS-19 Appendix B, revised 5- 2003). NPS-19 states that all records of natural and cultural resources and their management are considered mission-critical records, that is, necessary for fulfillment of the NPS mission. NPS-19 further states:

Mission critical records are permanent records that will eventually become archival records. They should receive the highest priority in records management activities and resources and should receive archival care as soon as practical in the life of the record.

Section N of NPS-19 Appendix B, which provides guidelines on natural resource-related records (including, specifically, products of I&M Programs), indicates all natural resource records are considered "permanent" and need to be retained either in an appropriate park museum facility or at the National Archives. It also indicates that non-archival copies of natural resource-related materials are "...potentially important for the ongoing management of NPS resources" and should not, in any instance, be destroyed.

The NPS Museum Handbook (NPS 2005) provides the overarching guidance for archival procedures. In particular, Part II, Appendix A: Mandates and Standards for NPS Museum Collections list the cultural and natural history laws, regulations and conventions for NPS museum collections and should be reviewed prior to object collections.

Network Objectives

To ensure high-quality long-term management and maintenance of program information, UCBN will implement record management procedures to:

- Avert the loss of information over time
- Ensure Network information can be easily obtained, shared and properly interpreted by a broad range of users.
- Standardize on Microsoft products as required by all Department of Interior (DOI) agencies (DOI, Assistant Secretary for Policy, Management and Budget, Findings and Determination, September 13, 2002)
- Update and maintain data to no more than two versions behind current version of software, or store the data in American Standard code for Information Interchange (ASCII) format, complete with data and file documentation.

General Guidelines

Process and Workflow

Both electronic and hardcopy files should be consolidated and packaged for archival when a project is complete or, for long-term projects, when milestones are reached. UCBN project leaders are responsible for packaging electronic files and data for the data manager, who is then responsible for product archiving, integration, backup, and distribution. Project leaders are also responsible for preparing materials for the curator. The workflow illustrated here refers to general procedures for making the transitions between project leaders, data managers and museum staff (Figure 9.1). The museum staff will follow a much more detailed procedure for actual curation.

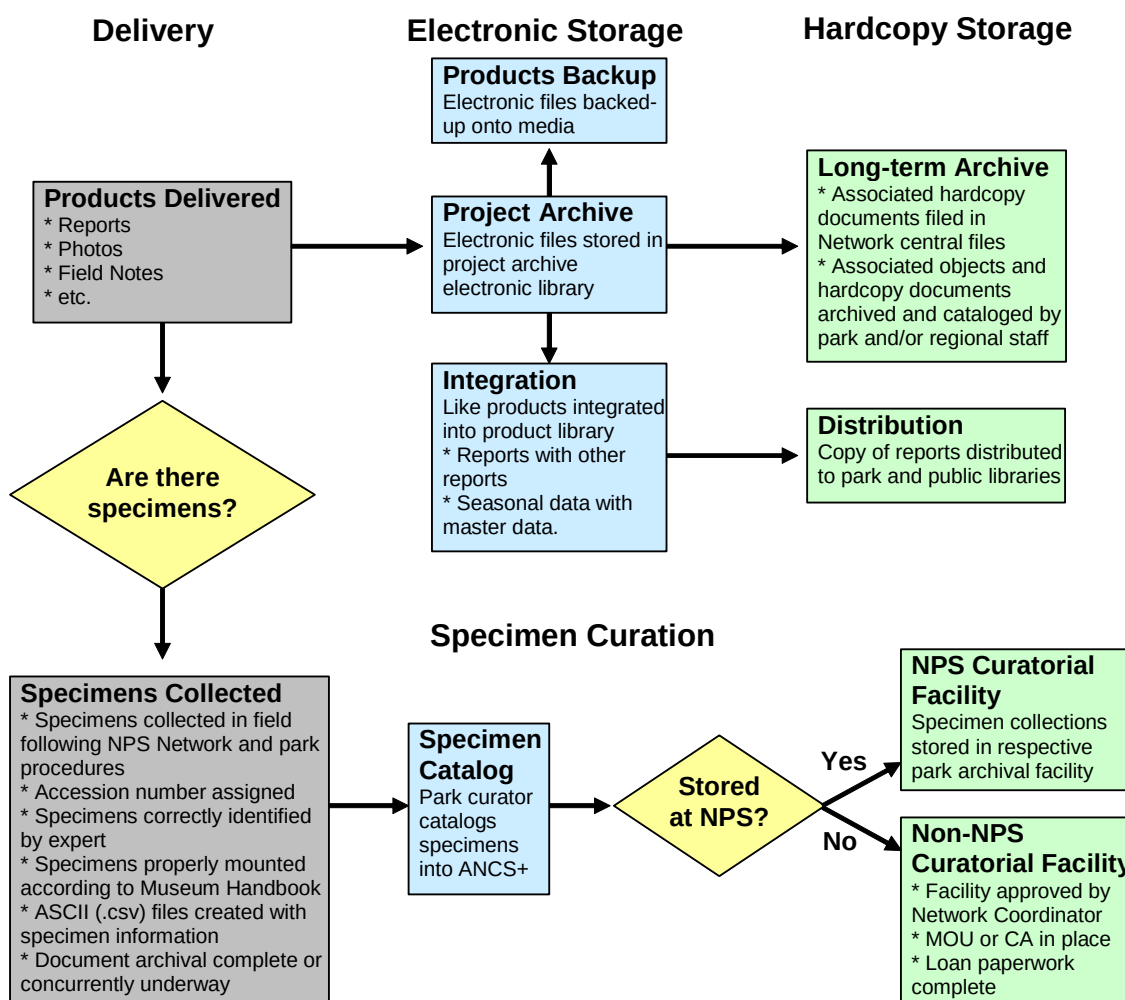


Figure 9.1. General workflow for storage and archiving of electronic (in blue) and hardcopy (in green) products.

For digital data, the project leader should prepare files as follows:

- Comply with file naming standards (see Chapter 6).
- Comply with documentation standards (see Chapter 7).

- Comply with Project Specifications guidance (see Product Specifications Guidance in appendices).
- Make a clear distinction between draft and final files. Draft files will be included in project backup, but will be deleted for final project archival.
- Make a clear distinction between public and sensitive information.
- Deliver all products and ensure acceptance by UCBN Coordinator and UCBN Data Manager.

The data manager will:

- Archive the project information on the UCBN Network Data Storage unit (see Chapter 3 for details on directory structure).
- Integrate deliverables into appropriate catalogs, electronic libraries and NPS clearinghouses, as described in this plan or in the monitoring protocols.

For non-digital data, the project leader should consult with the appropriate archival contact at the affiliated Park, and follow these general guidelines:

- Complete a coversheet listing contact information, project abstract and purpose, sensitivity, and a full inventory of non-digital data products being delivered.
- Clearly label materials with park acronym, date(s), accession number, project number.
- Comply with paper and folder type requirements (e.g., acid free paper).
- Print and document photos.
- Label voucher specimens according to the Museum Handbook and complete loan paper work if necessary. Specimens not residing in NPS repositories should be stored in a selected facility that meets NPS museum collection standards.
- Address specimen curation in vital signs protocols and specify name and address of facility, Memorandum of Understanding or Cooperative Agreement (if applicable), list of items to be curated or loaned.
- Create a CD or DVD of all related, electronic materials and label with park acronym, date created, date(s) of information, accession number, project number, and list of contents.

Park museum staff will catalog this information into ANCS+ and will properly store materials according to the Museum Handbook.

Archiving and Maintenance - Digital Data

In general, digital data maintained over the long-term will be one of two types: short-term data, for which collection and modification have been completed (i.e., inventory projects); and long-term monitoring data, for which data acquisition and entry will continue indefinitely.

Technological obsolescence is a significant cause of information loss and data can quickly become inaccessible to users if stored in out-of-date software programs or on outmoded media. Effective maintenance of digital files depends on proper management of a continuously changing infrastructure of hardware, software, file formats, and storage media. Major changes in hardware can be expected every 1-2 years and in software every 1-5 years (Vogt-O'Connor 2000). As software and hardware evolve, data must be consistently migrated to new platforms, or they must be saved in formats independent of specific platforms or software (e.g., ASCII delimited files).

Following the lead of the national I&M Program, UCBN has adopted MS Access as its database standard and ArcGIS as its spatial data management standard. The Network will remain current and compatible with NPS and national I&M program version standards for this software.

Short-term Data

Upon project finalization, a set of ASCII comma-delimited text files will be created for each data table comprising the data. These files will be accompanied by a text file that explains the contents, relationships and field definitions. The ASCII files are in addition to the native version of the data (typically in database or spreadsheet format) and will help ensure the data are usable in a wide range of applications or platforms.

The UCBN will employ a utility created by the Gulf Coast Network called `Access_to_ASCII.mdb`, which automates the creation of these ASCII files. Even so, all ASCII files created from databases will undergo quality control to ensure the number of records and fields correspond to the source data and conversion has not created errors or data loss. A second reviewer (preferably the project leader) will evaluate the ASCII files and documentation to verify tables, fields, and relations are fully explained and presented in a way that is useful to secondary users. All finalized files will be stored on the UCBN Network Data Storage unit in the appropriate protocol or project Close folder (see Chapter 3 for directory structure details).

In addition to creating ASCII files, the UCBN will also update completed and archived data that may be in older versions of MS Access, with the goal of having no data more than two versions behind the current version used by the Network. Converted databases will require additional QC; in particular, if the databases are being actively used for data entry or analysis. While forms, queries, reports and data entry will be thoroughly tested, there is a risk of reduced performance (e.g., complex data entry forms or reports may not function properly in an upgraded version). To the extent possible, proper functionality of data entry forms and reports will be maintained. However, the priority is to ensure basic table and relationship integrity. All previous versions of the data will be saved.

Long-term Data

Long-term monitoring data require regular updates and conversion to current database formats. All active or long-term databases will conform to current I&M software version standards.

Monitoring projects will also have variable long-term data archiving requirements. Raw data that are later manipulated or synthesized may need to be stored in perpetuity. Modifications to protocols will typically require complete data to be archived before modifications are implemented. Similarly, depending on the monitoring project, it may be necessary to preserve interim data “milestones” over the long term. Archived data sets or subsets destined for long-term archiving will be saved, whenever possible, in native formats in addition to ASCII text files. Specific data archiving requirements for ongoing projects are written into the data management SOPs for each monitoring project.

Spatial Data

Spatial data essential to UCBN will be maintained in a format that remains fully-accessible by the current ArcGIS version. ArcGIS has maintained compatibility with previous data formats,

though future conversions may be required. At this time there is no practical way to save GIS data in a software or platform-independent format.

Both uncorrected and corrected GPS data (e.g., .SSF and .COR files) will be archived in their native format in addition to the corresponding GIS files created.

Backups and Security

The risk of data loss can come from a variety of sources, including catastrophic events (e.g., fire, flood), user error, hardware failure, software failure or corruption and intentional acts of vandalism. Performing regular backups of data and arranging for off-site storage of backup sets are the most important safeguards against data loss.

All UCBN data reside on a Network Data Storage unit configured as RAID 5. The Network employs an automated backup system to a rotating set of external hard drives. One current hard drive is always stored off site. Details of this backup schedules will be developed in cooperation with IT specialists and documented in a Data Backup SOP (in development). The backup schedule will allow for daily backup of frequently-changing files (facilitating recovery of data from the previous 4 weeks) and a weekly backup (facilitating recovery of data from the previous 12 months). Off-site storage of backups will be located either at the home of the UCBN Coordinator or the UCBN Data Manager. All backups are performed and monitored by the UCBN Data Manager.

Backups of data stored on the personal computers of staff are the responsibility of each staff member. Every computer is equipped with an external hard drive and uses Microsoft XP built-in backup capabilities to perform daily backups of selected files and folders. Because the quantity of UCBN data is increasing rapidly, this backup schedule and system will need to be re-evaluated at least annually. Backup routines represent a significant investment in hardware, media and staff time; however, they are just a small percentage of the overall investment made in program data.

UCBN local and wide area networks conform to DOI security guidelines. Only I&M staff and NPS system administrators have permission to access files on the UCBN Network Data Storage unit and restrictions have been established on archived data files. Directories containing completed project data or interim versions of ongoing projects are designated as read-only for all staff with the exception of the data manager. In this way, any changes must be routed through the data manager, who is responsible for ensuring documentation and readme files associated with the data are updated.

Archiving and Maintenance - Non-digital Data

The following guidelines apply to hardcopy documents (e.g., final reports prepared by staff or contractors, program administrative documents, contracts and agreements, memoranda of understanding, etc.) related to UCBN administration, activities and projects. The guidelines also apply to physical items such as natural history specimens, photographs or audio tapes. In most instances, these documents and objects are essential companions to digital data archives described above.

Anyone submitting materials to an archive will provide essential cataloging information such as the scope of content, project purpose and non-record copies of documents presented using archival-quality materials (acid-free paper and folders, polypropylene slide pages, etc.).

Documents

All paper documents managed or produced by the UCBN will be housed in one of three locations:

- UCBN central files, Moscow, ID - These files contain project files, administrative documents and non-record copies of documents archived at regional or offices. Examples include: meeting minutes, correspondence, memoranda of understanding, contracts and agreements, research permits, interim and selected final reports produced by the program or under its auspices. In addition to maintaining paper records, the UCBN will maintain electronic versions when possible. The central hard copy files are maintained by the Network Coordinator and the Data Manager.
- Park-based Museums, Libraries and Central files – The UCBN will provide high-quality copies of park-related documents resulting from UCBN projects, along with electronic versions, to park resource management staff. Parks may choose to house these materials as they deem appropriate. The UCBN will not manage documents at the park level.
- Technical Information Center (TIC), Denver, CO - TIC provides temperature and humidity-controlled facilities, a professional archival staff and meets all museum standards set by NPS. This repository will be used for technical reports identified by the Network as high priority to maintain under archival conditions.

Specimens

The Network will provide specimens collected under the auspices of the UCBN to the park in which they were collected for inclusion in the park repository or to a repository approved by a park (where the specimens are considered on loan). The UCBN will provide park curators with associated data required for cataloging each specimen in comma-delimited format (.CSV) format for automated uploading into the Automated National Catalog System (ANCS+). Data will be provided to non-NPS curators in Excel format.

Photographs

All photographs taken as part of UCBN projects will be digital and will follow the UCBN Photo Database User Guidance (see appendices). Specific methods to be used by any given project may also be detailed in the project protocol. In general, digital photographs will be saved as high resolution (≥ 600 dpi) Tagged Image File Format (TIFF) and may be copied to low resolution (≤ 100 dpi) Joint Photographic Experts Group (JPEG) files. Accurate camera date and time information will be maintained and digital headers used instead of imprinted date/time on the image.

Every image will be entered into the UCBN photo database where attributes such as file name, keywords, project, photo description, photographer, date and location are cataloged. All photo files and the associated photo database are stored on the UCBN Network Data Storage unit.

Role of Curators

Curators for parks within the UCBN are an ongoing source of expertise, advice and guidance on archiving and curatorial issues. Project leaders will involve park curators to ensure all aspects of specimen preservation or document archiving are considered and any associated expenses are included in project budgets.

Chapter 10. Implementation

The I&M Program focuses attention on data consumers – managers who will make decisions based on the data provided or researchers who will use these data decades from now. Effective data management practices are essential to providing for these consumers. Data collection is expensive and ecological data often irreplaceable, making data management a critical aspect of any ecological data collection effort.

Good data management practices take time and money. While, some vital sign collection procedures and management practices are already in use and require minimal revisions, others may involve several iterations before the procedures and databases are acceptable and functional. The national I&M Program anticipates at least one third of each Network's efforts will involve data management. The data management plans for each of the 32 I&M Networks are the first comprehensive documents of their kind and establish the means by which managers, educators and the greater scientific community access information on Network and park resources.

Education and Training

Some practices specified in this DMP may be new to staff and cooperators. With few exceptions, however, it does not include any new requirement. Almost every obligation comes from law, Executive Orders, Director's Orders, or national I&M Program guidance. The DMP helps to put these requirements into context and provides operational guidance to achieve them.

Education and training on the practices and procedures implemented by the UCBN will be necessary to minimize some common data management problems of the past such as:

- *Unfinished projects.* Failure to finish a project can happen for a number of reasons beyond the staff's control. Unfinished projects, however, are the most difficult to manage. It takes several times longer to document an incomplete project than a finished one. The best data management practice is to finish the project or reach a logical milestone.
- *Lack of header information.* Often, sheets of information do not have a date, name or subject. Reading the material and determining its relevancy is extremely time-consuming. Often relevant information is discarded because it does not have adequate header information. A good data management practice is to use a memorandum format or standard header for all project files, no matter how insignificant.
- *Lack of documentation.* Most projects and data are inadequately documented. Without documentation, the data are often considered unusable. Documentation techniques should be applied throughout the duration of the project (see Chapter 7).
- *Lack of file naming standard and version control.* In many cases it is impossible to know which file contains the final product (see Chapter 6).
- *No archive or final storage.* Data tucked away in an individual's desk or personal computer is inaccessible to others and typically requires in-house knowledge to know the information even exists. With staff turnover, this information is often lost.

Measuring Success

The UCBN is just beginning to implement this plan. The coming years will involve learning, testing and refining. Templates, SOPs and guidance documents will be developed and tested and then applied to other vital signs monitoring protocols. Some bottlenecks are anticipated and their identification and elimination is an important part of successful implementation.

Goals for Years 1-5

- All Network staff understand the fundamentals of data management, including:
 - File management
 - Documentation
 - Quality assurance and quality control
 - Electronic storage
 - Archive storage
- Data management practices are improved by implementing:
 - Data models for vital signs monitoring protocols
 - Testing of data entry prior to field work
- Common SOPs and/or guidance documents for multiple protocols are developed
- Data management staff informs and directs data management aspects of the vital signs monitoring protocols implemented.

Goals for Years 5+

- The DMP is revised and streamlined for vital signs monitoring. Generalizations are reduced or eliminated.
- All implemented vitals signs monitoring protocols follow the DMP guidance.
- Databases and custom reporting continue to be enhanced.
- Framework and gateway for integration of data with other agencies or networks is established.

Plan Revisions

Revision and maintenance of the DMP will be incremental to reflect best practices and current information. Recommendations for changes may be forwarded to the UCBN data manager at any time and by any interested party or user of UCBN I&M data. The data manager is responsible for on-going maintenance (including documentation of any changes – see Change History) of the DMP and changes will be made at his/her discretion only after thorough review and assessment of anticipated ramifications. The amount of time required for adequate review and assessment will likely vary greatly in accordance with the nature of the suggested change. When enough revisions have been made to merit an additional round of review, this process will be undertaken with the Network Science Advisory Committee and the national I&M Data Manager. Otherwise, the plan will be scheduled for a more formal revision and review on a 5-year basis.

Literature Cited

- Edwards, D. 2000. Data quality assurance. Pages 70-91 *in* Michener, W.K. and J.W. Brunt, eds. Ecological data: Design, management and processing. Blackwell Science Ltd., Methods in Ecology Series, Malden, MA.
- Garrett, L.K., T.J. Rodhouse, G.H. Dicus, C.C. Caudill and M.R. Shardlow. 2007. Upper Columbia Basin Network vital signs monitoring plan. Natural Resource Report NPS/UCBN/NRR—2007/002. National Park Service, Fort Collins, Colorado.
- Michener, W.K. 2000. Metadata. Pages 92-116 *in* Michener, W.K. and J.W. Brunt, eds. Ecological data: design, management and processing. Blackwell Science Ltd., Methods in Ecology Series, Malden, MA.
- Michener, W.K. and J.W. Brunt, eds. 2000. Ecological data: design, management and processing. Blackwell Science Ltd., Methods in Ecology Series, Malden, MA.
- National Park Service. 2002. Director's Order #11B: Ensuring quality of information disseminated by the National Park Service. Available online at: <http://home.nps.gov/applications/npspolicy/DOrders.cfm>
- National Park Service. 2004. Director's Order #66: Draft Reference Manual 66B – Handling Protected Information.
- National Park Service. 2005. The Museum Handbook. National Park Service, Washington, DC. Available online at: <http://www.cr.nps.gov/museum/publications/handbook.html>
- Olson, R.J. and R.A. McCord. 1998. Data Archival. Pages 53-57 *in* Michener, W.K., J.H. Porter and S.G. Stafford. Data and information management in the ecological sciences: a resource guide. LTER Network Office, University of New Mexico, Albuquerque, NM.
- Palmer, C.J. 2003. Approaches to quality assurance and information management for regional ecological monitoring programs. Pages 211-225 *in* Busch, D.E. and J.C. Trexler, eds. Monitoring ecosystems: Interdisciplinary approaches for evaluating ecoregional initiatives. Island Press, Washington, DC.
- Palmer, C.J. and E.B. Landis. 2002. Draft Lake Mead National Recreational Area Resource Management Division: Quality System Management Plan for Environmental Data Collection Projects.
- Tessler, S. and J. Gregson. 1997. Draft Data Management Protocol. NPS I&M Program. (<http://www1.nrintra.nps.gov/im/IMProto/joe40001.htm>).
- Vogt-O'Connor, D. 2000. Planning digital projects for preservation and access. National Park Service Conserve O Gram 14:4. NPS Museum Management Program, Washington, DC.

List of Acronyms and Abbreviations

ANCS+	Automated National Catalog System
ASCII	American Standard code for Information Interchange
CDM	Conceptual Data Model
CSDGM	Content Standard for Digital Geospatial Metadata
DOI	Department of Interior
EPMT	Exotic Plant Management Team
FGDC	Federal Geographic Data Committee
FOIA	Freedom of Information Act
FTP	File Transfer Protocol
GIS	Geographic Information System
GPS	Global Positioning System
HTML	HyperText Markup Language
I&M	Inventory and Monitoring
DMP	Data Management Plan
IT	Information Technology
JPEG	Joint Photographic Experts Group
LAN	Local Area Network
LDM	Logical Data Model
MS	Microsoft
NAD	North American Datum
NBII	National Biological Information Infrastructure
NPOMA	National Parks Omnibus Management Act of 1998
NPS	National Park Service
NRDT	Natural Resource Database Template
NRPC	Natural Resource Program Center
OMB	Office of Management and Budget
PDA	Personal Data Assistant
PDM	Physical Data Model
PWR	Pacific West Region
QA/QC	Quality Assurance Quality Control
RAID	Redundant Array of Independent Disks
SOA	Service Oriented Architecture
SOP	Standard Operating Procedure
T&E	Threatened and Endangered
TIC	Technical Information Center
TIFF	Tagged Image File Format
UCBN	Upper Columbia Basin Network
WAN	Wide Area Network
XML	Extensible Markup Language

Glossary of Terms Used in the Data Management Plan

Cheat sheets: simple, sequential procedures for conducting specific data management practices (e.g., cheat sheet on linking digital images to Access databases); usually included as supplementary sections in overall data management plan, guidance documents or SOPs.

Guidance documents: general descriptions of data management practices and procedures; generally standalone documents that inform the ‘what’ and ‘how’ of data management practices; can reference SOPs and standards.

Inventory study/project: a field sampling and/or data mining effort designed to gather data for a natural resource inventory; may or may not be repeated.

Long-term monitoring study/project: a monitoring protocol designed to assess the status of a natural resource vital sign over time.

Monitoring Protocol: formal document describing how a vital sign will be monitored; composed of a narrative section, standard operating procedures and supplementary information (databases, reports, tools, hardcopy materials); a Monitoring Protocol is a component of a vital sign monitoring study plan.

Network Data Storage unit: dedicated data storage connected directly to a computer network to provide centralized data access and storage.

Policy/ies: Rationale(s)/guiding principles for data management practices and procedures (the ‘Why?’ of data management planning).

Project: general term for a formal data gathering effort; ‘formal’ implies an organized, documented approach that includes, at minimum, a study plan/proposal detailing the protocol(s) followed; projects can be inventory studies, pilot studies, short-term/park-specific studies or longer-term monitoring projects.

Protocol: a set of sampling procedures and/or experimental processes followed in a project

RAID 5: a redundant array of independent disks (RAID) refers to a data storage scheme using multiple hard drives to share or replicate data among the drives; level 5 uses block-level striping with parity data distributed across all member disks.

Short-term monitoring study/project: pilot study designed to guide, test and evaluate monitoring protocol development.

Standard Operating Procedures (SOPs): detailed step-by-step instructions for carrying out procedures (the ‘How?’ of data management planning); dynamic in nature and frequently updated; referenced by overall data management plan, monitoring protocols, and guidance documents.

Standard: an data management practice or policy implemented Servicewide or Network-wide (e.g., recommended naming conventions, FGDC metadata format); referenced by overall data management plan, monitoring protocols and SOPs.

Tips: small ‘how-to’ sections within larger data management documents; may evolve into full cheat sheets and/or SOPs.

The Department of the Interior protects and manages the nation's natural resources and cultural heritage; provides scientific and other information about those resources; and honors its special responsibilities to American Indians, Alaska Natives, and affiliated Island Communities.

NPS/UCBN/NRR—2007/0120

**National Park Service
U.S. Department of the Interior**



Natural Resource Program Center

Natural Resource Program Center
1201 Oakridge Drive, Suite 150
Fort Collins, Colorado 80525

www.nature.nps.gov

EXPERIENCE YOUR AMERICA™