National Park Service
U.S. Department of the Interior

Northeast Region
Philadelphia, Pennsylvania

Eastern Rivers and Mountains Network
Data Management Plan

Natural Resource Report NPS/NER/NRR—2009/020

ON THE COVER
A U.S. Geologic Survey Geodetic Control Point, NPS.
ERMN GIS Intern performing GPS work, used with permission from subject, NPS.
View of New River Gorge National River, NPS.

Eastern Rivers and Mountains Network
Data Management Plan

Natural Resource Report NPS/NER/NRR—2009/020

Nathan Piekielek

Eastern Rivers and Mountains Network
422 Forest Resources Building
Pennsylvania State University
University Park, PA 16802

December 2006 (Revised December 2009)

U.S. Department of the Interior
National Park Service
Northeast Region
Philadelphia, Pennsylvania

The Northeast Region of the National Park Service (NPS) comprises national parks and related areas in 13 New England and Mid-Atlantic states. The diversity of parks and their resources are reflected in their designations as national parks, seashores, historic sites, recreation areas, military parks, memorials, and rivers and trails. Biological, physical, and social science research results, natural resource inventory and monitoring data, scientific literature reviews, bibliographies, and proceedings of technical workshops and conferences related to these park units are disseminated through the NPS/NER Technical Report (NRTR) and Natural Resources Report (NRR) series. The reports are a continuation of series with previous acronyms of NPS/PHSO, NPS/MAR, NPS/BSO-RNR, and NPS/NERBOST. Individual parks may also disseminate information through their own report series.

Natural Resources Reports are the designated medium for information on technologies and resource management methods; "how to" resource management papers; proceedings of resource management workshops or conferences; and natural resource program descriptions and resource action plans.

Technical Reports are the designated medium for initially disseminating data and results of biological, physical, and social science research that addresses natural resource management issues; natural resource inventories and monitoring activities; scientific literature reviews; bibliographies; and peer-reviewed proceedings of technical workshops, conferences, or symposia.

Mention of trade names or commercial products does not constitute endorsement or recommendation for use by the National Park Service.

The statements, findings, conclusions, recommendations, and data in this report are solely those of the author(s), and do not necessarily reflect the views of the U.S. Department of the Interior, National Park Service.

Print copies of reports in these series, produced in limited quantity and only available as long as the supply lasts, or preferably, file copies on CD, may be obtained by sending a request to the address on the back cover. Print copies also may be requested from the NPS Technical Information Center (TIC), Denver Service Center, PO Box 25287, Denver, CO 80225-0287. A copy charge may be involved. To order from TIC, refer to document NPS 962/100731.

This report may also be available as a downloadable portable document format file from the Internet at http://www.nps.gov/nero/science/.

Please cite this publication as:

N. B. Piekielek. 2006 (Revised December 2009). Eastern Rivers and Mountains Network Data Management Plan. Natural Resource Report NPS/NER/NRR—2009/020. National Park Service. Philadelphia, PA.

NPS 962/100731 December 2006 (Revised December 2009)

Table of Contents

Page

Figures vii

Tables viii

Appendixes ix

Executive Summary xi

Acknowledgments xxiii

Chapter 1. Introduction 1

1.1. Eastern Rivers and Mountains Network 2

1.2. ERMN Data Management Goals and Objectives 3

1.3. Data Defined 3

1.4. Sources of Natural Resource Data 4

1.5. ERMN Data Management Plan Scope and Organization 5

1.6. Intended Audience 6

1.7. Revisions 6

1.8. Credits 6

Chapter 2. Data Management Roles and Responsibilities 7

2.1. Definition of Roles Versus Responsibilities 7

2.2. Roles and Responsibilities of Network Staff 7

2.3. Coordination 13

2.4. Documentation 14

2.5. Awareness and Accountability 15

2.6. Communication 15

2.7. Credits 15

Table of Contents (Continued)

Page

Chapter 3. Data Life Cycle and Work Flow 17

3.1. Project Workflow 17

3.2. Data Life Cycle 20

3.3. Integrating and Sharing Data Products 23

3.4. Credits 25

Chapter 4. Infrastructure and System Architecture 27

4.1. Network and System Resources 27

4.2. ERMN Digital File Storage and Maintenance 28

4.3. National Information Management System Architecture 30

4.4. Credits 32

Chapter 5. Data Acquisition and Processing 33

5.1. Eastern Rivers and Mountains Network Data 33

5.2. Other National Park Service Data 37

5.3. External Data 37

5.4. Credits 38

Chapter 6. Quality Assurance and Quality Control (QA/QC) 39

6.1. Quality Assurance and Quality Control Defined 39

6.2. General Operations 40

6.3. Monitoring Protocol Development 42

6.4. Data Collection 42

6.5. Data Entry and Import 43

6.6. Data Verification (Quality Control Part 1) 44

6.7. Data Validation (Quality Control Part 2) 45

Table of Contents (Continued)

Page

6.8. Data Quality Review and Communication 47

6.9. Conformance Guidelines and Revisions 47

6.10. Credits 48

Chapter 7. Data Documentation 49

7.1. NPS Standards 49

7.2. Network Standards 49

7.3. Project Documentation 49

7.4. Vital Signs Protocol Documentation 51

7.5. Additional ERMN Documentation 51

7.6. Credits 52

Chapter 8. Data Analysis and Reporting 53

8.1. Data Analysis 53

8.2. Data Reporting 53

Chapter 9. Data Dissemination 55

9.1. Data Ownership 55

9.2. Data Sensitivity 57

9.3. Access Restrictions to Sensitive Data 59

9.4. Data Distribution Mechanisms 61

9.5. Feedback Mechanisms 63

9.6. Credits 63

Table of Contents (Continued)

Page

Chapter 10. Data Maintenance, Storage and Archiving 65

10.1. Digital Archiving and Storage 65

10.2. Digital Storage and Archiving Procedures 68

10.3. ERMN Hardcopy and Digital Reports 70

10.4. Specimens 71

10.5. Photographs 71

10.6. Credits 72

Chapter 11. Literature Cited 75

Figures

Page

Figure 1.1. Map of ERMN parks. 2

Figure 2.1. Intersection of data management responsibilities. 13

Figure 3.1. Conceptual model of project workflow. 18

Figure 3.2. Diagram of typical project data life cycle. 22

Figure 3.3. Data flow diagram for water quality data. 26

Figure 4.1. ERMN project folder structure. 28

Figure 4.2. ERMN park data folder. 29

Figure 4.3. Model of national level application architecture. 30

Figure 5.1. Inventory data flow model. 34

Figure 5.2. Data flow model for monitoring projects. 35

Figure 5.3. ERMN project tracking database. 36

Figure 6.1. QA/QC controls and project workflow. 41

Tables

Page

Table 1.1. Categories of data products. 3

Table 2.1. Categories of data stewardship. 9

Table 2.2. Summary of roles and responsibilities by position. 10

Table 3.1. Repositories for ERMN products. 24

Table 10.1. Principal ERMN repositories for project products. 73

Appendixes

No. of Pages

Appendix A. NER I&M Product Specifications. 24

Appendix B. NER Natural and Social Science Study Proposal and Deliverable Guidelines. 87

Appendix C. NER Technical Report Template. 8

Appendix D. NER Inventory Product Review Work Flow. 12

Appendix E. NER Vegetation Mapping Project Work Flow. 24

Appendix F. NER NPSpecies Data Entry Standards. 21

Appendix G. NER NatureBib Data Management and Data Entry Standards Manual. 30

Appendix H. NER NatureBib Proposed Update Protocol. 3

Appendix I. NPS Director's Orders II A-C. 32

Appendix J. NER NPSpecies Certifier's Guide. 10

Appendix K. NER NatureBib Data Management Plan. 5

Appendix L. NER NPSpecies Data Management Plan. 6

Appendix M. Scanning Specifications and Personnel Commitment. 7

Appendix N. Proposed NPS Image Metadata Standard. 23

Appendix O. NER GPS Metadata Questionnaire Form. 3

Appendix P. ERMN NPSpecies Museum Voucher Fields. 2

Appendix Q. Multi-Park Project Collections. 5

Executive Summary

Collecting data on specific natural resource parameters is the first step toward improving our understanding of National Parks ecosystems. We use these data in tandem with sound data management practices to analyze, synthesize, and model aspects of ecosystems. We use analysis results and interpretations to make decisions about the Park's vital natural resources and how to manage them in accordance with the National Park Service (NPS) mission. Thus, data collected and maintained through sound data management practices become information through analyses, syntheses, and modeling. This can only be achieved through the development of a modern information management infrastructure (e.g., staffing, hardware, software) and procedures to ensure that relevant natural resource data collected by NPS staff, cooperators, researchers, and others are entered, quality-checked, analyzed, reported, archived, documented, cataloged, and made available to others for management decision-making, research, and education.

Eastern Rivers and Mountains Network Data Management Program Goals

The goals of Eastern Rivers and Mountains Network (ERMN) data management program are to make data and information:

- Available – We have the means of knowing that the data exist. Data are accessible and easily located.
- Usable – Data are stored in a stable, reliable, and interpretable data retrieval system.
- Shareable – Data products are complete, certified for quality assurance, screened for sensitive information, formatted for use, and documented for interpretation by others.
- Integrated – Data products are consistent with data exchange standards, interoperable with related natural resource data sets, and are collected and stored in a way that optimizes the tradeoff between meeting local needs and achieving compatibility with other agencies and partners.
- Interpreted – Data have been reviewed, summarized, and transformed into useful information.

The ERMN data management strategy holds that all data and derived information generated or otherwise used by the program will meet a high level of quality standards. Further, all data and information the ERMN program deems necessary to meet its objectives, and that are not otherwise maintained, will be archived, documented, and made easily available and accessible. Data and information will be accompanied by supporting documentation (metadata) that provide context, value, utility, and longevity, thereby facilitating a broad understanding of ERMN program products for current and future end-users.

Data Defined

Natural resource data are the building blocks on which our evolving ecological understanding of park resources is built. The term "data" refers not only to tabular and spatial data, but also to other information recorded and stored alongside these data such as documentation. Data products fall into five general categories: raw data, derived data, documentation, reports, and administrative records (Table ES.1).

Table ES.1. Categories of data products.

Category	Examples
Raw data	GPS files, field forms and notebooks, photographs and sound/video recordings, telemetry or remote-sensed data files, biological voucher specimens etc.
Compiled/derived data	Relational databases, tabular data files, GIS layers, maps, species checklists, analyzed data
Documentation	Data collection protocols, data processing/analysis protocols, record of protocol changes, data dictionaries, FGDC/NBII metadata, data design documentation, quality assurance reports, catalogs of specimens/photographs etc.
Reports	Annual progress report, final report (technical or general audience), periodic trend analysis report, publications etc.
Administrative records	Contracts and agreements, study plans, research permit/applications, other critical administrative correspondences etc.

There are also data from a variety of sources including those that originate from the I&M program and those that do not. Our highest priority is to produce and curate high-quality, well-documented data originating from the I&M Program. As time and resources permit, we will work toward raising the level of data management for current projects, legacy data, and data originating outside the I&M Program.

Roles and Responsibilities

Data management can not be successful unless everyone takes it seriously; it is therefore *everyone's* responsibility. Data management is sometimes more about managing people than it is about managing data. A data manager by profession can provide the tools, education and encouragement to facilitate good data management, but it is primarily the actions of others that will dictate the quality of data management. Table ES.2 shows primary data management responsibilities by position in the ERMN.

Data Life Cycle and Workflow

There are two main types of projects in which the I&M program is involved: short-term and long-term. A primary difference between short and long-term projects is the importance of adhering to established detailed standards for long-term projects to ensure internal consistency and comparability through time. While the need to follow standards is still present for short-term projects, sometimes the cost of compliance will outweigh the benefits due to the scope, budget, and level of NPS influence over the project. 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.

Projects can be divided into five primary stages: planning and approval; design and testing; implementation; product integration; evaluation and closure (Figure ES.1.). Each stage is characterized by a set of activities carried out by staff involved in the project. Primary responsibility for these activities rests with different individuals according to the different phases of a project.

Table ES.2. ERMN staff data management roles and responsibilities.

Role	Primary responsibilities related to data management
Project Crew Member (Biotech GS-4/5)	Record and verify measurements and observations based on project objectives and protocols. Document methods, procedures and anomalies
Project Crew Leader (Biotech GS-7)	Supervise crew members to ensure their data collection and management obligations are met
Botanist (Database Manager)	Perform assigned level of data management, including data entry, conversion, and documentation
Network Data Manager	Provide overall network planning, training, and operational support for the awareness, coordination, and integration of data and information management activities
Plant Ecologist/Hydrologist	Ensure useful data are collected and managed by integrating natural resource science in network activities and products, including objective setting, sample design, data analysis, synthesis, and reporting
Park Resource Specialist	Make decisions about data with regard to validity, utility, sensitivity, and availability.
Park or Other Curator	Oversee all aspects of the acquisition, documentation, preservation, and use of park collections
Statistician/Biometrician	Analyze data and present information, make suggestions on systematic data anomalies
Network Coordinator	Ensure programmatic data and information management requirements are met as part of overall network business

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 we can model conceptually. Figure ES.2. models a typical data life cycle in the ERMN complete with relationships to national data servers and repositories.

Infrastructure and System Architecture

A complex system architecture and reliable computer infrastructure are necessary in order to support the data management program and its goals. The ERMN is in a unique situation in its duty-station at The Pennsylvania State University and relies on local university computer and hardware infrastructure on a daily basis to accomplish its mission. At the same time, ERMN staff also rely on data management system architecture developed and maintained at the national level of the I&M program for a variety of cataloging and data services. Data are generated, managed and maintained at the local network level and sometimes uploaded to the national level. Aspects of the local ERMN system architecture are still under development, however. Figure ES.3 shows generically how these two architectures will integrate with one and other.

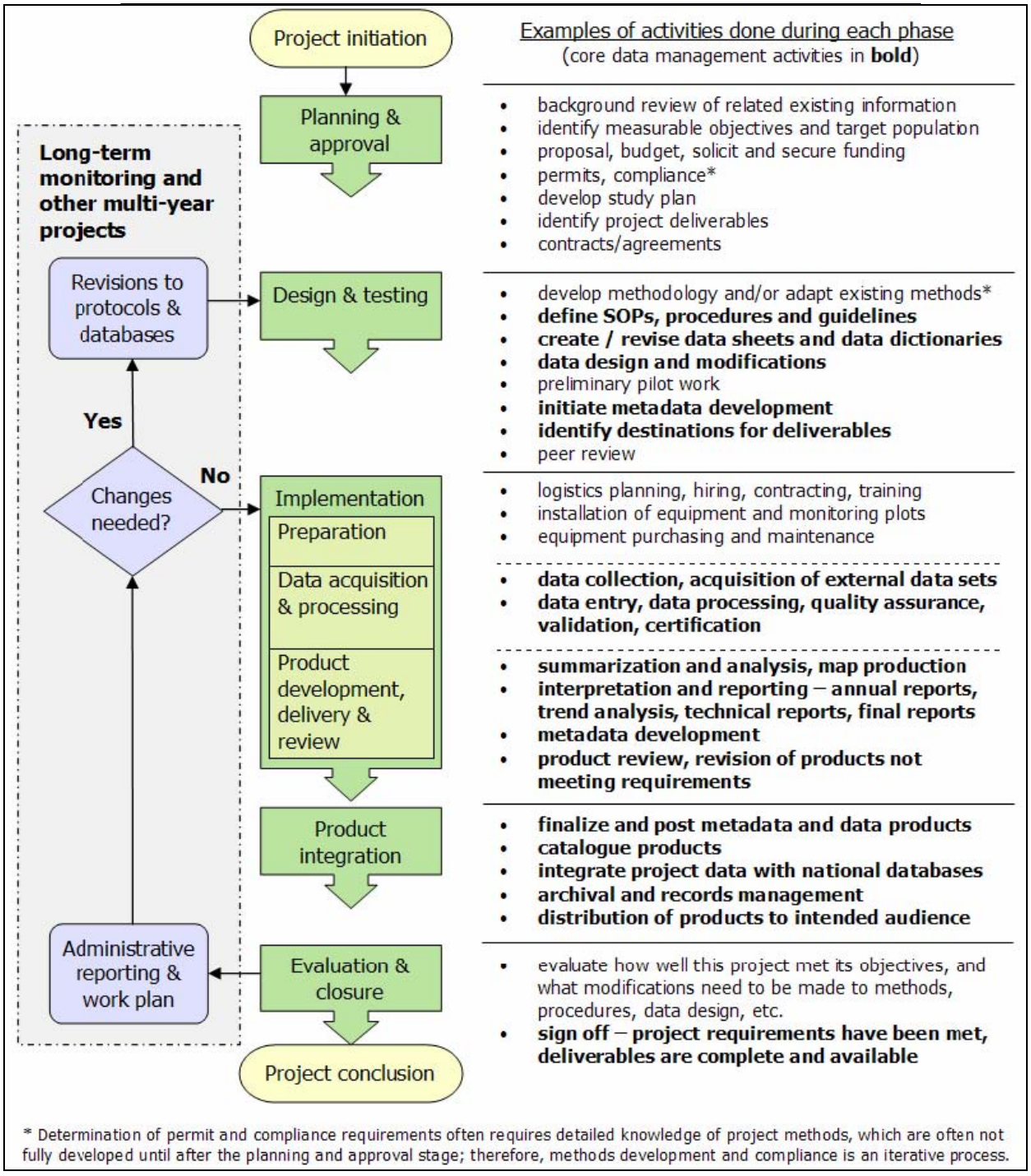

Figure ES.1. Project and data workflow.

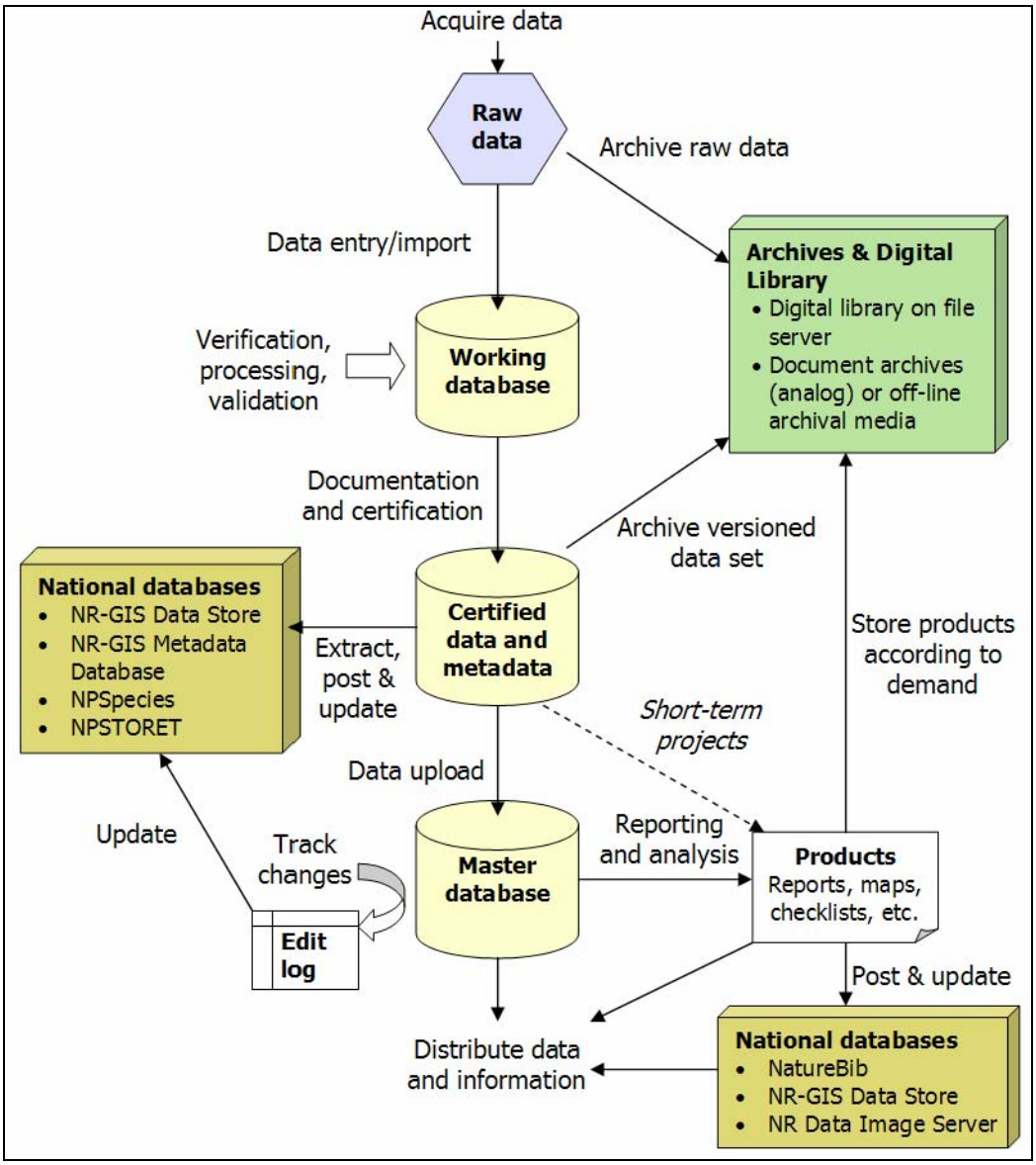

Figure ES.2. Diagram of typical data life cycle.

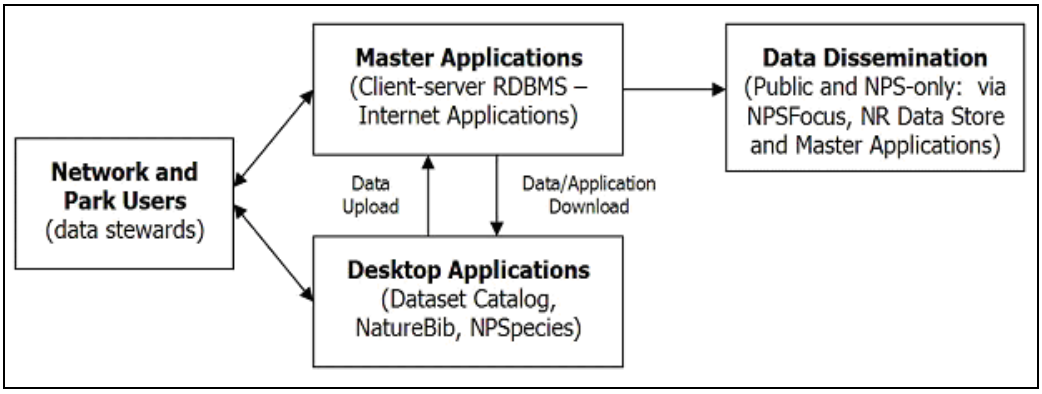

Figure ES.3. Sample conceptual model of I&M Program system architecture.

Data Acquisition and Processing

Steps for acquiring and processing natural resource data will vary depending on the data source, which can be placed into three general categories:

- Eastern Rivers and Mountains Network data: data resulting from projects that are initiated, sponsored, or funded by the ERMN
- Other National Park Service (NPS) data: data resulting from projects that are initiated, sponsored, or funded by park units, or by regional or national NPS programs
- External data: data produced or managed by agencies, organizations, or individuals other than the NPS

ERMN staff has the most control over data produced by and for the network. They will therefore ensure that data meet all specifications laid out by the network, program, NPS, DOI, and federal government. These standards can include those for documentation, cataloging, and dissemination. Data falling into the second category of "other NPS data" is often managed by the NPS entity most closely associated with the dataset. However, sometimes there are not the resources or expertise at these entities to manage the data properly. In these cases, the ERMN will evaluate the dataset's utility to the program or to natural resource management at an ERMN park and manage these data as needed. ERMN staff will take the following action where appropriate, for data falling into the third category of "external data":

Acquire data→Evaluate→Transform→Catalog→Make Available

Quality Assurance and Quality Control (QA/QC)

High quality data and information are mandated by national NPS directives and orders and are vital to the credibility and success of the I&M program. Quality can be defined as incorporating three key components–objectivity, utility, and integrity.

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

Utility refers to the usefulness of the information to its intended users from the perspectives of both the Network, perhaps most importantly natural resource managers, and the general public.

Integrity refers to the security of information which could entail protection from unauthorized access or revision to ensure that the information is not compromised through corruption or falsification.

Data quality is planned for throughout the entire data life cycle—from training, to data collection, entry and input, verification and validation, documentation, communication, review and all aspects of general operations. A variety of methods can be employed at each step along the way. As you apply more methods and move further along in the cycle, the more confidence you can have in the quality of your data as depicted in Figure ES.4.

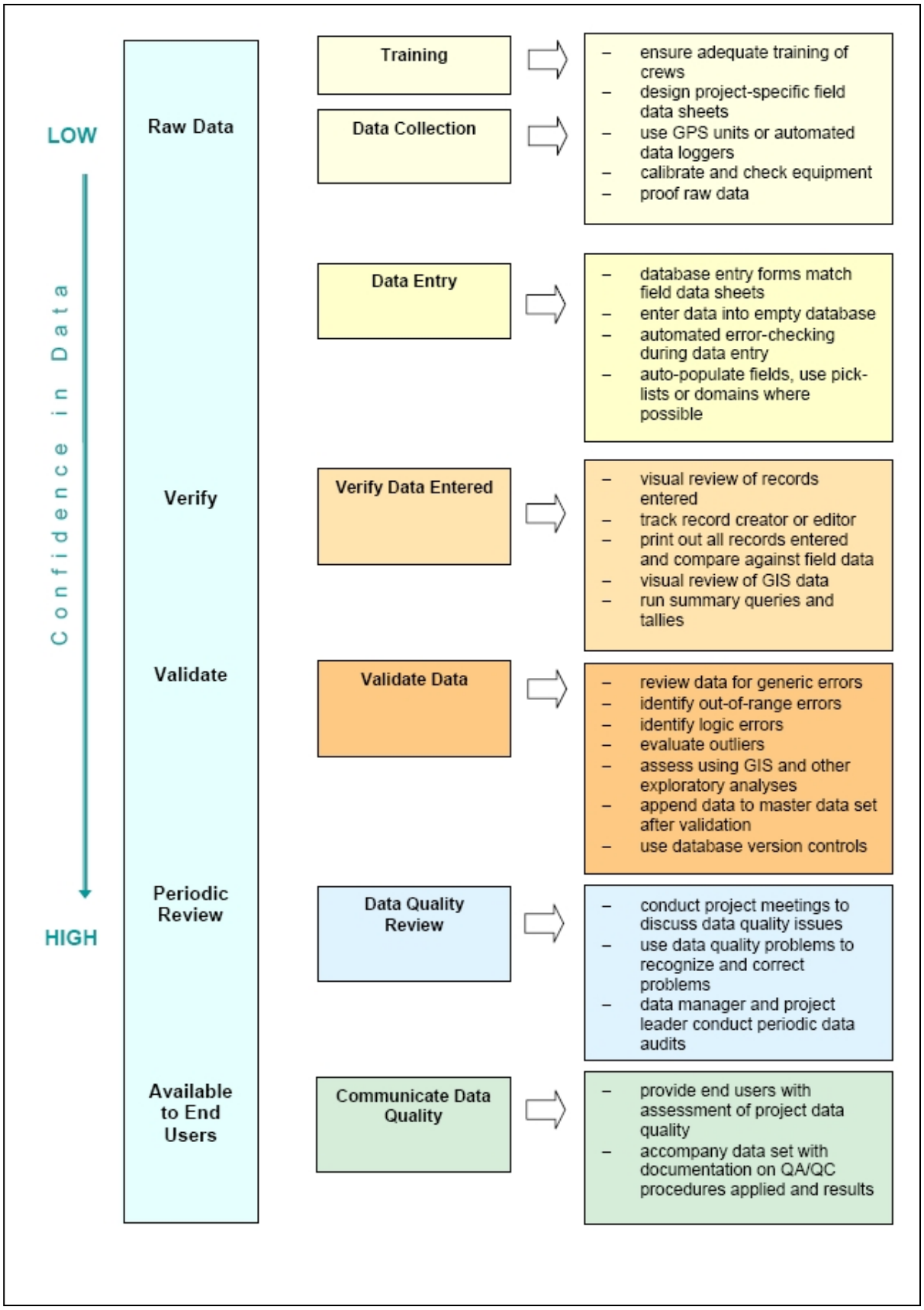

Figure ES.4. QA/QC controls and project workflow.

As you become more familiar with and continually review the quality of your data, the effectiveness of your QA/QC procedures will become evident. Nevertheless, a formal and coordinated review and revision of QA/QC procedures should take place every several years to ensure the quality of data generated in the future.

Data Documentation

Documenting data is the most important step toward ensuring that data sets are usable well into the future. Data longevity is directly related to the comprehensiveness of their documentation. There are a number of NPS and ERMN standards for documentation of spatial and tabular data, as well as projects, administrative records, vital signs protocols, and SOPs.

Overarching NPS standards include:

- Executive Order 12906, 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).
- FGDC CSDGM extensions, such as the Biological Data Profile, Remote Sensing Extension, and Shoreline Data Profile are not required, but recommended where appropriate.
- NPS Geographic Information System (GIS) Committee requires all GIS data layers be described with FGDC standards and the NPS Metadata Profile.

ERMN standards include:

- Projects will be documented and organized in their appropriate "Projects" folder using standard folder organization as described in chapter 4.
- Project finances will be tracked in the ERMN project tracking database as described in chapter 4.
- Spatial data files will be documented in compliance with FGDC standards using the CSDGM.
- Biological databases will be documented in compliance with FGDC standards using the NBII biological metadata profile.

Data documentation is perhaps the most undervalued data management activity. As such, it is the duty of the Network Data Manager to communicate, educate, and champion the importance of documentation, as well as sometimes to just make sure it gets done!

Data Dissemination

One of the most important goals of the Inventory and Monitoring Program is to *integrate natural resource inventory and monitoring information into National Park Service planning, management, and decision making*.

To accomplish this goal, procedures must be developed to ensure that relevant natural resource data collected by NPS staff, cooperators, researchers, and the public are entered, quality-checked, analyzed, documented, cataloged, archived, and made available for management decision-making, research, and education. Providing well-documented data in a timely manner to park managers is especially important to the success of the Program. The ERMN Program will make certain that:

- Data are easily discoverable and obtainable.
- No data that have not been subjected to full quality control are released.
- Distributed data are accompanied by complete metadata which clearly establishes the data as a product of the NPS I&M Program.
- Sensitive data are identified and protected from unauthorized access and inappropriate use.
- A complete record of data distribution/dissemination is maintained.

Data dissemination can be a complex task complete with considerations of data ownership, sensitivity, and access. All data and information generated by the I&M program should be property of the NPS and managed as such. This includes working closely with cooperators, contractors, experts, and park staff to assess sensitivity of data to public release and the potential for misuse. Data should be made accessible to the widest possible audience whenever prudent, providing for federal requirements to both protect (sensitive data) and release (according to the Freedom of Information Act) data.

Data Maintenance, Storage, and Archiving

Long-term maintenance and management of digital information are vital to the Inventory and Monitoring Program. This includes procedures for maintaining and managing digital data, documents, and objects that result from Network projects and activities. These procedures will help ensure the continued availability of crucial project information and permit a broad range of users to obtain, share, and properly interpret that information.

The ERMN has entered into a long-term agreement with North Carolina State University to serve as our primary digital archive. This archive, along with a variety of ERMN and I&M national program resources, make up the various repositories that store and safeguard ERMN products for the future. Table ES.3 shows the various repositories and the ERMN products that can be found therein.

Feedback and Revisions

The ERMN Data Management (DM) Plan (Piekielek 2006) outlines, in detail, the above strategies that have been and will be adopted by the ERMN. Appendixes to the DM Plan provide support and instruction on policies, specifications, and manuals. Whereas this chapter and the DM Plan outline general concepts and approaches, data management SOPs associated with each vital signs monitoring protocol will provide detailed step-by-step instructions on specific data management procedures and methods for long-term monitoring projects. The DM Plan, appendixes, and SOPs work in tandem to present and document the evolving ERMN data management program.

Table ES.3. Principal repositories for project products.

Project Products	Repository
Tabular Data and Documentation	
Final Products	NCSU Digital Archive
	ERMN Physical Archive (on media)
	NPS Data Store
	NPSpecies
Draft Products	ERMN Physical Archive (Projects Folder on media)
Spatial Data Final Products	 NCSU Digital Archive ERMN Physical Archive (on media)
	NPS Data Store
Draft Products	ERMN Physical Archive (Projects Folder on media)
Technical Reports digital	 NPS Data Store, NPS Focus, NERO Science Website, E-TIC
hard copy	ERMN library
	Park library
	ERMN physical archive TIC
bibliography Photographs Aerial hard copy Aerial digital Project Related	NatureBib NCSU physical archive NCSU digital archive NCSU digital archive ERMN physical archive (on media)

Revisions to this chapter and its associated ERMN DM Plan will be made periodically or as needed, every several years. The Network Data Manager is responsible for initiating and coordinating revisions. Questions and comments about this chapter and the DM Plan are welcome, as are suggestions on how to make it better, more clear, or more readable. Feedback should be directed to the Network Data Manager.

Efficient and effective data management methods, strategies, and procedures are vital to meeting I&M goals and maintaining programmatic support and relevance well into the future.

Acknowledgements

The majority of material in this document has been developed by I&M data managers across the country working together and sharing ideas over the course of the last several years. At the end of each chapter I try to acknowledge significant authors and networks that were responsible for the development of information therein, though it is sometimes difficult to credit all people involved since there have been iterations of improvements made through the years. A great deal of credit goes to the founders of the I&M program for having the foresight to emphasize the importance of data management to this program and to both Lisa Nelson and Margaret Beer for their efforts to organize data managers and data management planning efforts throughout the early years of this program.

The Inventory and Monitoring Program was made possible through the National Park Service Natural Resource Challenge.

Chapter 1. Introduction

The Inventory and Monitoring Program (I&M; www.nature.nps.gov/im) represents a long-term commitment by the National Park Service (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 Omnibus Management Act (1998) charges the Secretary of the Interior to:

> *"continually improve the ability of the National Park Service to provide state-of-the-art management, protection, and interpretation of and research on the resources of the National Park System,"* and to *"… assure the full and proper utilization of the results of scientific studies for park management decisions."* Section 5934 of the Act requires the Secretary of the Interior to develop a program of *"inventory and monitoring of National Park System resources to establish baseline information and to provide information on the long-term trends in the condition of National Park System resources."*

To carry out this mission, the NPS initiated a servicewide natural resource Inventory and Monitoring Program encompassing approximately 270 park units with significant natural resources. These park units are grouped into 32 networks based on similar geography and ecology. Each network works towards the following goals of the I&M Program:

- 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; and
- Share National Park Service accomplishments and information with other natural resource organizations and form partnerships for attaining common goals and objectives.

The last two of these goals can only be achieved through the development of a modern information management infrastructure and good data management procedures. Data collected by NPS staff, cooperators, researchers, and others must be entered, quality-checked, analyzed, reported, archived, documented, cataloged, and made available to others for management decision-making, research, and education. This Data Management Plan serves as the overarching strategy for achieving these goals. The plan supports the Eastern Rivers and Mountains Network Monitoring Plan (Marshall et al. 2006).

1.1. Eastern Rivers and Mountains Network

The Eastern Rivers and Mountains Network (ERMN) was created by charter in October 2002 and has a mission to inventory and monitor important natural resources in a network of nine parks in Pennsylvania, West Virginia, New York, and New Jersey. Figure 1.1 depicts the nine parks which include Delaware Water Gap National Recreation Area, Upper Delaware National Scenic and Recreational River, Allegheny Portage Railroad National Historic Site, Johnstown Flood National Memorial, Fort Necessity National Battlefield, Friendship Hill National Historic Site, New River Gorge National River, Gauley River National Recreation Area, and Bluestone National Scenic River.

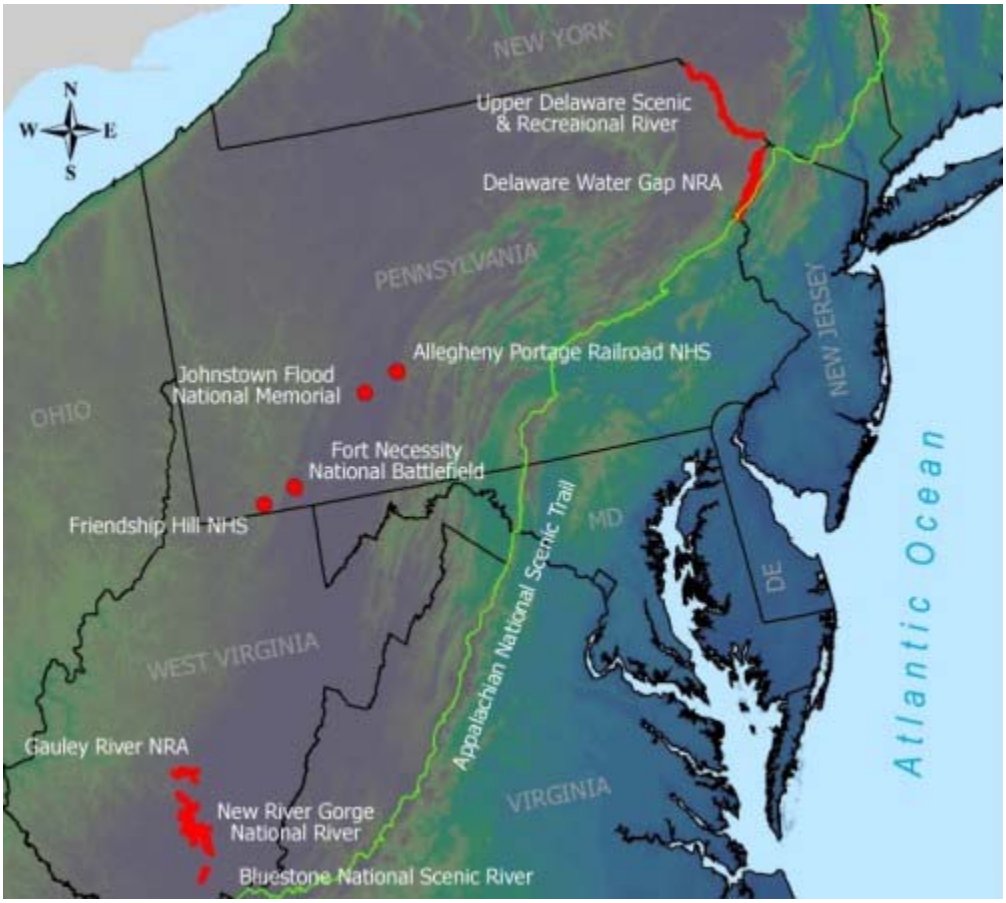

Figure 1.1. Map of ERMN parks.

1.2. ERMN Data Management Goals and Objectives

The overarching goal of the ERMN data management system is to provide timely and usable scientific information about the status and trends of park resources to park managers and others. The success of our program hinges upon our ability to produce, manage, and deliver this information to its intended audience. This system should ensure that natural resource information is available, usable, shareable, integrated, and interpreted:

- Available – We have the means of knowing that the data exist. Data are accessible and easily located.
- Usable – Data are stored in a stable, reliable, and interpretable data retrieval system.
- Shareable – Data products are complete, certified for quality assurance, screened for sensitive information, formatted for use, and documented for interpretation by others.
- Integrated – Data products are consistent with data exchange standards, interoperable with related natural resource data sets, and are collected and stored in a way that optimizes the tradeoff between meeting local needs and achieving compatibility with other agencies and partners.
- Interpreted – Data have been reviewed, summarized, and transformed into useful information.

1.3. Data Defined

Natural resource data are the building blocks on which our evolving ecological understanding of park resources is built. The term "data" refers not only to tabular and spatial data, but also to other information recorded and stored alongside these data, such as documentation. Data products fall into five general categories: raw data, derived data, documentation, reports, and administrative records (Table 1.1).

Table 1.1. Categories of data products.

Category	Examples
Raw data	GPS files, field forms and notebooks, photographs and sound/video recordings, telemetry or remote-sensed data files, biological voucher specimens, etc.
Compiled/derived data	Relational databases, tabular data files, GIS layers, maps, species checklists, analyzed data
Documentation	Data collection protocols, data processing/analysis protocols, record of protocol changes, data dictionaries, FGDC/NBII metadata, data design documentation, quality assurance reports, catalogs of specimens/photographs, etc.
Reports	Annual progress reports, final report (technical or general audience), periodic trend analysis report, publications, etc.
Administrative records	Contracts and agreements, study plans, research permit/applications, other critical administrative correspondences, etc.

Each data category listed in Table 1.1 requires some level of management in order to place data products in their appropriate context and preserve their meaning for the future. This requires a holistic view of how natural resource data are generated, processed, finalized, and made available. Such an understanding allows us to plan for an efficient and effective data management strategy.

1.4. Sources of Natural Resource Data

There are many potential sources of important information about the condition of natural resources in our parks. The types of work that may generate natural resource data about park resources include:

- Inventories;
- Monitoring;
- Protocol development pilot studies;
- Special focus studies;
- External research projects;
- Monitoring or research studies done by other agencies on park or adjacent lands;
- Resource impact evaluations related to park planning and compliance with regulations; and / or
- Resource management and restoration work.

Our highest priority is to produce and curate high-quality, well-documented data originating from the I&M Program. As time and resources permit, we will work toward raising the level of data management for current projects, legacy data, and data originating outside the Inventory and Monitoring Program as follows:

1. We will place greatest emphasis on projects that are just beginning to be developed and implemented, because inserting good data management practices from the very beginning is the most effective strategy.
2. Updating data management for ongoing projects and conversion of legacy data will be prioritized according to the value of the information they provide.

Because the I&M Program focuses on long-term monitoring and natural resource inventories, our first priority should be the data and information that we derive from these primary efforts. However, we can easily apply the same standards, procedures, infrastructure, and attitudes about data management to other natural resource data sources. High-profile data sets that provide crucial information to park management will be prioritized for data management regardless of funding source.

This plan covers four major categories of data that are coordinated or managed by the ERMN (see chapter 4 for more information). These are:

1. Data managed in service-wide databases. These databases are continually evolving, but at their core include: NPSpecies, NatureBib, and the NPS Data Store.
2. Data developed or acquired directly by the network as a result of inventory, monitoring, or other projects. This category includes project-related protocols, reports, spatial data, and associated materials such as field notes and photographs provided to the ERMN by contractors or developed by ERMN staff. Projects can be short-term (i.e., one to two years duration) or long-term (e.g., ongoing monitoring).
3. Data that, while not developed or maintained by ERMN, are used as data sources or provide context to other data sets. Examples of this category include: base GIS data developed by parks other agencies or organizations; national or international taxonomic classification systems; climate or hydrologic data collected by regional or national entities; data and products developed under the auspices of a research permit.
4. Data acquired and maintained by network parks that the ERMN assists in managing. Because of the lack of data management expertise in many network parks, the ERMN provides data management assistance for high-priority data sets or those that may benefit from standardized procedures. Examples include park observations databases for wildlife and rare plant data, legacy or current park natural resource monitoring data sets, and data on exotic invasive plant species.

1.5. ERMN Data Management Plan Scope and Organization

The Eastern Rivers and Mountains Network Data Management Plan outlines how the network intends to implement and maintain a system that will serve the data and information management needs of our Inventory and Monitoring Program. This plan reflects our commitment to establishing and maintaining a robust system for data management to ensure the availability and usability of high-quality natural resource information.

The ERMN Data Management Plan describes how our Network will:

- support I&M Program objectives;
- acquire and process data;
- assure data quality;
- document, analyze, summarize, and disseminate data and information;
- maintain nationally developed data management systems; and
- maintain, store, and archive data.

We have developed this data management plan in modular format in the hope of increasing readability and usability among a wider audience. This format allows users to more easily locate and access subdocuments pertaining to a particular element of information management guidance or process. Individual procedural documents within the plan can be provided as standalone documents to park and regional data management staff, cooperators, and project

leaders. This format also allows for a more efficient review, as technical experts can be provided with applicable portions of the plan based on their expertise (GIS, databases, metadata, GPS, etc.). There are three document types that make up this plan:

- ERMN Data Management Plan;
- Guidance Documents (appendixes); and
- Standard Operating Procedures (SOPs).

The body of this plan is composed of chapters that provide overview information, whereas more specific information, including procedures, is found in guidance documents and SOPs associated with specific monitoring (or other) protocol, many of which are included as appendixes to this plan.

1.6. Intended Audience

This plan is intended for everyone. It's written so as to be accessible to a very broad and sometimes nontechnical audience. It is for anyone who wants to understand more about the data management philosophy, approach, and strategies employed by the ERMN. The I&M Program has a complex data management framework, complimented by a variety of different network level solutions, creating an often difficult matrix of relationships to understand. The body of the plan is general in its treatment of many program topics, but specific enough to set the stage for the more detailed and technical complimentary documents.

1.7. Revisions

Feedback, comments, and questions about this plan and its associated documents are always welcome and should be directed to your friendly local Network Data Manager. The body of this plan is meant to be relatively static and, as such, will be reviewed and revised only once every several years, or as deemed necessary. The Network Data Manager will be responsible for initiating revisions to the body of this document. Complimentary documents, some of which are currently in draft form, will be much more dynamic and will be reviewed and revised on a shorter (likely annual) basis by the Data Manager in cooperation with Project Leaders (in the case of monitoring protocol SOPs) and others (in the case of other supporting documents). This plan is a living document that will necessarily change throughout time if it's meant to remain relevant.

1.8. Credits

Materials and information presented in this chapter were adapted from Mark Hart and Ulf Gavert of the Great Lakes Monitoring Network (Hart et al. 2005), as well as from John Boetsch, Bret Christoe, and Ronald Holmes of the Olympic National Park prototype park, Mount Rainier National Park and North Cascades Network (Boetsch et al. 2005).

Chapter 2. Data Management Roles and Responsibilities

The success of any long-term ecological monitoring program depends, in part, on the skills and abilities of those working on the monitoring project to convert observations into data, and data into meaningful information, and then finally maintaining that information through time (i.e. perform sound data management practices). The importance of data management was expressed by the founders of the I&M program by hiring one professional position into each monitoring network to champion data management. What's more, the I&M program expects that at least one-third of network financial resources (including staff time) be dedicated to data management. In the ERMN, no individual referenced in the ERMN Monitoring Plan, with the exception of the Network Coordinator, is identified as having less than one-third of their time ear-marked to be spent on data management activities (Marshall et al. 2006).

The following chapter presents a generic description of roles and responsibilities of individuals working on a monitoring project in the ERMN. Each individual monitoring protocol will have more details and instruction (standard operating procedures) regarding data management roles and responsibilities for each monitoring activity and will appear as appendixes to this document when they are available.

2.1. Definition of Roles versus Responsibilities

The network considers a role as a function or position (e.g., Data Manager), while a responsibility is a duty or obligation (e.g., review data records). Data stewardship is not treated as a role because it involves all project staff sharing the responsibilities for managing data and information resources. Using the term 'Data Steward' seems appropriate only when a list of specific responsibilities is associated with every role in a given project or program. Establishing a single project data steward role is avoided because an unmanageable burden (real or perceived) is placed on a single position/person, and others with data management responsibilities are more likely to under-perform. As has already been stated and is reflected in percentages of each position's time dedicated to data management in the ERMN Monitoring Plan, data management is everyone's responsibility!

2.2. Roles and Responsibilities of Network Staff

The following should serve as an overview of each individual's data management role and responsibilities by position title and as they fit into the monitoring process. These are not, however, static, and will likely change across protocol and through time.

2.2.1. Roles and Responsibilities by Category

Managing data for multiple monitoring protocols involves timeframes of several years to decades and includes changes in information technology, turnover in staff, new scientific insights, and shifting priorities. This requires that every individual understand and perform data stewardship responsibilities in the production, analysis, management, and/or end use of the data. As coordinator of these activities, the fundamental role of the network data manager is to understand program and project requirements, create and maintain data management infrastructure and standards, and educate, communicate, and work with all responsible individuals. In practice, the

challenges of data management become more people-management challenges than technical challenges of relational databases and enterprise data solutions. Table 2.1 places each potential ERMN position identified in the Monitoring Plan into a category associated with their primary data stewardship responsibilities and serves as a good starting point for placing yourself and your data management responsibilities into the monitoring process.

2.2.2. Roles and Responsibilities by Position Title

An increasing demand for more detailed, higher quality data and information about natural resources and ecosystem functions requires a group of people working together to effectively 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 reasonably do alone. Knowledgeable individuals from scientific, administrative, and technological disciplines must work together to ensure that data are collected using appropriate methods, and that resulting data sets, reports, maps, models, and other derived products are well managed. Data sets and related products must be credible, representative, and available for current and future needs. A typical data set used for inventory and monitoring is handled, viewed, and stewarded by most or all of the people serving in the roles (positions) listed in Table 2.2. These roles are listed 'from the ground up' to help demonstrate the hierarchy and overlap of responsibilities. For example, a project leader is ultimately responsible for the activities listed in the field level roles of crew leader and crew member, and the network coordinator ensures that the network data manager and ecologist achieve the required performance level. With one-third of network resources devoted to overall data management, analysis, and reporting, the network intends to avoid cases in which a single position is assigned unattainable duties for multiple roles resulting in a lack of attention to one or more aspects of the project. Meeting this goal requires awareness and support from park managers to plan for, fund, and provide performance-level accountability for the resources required to adequately manage their park's data and information assets. Learning, understanding, and acting on the responsibilities listed for all categories of data stewardship will help prevent the critical aspects of data management from being overlooked or under-funded.

Table 2.1. Categories of data stewardship.

Stewardship Category	Related Activities	Principal jobs or positions
Note: Each position is listed in only one category according to overriding responsibilities. However, most positions contribute in each category.		
Production	Creating data or information from any original or derived source. This includes recording locations, images, measurements, and observations in the field, digitizing source maps, keying in data from a hardcopy source, converting existing data sources, image processing, and preparing and delivering informative products, such as summary tables, maps, charts, and reports	Project Crew Member (Biotech GS-4/5) Project Crew Leader (Biotech GS-7)
Analysis	Using data to predict, qualify, and quantify ecosystem elements, structure, and function as part of the effort to understand these components, address monitoring objectives, and inform park and ecosystem management	Plant Ecologist/Hydrologist Park Resource Specialist Consulting Statistician
Management	Preparing and executing policies, procedures, and activities that keep data and information resources organized, available, useful, compliant, and secure.	Network Data Manager Botanist (Database Manager) National I&M Data Manager Park or Other Curator
End Use	Obtaining and applying available information to develop knowledge that contributes to understanding and managing park resources. Providing feedback for improvements in data content and quality	Network Coordinator Park Managers and Superintendents Others

Table 2.2. Summary of roles and responsibilities by position.

Role	Primary responsibilities related to data management
Project Crew Member (Biotech GS-4/5)	Record and verify measurements and observations based on project objectives and protocols. Document methods, procedures, and anomalies.
Project Crew Leader (Biotech GS-7)	Supervise crew members to ensure their data collection and management obligations are met, including data verification and documentation.
Botanist (Database Manager)	Perform assigned level of technical data management, including data entry, data conversion, and documentation. Work on overall data quality and stewardship with project leaders, resource specialists, and the Network data manager.
Network Data Manager	Provide overall network planning, training, and operational support for the awareness, coordination, and integration of data and information management activities, including people, information needs, data, software, and hardware. Serve as Point of Contact for National Park Service database applications (NPSpecies). Coordinate internal and external data management activities, especially in collaboration with project leaders. Ensure that metadata is created to meet NPS standards and published on an NSDI clearinghouse.
Plant Ecologist/Hydrologist	Oversee and direct operations, including data management requirements, for one or more network projects. Maintain communication with project staff, Network Data Manager, and resource specialist regarding data management. Determine what others inside and outside the NPS are doing relative to the project and establish a minimum set of standard data items to promote comparability with other efforts. Ensure useful data are collected and managed by integrating natural resource science in network activities and products, including objective setting, sample design, data analysis, synthesis, and reporting.
Park Resource Specialist	Understand the objectives of the project, the resulting data, and their scientific and management relevance. Make decisions about data with regard to validity, utility, sensitivity, and availability. Describe, publish, release, and discuss the data and associated information products.
Park or Other Curator	Oversee all aspects of the acquisition, documentation, preservation, and use of park collections.
Statistician/Biometrician	Analyze data and present information, make suggestions on systematic data anomalies.
Network Coordinator	Ensure programmatic data and information management requirements are met as part of overall network business .
I&M Data Manager (National Level)	Provide servicewide database design, support, and services, including receiving and processing to convert, store, and archive data in servicewide databases.
Other End Users	These 'information consumers' include park managers and superintendents, researchers, staff from other agencies, and the public. End users are responsible for informing the scope and direction of science information needs and activities, applying data and derived products appropriately, and providing feedback for improvements. Park superintendents and program managers, in particular, are responsible for providing the money and scheduling the time required for project leaders and staff to meet stated data quality and other data management objectives.

2.2.3. Data Manager Roles and Responsibilities

The data manager is the champion and coordinator of data management activities within the network. They are the salesperson of sound data management principles and concepts to the other staff in the network and their ability to convince others of the importance of data management will be directly reflected in the quality of data and information produced by the network. More specifically, data managers will:

- develop and maintain the infrastructure for metadata creation, project documentation, and project data management;
- create and maintain project databases in accordance with best practices and current program standards;
- provide training in the theory and practice of data management tailored to the needs of project personnel;
- develop ways to improve the accessibility and transparency of data and information; and
- establish and implement procedures to protect sensitive data.

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

- define the scope of project data and create a data structure that meets project needs;
- become familiar with how data are collected, handled, and used;
- review quality control and quality assurance aspects of project protocols and standard procedure 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, as well as other automated "push-button" procedures, developed as needed;
- create a user interface that streamlines the process of data entry, review, validation, and summarization that is consistent with the capabilities of project staff;
- develop automated database procedures to improve the efficiency of the data summarization and reporting process;
- make sure that project documentation is complete, complies with metadata requirements, and enhances the interpretability and longevity of the project data;
- ensure regular archival of project materials; and
- inform project staff of changes and advances in data management practices.

The network data manager is by no means responsible for performing all, or even the majority, of data management tasks—they are responsible for ensuring that all tasks are done well, in an efficient manner, and in accordance with established standards.

2.2.4. Project Leader Roles and Responsibilities

For ease of presentation, project leaders in the ERMN are assumed to be either the Plant Ecologist or the Hydrologist positions identified in the ERMN Monitoring Plan. In actuality though, the Network Coordinator, Data Manager, and Botanist will all likely serve as project (or protocol) leads for the ERMN, as well. The following can be expected of each individual while serving as a Project Leader no matter what their position title.

The project leader is accountable for data quality during all phases of the project, including collecting, entering, handling, reviewing, summarizing, and reporting data. Developing project documentation and metadata are crucial elements of the project leader's role. Specifically, the role of the project leader includes the following:

- Complete project documentation describing the who, what, where, when, why, and how of a project;
- Develop, document, and implement standard procedures for field data collection and data handling;
- Enact and supervise quality assurance and quality control measures for the project;
- Supervise and certify all field operations, including staff training, equipment calibration, species identification, and data collection;
- Supervise or perform data entry, verification, and validation;
- Maintain concise explanatory documentation of all deviations from standard procedures;
- Ensure documentation of important details for each field data collection period;
- Maintain hard copies of data forms and send original data forms to archive on a regular basis;
- Work with program coordinators to identify analysis and reporting mechanisms and to establish a schedule for regular project milestones, such as data collection periods, data processing target dates, and reporting deadlines;
- Produce regular summary reports and conduct periodic trend analysis of data, store the resulting reports, and make them available to users; and
- Act as the main point of contact concerning data content.

In cases where the data manager is not the project leader (which is assumed to be most cases, with the possible exception of the ERMN Landscape Dynamics Monitoring Protocol), the project leader will work with the data manager to do the following (depicted graphically in Figure 2.1):

- Develop quality assurance and quality control procedures specific to project operations;
- Identify training needs for staff related to data management philosophy, database software use, quality control procedures, etc;
- Coordinate changes to the field data forms and the user interface for the project database;
- Fully document and maintain master data;
- Identify sensitive information that requires special consideration prior to distribution;
- 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 how project data will be transformed from raw data into meaningful information and create data summary procedures to automate and standardize this process;
- Identify and prioritize legacy data for conversion; convert priority data sets to a modern format; and
- Increase the interpretability and accessibility of existing natural resource information.

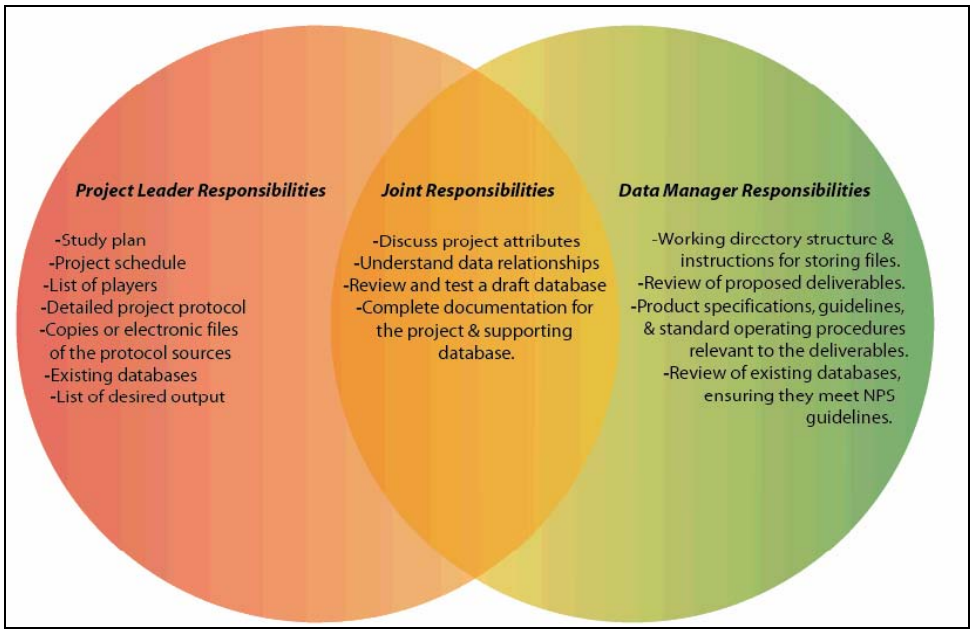

Figure 2.1. Intersection of data management responsibilities

2.3. Coordination

As should already be clear to the reader based on above sections, data management coordination among network staff is imperative. Given common goals, supervision, and source of resources, this should be an achievable goal; however, there are others with whom network data management should be coordinated.

2.3.1. Coordination with Park and Regional Staff

The network data manager works with national Inventory and Monitoring Program data management staff and regional resource information management personnel to maintain awareness and involvement in servicewide and regional databases and data management policy and guidance. The network data manager works locally with network personnel, park staff, and cooperators to promote and develop workable standards and procedures that result in integration and availability of data sets.

Key contacts for the network data manager include park GIS and data managers and the project leaders for each monitoring or inventory project. Regular and productive communication among these personnel leads to common understanding and better synchronization of network and park data management activities. Park and network staff coordinate on resource information management using a variety of methods, including personal visits, phone calls, email, joint meetings and training sessions, as well as the meetings and work of the network's Technical Steering Committee, Science Advisory Committee, and Board of Directors. The development of network planning materials, including inventory study plans and monitoring protocols, includes involvement and input from park scientists and resource information management staff.

In the Northeast Region (NER) of the NPS there is substantial data management (and other) coordination among the four I&M networks. Staff of other NER I&M networks, as well as

regional staff of the Philadelphia Science Office (PHSO), should be looked to as a data management resource and kept abreast of major data management changes and decisions.

2.3.2. Coordination with Other I&M Program Data Management Staff

Thirty-two Network data managers have been hired as part of the Inventory and Monitoring Program across the country, in addition to a National Program Lead Data Manager and a host of other data management staff located in Fort Collins, CO. I&M Program data management staff regularly coordinate with each other via annual meetings, conference calls, workgroups, a listserv, Web sites, and informal communications. Data managers share their workload by working closely with one another when appropriate. This model of cooperation and communication has been very effective; a data manager should never assume that the problem with which they are faced is unique or novel, because the chances are good that a peer has faced the exact same problem already and has developed an effective solution. The challenge becomes identifying the correct person to contact and networking. Routine face-to-face contact between data managers through meetings, workshops, and trainings is imperative if networking data management across the country is going to continue to succeed.

2.3.3. Coordination with Others

No data management system is flexible enough so-as to be compatible with all other data management systems. For this reason, data management should be coordinated amongst people who would like to be able to share data today and into the future. What's more, we are all often faced with the same problems, and you may just find that another organization or agency has come up with a more elegant or efficient solution to a problem with which you are faced. Though it falls relatively low on the priority list of tasks, data management coordination with others should be pursued whenever possible and this responsibility will, at times, fall on everyone within the network.

There are many organizations external to the NPS with whom the ERMN would like to share data, including state Departments of Environmental Protection, Natural Heritage Programs, and Breeding Bird Atlases, etc. However, one of the most important may be the Delaware River Basin Commission (DRBC) (http://www.state.nj.us/drbc/). The DRBC was established in 1961 to protect water quality and aquatic habitats along roughly 330 miles of the Delaware River, including the major resource of Delaware Water Gap National Recreation Area (DEWA) and the Upper Delaware Wild and Scenic River (UPDE). They have their own monitoring network and the regulatory teeth entrusted to them by President Kennedy and the governors of the four states involved when they signed the legislation to create what became the DRBC.

2.4. Documentation

If there is one aspect of data management responsibility that falls securely on the shoulders of one individual, it's in data documentation. The most reasonable person to document any data set is the creator of that data set. Data documentation establishes the basis for the appropriate use of data in resulting analysis and products, both short term and long term. More information is provided in the Data Documentation Section of this document; however, specific procedures to

document data will all be found in appendixes, protocols, and data management standard operating procedures.

2.5. Awareness and Accountability

Successful data stewardship requires that people involved in network activities learn and understand the expectations for continuous data management and are held accountable to perform the duties required to meet these expectations. This is equally important for network staff, park employees, and contractors or cooperators. All project or monitoring participants should receive training, briefings, materials, and additional regular communication about data stewardship from supervisors, project leaders, and data managers. The purpose is to promote the appropriate level of understanding about how their efforts relate to park and network management objectives, National Park Service and Department of Interior policies, and other federal government requirements. Other relevant context and linkages can also be discussed to help establish a sense of ownership and accountability among project staff; it is important that people understand why they are spending time doing data management activities. Inventory and Monitoring project leaders should have a good understanding of resource information management issues and requirements and an awareness of the challenges and limitations associated with field data collection, including the use of GPS technology. This understanding can be achieved through training, detailed and regular briefings, and accompanying field crews to perform data collection at reasonable intervals. It is the network data manager's responsibility to champion data management awareness and accountability within the network; however, they need the support of all network staff (especially the coordinator and other managers) to achieve this.

2.6. Communication

Always the "wild card" where coordination is concerned, communication is key. At the network level, staff should discuss data management almost as often as rerun episodes of Seinfeld; it should become a popular topic of communication, so effortless that most will not even realize they are discussing data management. In a staff meeting model, I don't envision an agenda item titled "Data Management," but rather it is an issue that permeates all other topics. In a long-term ecological monitoring program it is difficult to partition data management into its own little cubicle, because as this chapter has hopefully demonstrated, it is something that's done by almost everyone at almost every step along the path of monitoring. It is the Network Data Manager's responsibility to personalize data management, to provide the vocabulary to others and facilitate communication along the way. In my limited experience, more than any other aspect of natural resource study and research, data management could be dramatically improved if people would simply talk about it, plan for it, and ask questions.

2.7. Credits

Much of the material presented in this section was adapted from Rob Daley and the Greater Yellowstone Network Data Management Plan (Daley 2005).

Chapter 3. Data Life Cycle and Work Flow

This chapter considers the general workflow of projects that produce natural resource data and gives an overview of how natural resource data are generated, processed, finalized, and made available. Data management activities that relate to the various stages of a project are highlighted. By describing the progressive stages of a project and the life cycle of the resulting data, we can more easily communicate the overall objectives and specific steps of the data management process. In addition, this awareness helps us to manage the staffing resources and infrastructure required to produce, maintain, and deliver quality data and information.

3.1. Project Workflow

There are two main types of projects in which the I&M program is involved:

- Short-term projects, which may include individual park research projects, inventories, or pilot work done in preparation for long-term monitoring.
- Long-term projects, which will mainly be the implemented monitoring projects central to the I&M program, but which may also include multi-year research projects and monitoring performed by other agencies and cooperators. Long-term projects will often require a higher level of documentation, peer review, and program support.

A primary difference between short- and long-term projects is the importance of adhering to standards for long-term projects to ensure internal consistency through time. While the need to follow standards is still present for short-term projects, sometimes the cost of compliance will outweigh the benefits due to the scope, budget, and level of NPS influence over the project. Nevertheless, both short-term and long-term projects share many workflow characteristics and both generate data products that must be managed and made available.

Projects can be divided into five primary stages: planning and approval; design and testing; implementation; product integration; evaluation and closure (Figure 3.1). Each stage is characterized by a set of activities carried out by staff involved in the project. Primary responsibility for these activities rests with different individuals according to the different phases of a project.

3.1.1. Planning and Approval

At this stage of the project many of the preliminary decisions are made regarding project scope and objectives. In addition, funding sources, permits, and compliance are all addressed. 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 (see Appendix A and B for detailed descriptions of Northeast Region I&M product specifications).

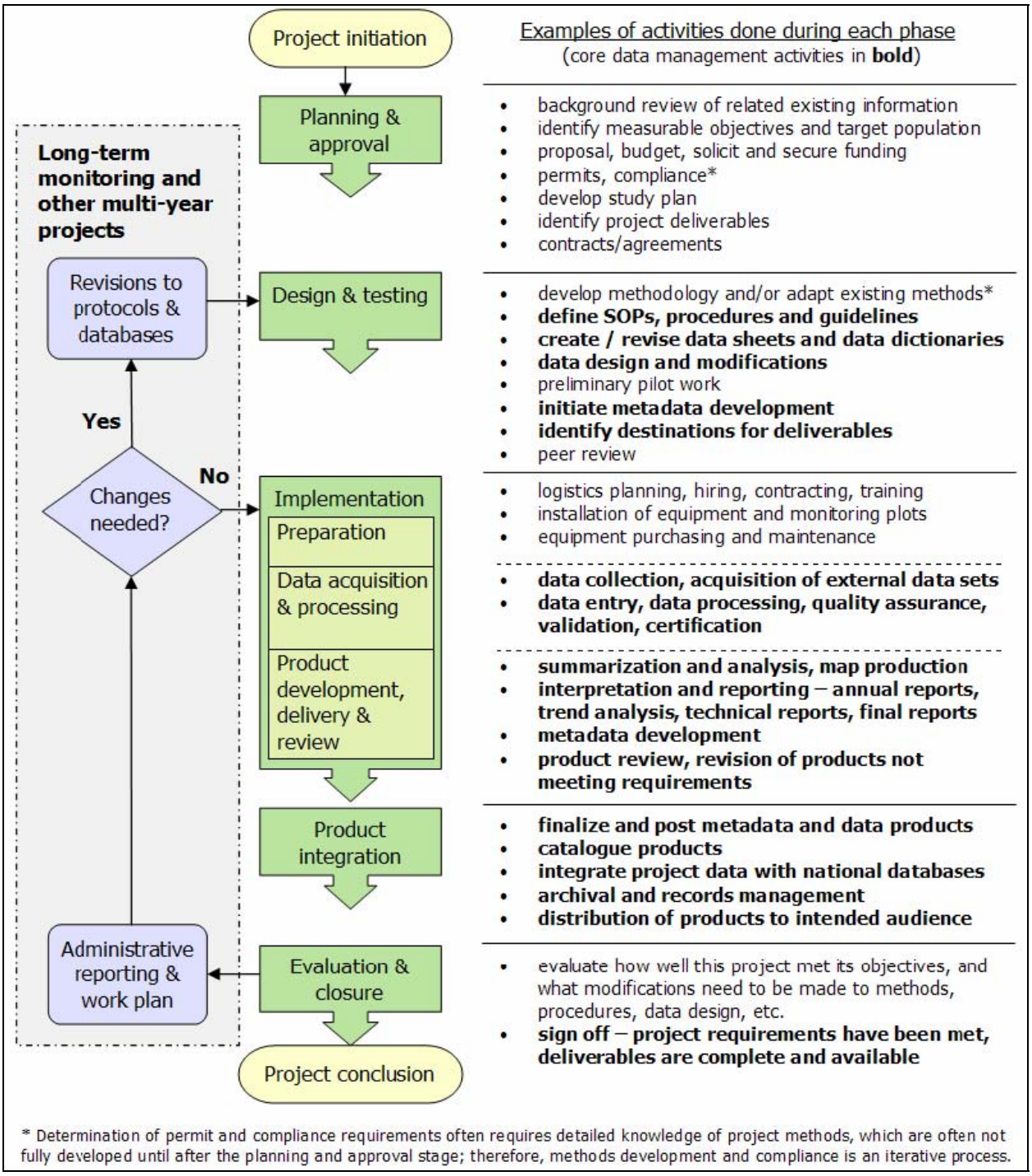

Figure 3.1. Conceptual model of project workflow.

3.1.2. Design and Testing

During this phase, all of the details are worked out 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 to 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 the development of the data design and data dictionary, where the specifics of database implementation and parameters that will be collected 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 it is likely that additional changes will need to be made 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.

3.1.3. Implementation

During the implementation phase, data are acquired, processed, error-checked, and documented. This is also when products such as reports, maps, GIS themes, and other products are developed and delivered. The project leader oversees all aspects of implementation – from logistics planning, contracting, 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 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 is possible, these roles should be worked out in advance of implementation.

Toward the end of this phase, project staff members work to develop and finalize the deliverables that were identified in the project planning documents (i.e., protocol, study plan, contract, agreement, or permit). See Appendixes C through E for descriptions of the product review process for inventory and vegetation map products as currently implemented in the NER, complete with roles, responsibilities, and data review procedures. In general, all raw and derived data products, metadata, reports, and other documentation should be delivered to the data steward assigned to the project. Administrative records should be delivered to appropriate park and network staff as specified. All project deliverables should be developed and delivered according to product specifications, which should be stipulated in all protocols, contracts, agreements, and permits. Products that do not meet program requirements will be returned for revision in an iterative process. As has already been mentioned, however, sometimes the cost of enforcing compliance on short-term projects is not worth staff time. Occasionally, network data management staff will make revisions to data products themselves when they feel this would be the most efficient way to modify product deliverables.

3.1.4. Product Integration

During this phase, data products and other deliverables are integrated into national and network databases, metadata records are finalized and posted to clearinghouses, and products are distributed or otherwise made available to their intended audience. Another aspect of integration is merging data from a working database to a master database maintained on the network server or other enterprise level data solution. This occurs only after the annual working data set has been certified for quality by the project leader. Certain projects may also have additional integration needs, such as when working jointly with other agencies on a common database.

Product integration includes creating records for reports and other project documents in NatureBib, posting imaged documents to the appropriate repositories, posting metadata records that have been completed and submitted by project leaders, and updating NPSpecies to reflect any new species occurrence information derived from the project. This will allow the information from the project to be searchable and available to others via servicewide search engines.

3.1.5. Evaluation and Closure

Upon project closure, records are updated to reflect the status of the project and its associated deliverables in a network project tracking application. In the ERMN, this database tracks only funds, not deliverables, but you can at least see that the last invoice (which is typically withheld until all deliverables are considered to be final) has been paid. 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, and reporting on, the project. For non-cyclic projects, this phase represents the completion of the project. After products are catalogued and made available, program administrators, project leaders, and data managers should work together to assess how well the project met its objectives and to determine what might be done to improve various aspects of the methods, implementation, and formats of the resulting information. For monitoring protocols, careful documentation of all changes is required. Changes to methods, SOPs, and other procedures should be maintained in a tracking table associated with each document. Major revisions may require additional peer review.

3.2. Data Life Cycle

During various phases of a project, 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 we can model to facilitate communication and aid in planning efforts. 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 the detailed information about project data needed for proper use and interpretation (see Chapter 7).
- Master database – Project-specific database for storing the full project data set, 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 copies of original field forms, data, metadata, and other products generated by projects. Archives are for hard-copy items and off-line storage media; whereas the digital library is maintained on a server or other hard drive.

Although the data life cycle may vary depending on specific project needs and objectives, a typical data life cycle proceeds as follows (depicted graphically in Figure 3.2):

1. Acquire data – For data recorded by hand in the field, data forms should be reviewed regularly (at least daily) for completeness and validity in order to capture errors as close to their origin as possible.
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. Archiving or scanning of hard copy data forms may occur at the end of a season as a means of retaining all marks and edits made during the verification and validation steps.
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 set. Certification is a confirmation by the project leader that the data have passed all quality assurance requirements and are complete and documented. It also means that data and metadata are ready to be posted and delivered.
6. Archive versioned data set – 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).
7. Extract data, post data, and update national databases – To make data available to others, certified data and metadata are posted to national repositories such as the NPS Data Store. In addition, national databases such as NPSpecies and NPSTORET are updated

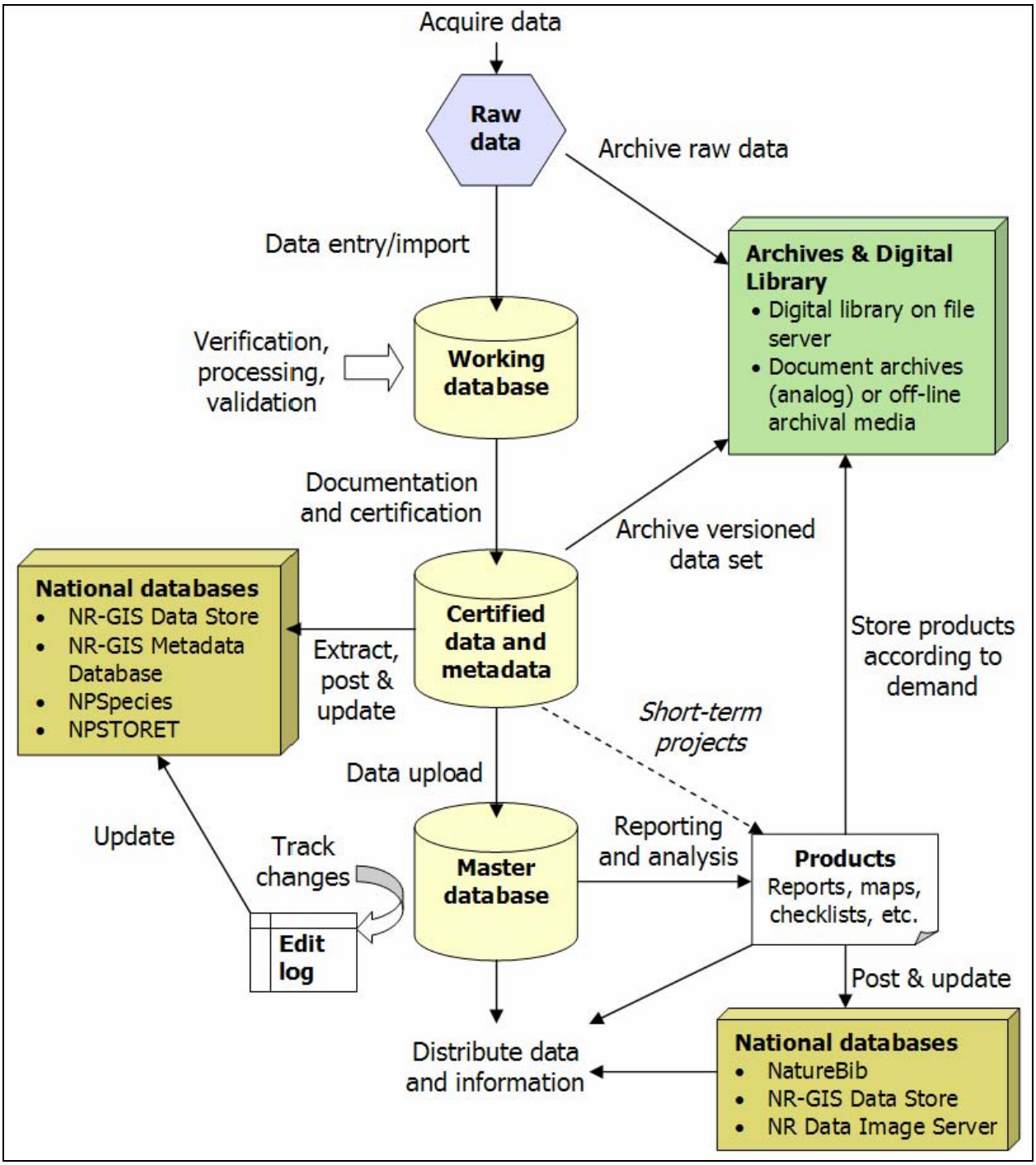

Figure 3.2. Diagram of typical project data life cycle.

with data extracted from the certified data set. Note that, at this time, data and data products may not be posted if they contain protected information about the nature or location of rare, threatened, or endangered species, or other natural resources of management concern (see Appendix F for NPSpecies data entry standards).

8. Upload data – Certified data are uploaded from the working database to a master project database. This step might be skipped for short-term projects where there is no need to distinguish working data for the current season from the full set of certified project data.
9. 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.
10. Store products – Reports and other data products are stored according to format and likely demand – in the digital library, on off-line media, or in the physical archives.
11. Post products and update national databases – To make data available to others, reports and other products are posted to national repositories such as NPS Data Store. In addition, products are catalogued in NatureBib. Data products may not be posted if they contain protected information about the nature or location of rare, threatened, or endangered species, or other natural resources of management concern (see Appendix G for NatureBib record entry standards and Appendix H for a proposed NER NatureBib workflow).
12. 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 to cooperators. In all cases, distribution will follow legal requirements under the Freedom of Information Act (FOIA) and limitations established to protect information about sensitive resources. Note that, when in doubt, you should be conservative, as clauses in FOIA may require you to release data to the public at large if you've released it to anyone at all (see Appendix I for Director's Orders 11 for more details on information management and Web publishing in the NPS).
13. 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.

3.3. Integrating and Sharing Data Products

Once project data and data products have been finalized, they need to be secured in long-term storage and made available to others. To accomplish this requires that we use a range of information systems, such as product repositories, clearinghouses, and Web applications. Each of these systems has a different purpose and function. Table 3.1 shows each repository and the kinds of data objects found therein.

Table 3.1. Repositories for ERMN products.

Repository	Item
Local Digital and Physical Library	Project data, metadata, and other products Raw and certified data sets Hard copies of field data forms Metadata, protocols, SOPs Bound and digital copies of reports and administrative records Digital photographs, derived products
North Carolina State University Physical and Digital Archive	Final Inventory Project Files Final Vegetation Mapping Project Files Final Vegetation Mapping Aerial Photo Mosaic TIFF files Vegetation Mapping Aerial Photo Positives
Local 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 about water quality, park species lists and taxonomic documentation, park resource bibliographies
National Image Servers E-TIC and NPSFocus	Copies of digital reports and other documents (catalogued in NatureBib)
NPS Data Store	Static copies of digital data sets (non-sensitive) and metadata

3.3.1. Data Distribution

The process of product distribution involves several steps. As products are finalized, they can be sent to the appropriate person for integration, posting, and distribution. In most cases, it will be the Data Manager or database specialist who reviews the product for conformance with format standards then stores the product in the appropriate repository. After storing the products, their existence is documented by posting non-sensitive data and metadata for all projects and by updating other records where appropriate. The ERMN posts biological metadata records as the "master" metadata record for a project dataset, as opposed to spatial metadata records (these are zipped together with the spatial data itself). At this point, data discovery is accomplished, as metadata are then indexed by the NPS Datastore. Note that the NPS Datastore houses only static records, so monitoring data sets will need to be uploaded on a regular (probably annual) basis. These metadata records provide pointers to data and project products which are zipped together by ERMN staff in "Data" zip and "Report" zip files. At the time of this writing the NPS Data Store allows only the posting of two files per project. Distribution follows, as the search functions in the Datastore allow users to find data sets, and either:

A) in the case of sensitive data, using the information in the standard order process of the project's metadata record, request data from the data steward contact (usually the Network Data Manager); or

B) download the data directly from the NPS Data store.

When public versions of both the NatureBib and NPSpecies databases come online, they will both provide data discovery tools for the public and direct links to the NPS Data Store, as well. There is also an ArcIMS project in development that may provide some public I&M data discovery and visualization tools for the ERMN.

3.3.2. Integration with National Databases

In addition to storing and distributing data products, product integration also involves updates to national databases such as NatureBib, NPSpecies, and NPSTORET. All of these databases have local desktop versions which can be updated with data collected during the course of a project; although, the ERMN does not recommend using the desktop version of NatureBib. Desktop databases can then be uploaded and synchronized with the national databases on a regular basis.

Recently added features, including full-text downloads for non-copyright sensitive materials and enhanced reference download capabilities, make the NatureBib database a more useful I&M tool. ERMN staff posts full-text NPS reports to the NatureBib Data Store, along with peer-reviewed articles, dissertations, and theses whenever we can get permission from authors and publishers; this is typically done when final copies of documents are made available. However, some problems with redundancy arise, in that NER technical reports are also made available in full-text form at the regional science Website (http://www.nps.gov/nero/science/index.htm) and often also on the NPS Electronic Technical Information Center's site, which is searchable through the catch-all NPSFocus search engine. ERMN reports are certainly available and discoverable; however, there is a potential versioning and redundancy problem in the way that reports and other documents are currently synchronized with national databases and repositories. For more information on NatureBib in the NER see Appendix K.

To update NPSpecies, data on the distribution and occurrence of species in ERMN parks are compiled and added to the database upon delivery of data and data products. Synchronization occurs after the data is certified and sent to National I&M Program data management staff. For more information on the NPSpecies data workflow and roles and responsibilities see Appendix L.

For NPSTORET, any project collecting water quality data must feed its data into the NPSSTORET database, which, in turn, is a part of a larger EPA maintained STORET database of water quality information. Data collected by the network can be either extracted and uploaded or entered directly into NPSTORET. All water quality data must be managed according to guidelines from the NPS Water Resources Division. Figure 3.3 shows the flow of water quality data from the network to the EPA STORET database.

3.4. Credits

Much of the information presented above was adapted from the North Coast and Cascades Data Management Plan (Boetsch et al. 2005); including authors John Boetsch, Bret Christoe, and Ronald Holmes of the Olympic Prototype Park, Mount Rainier Prototype Park, and North Coast and Cascades Network.

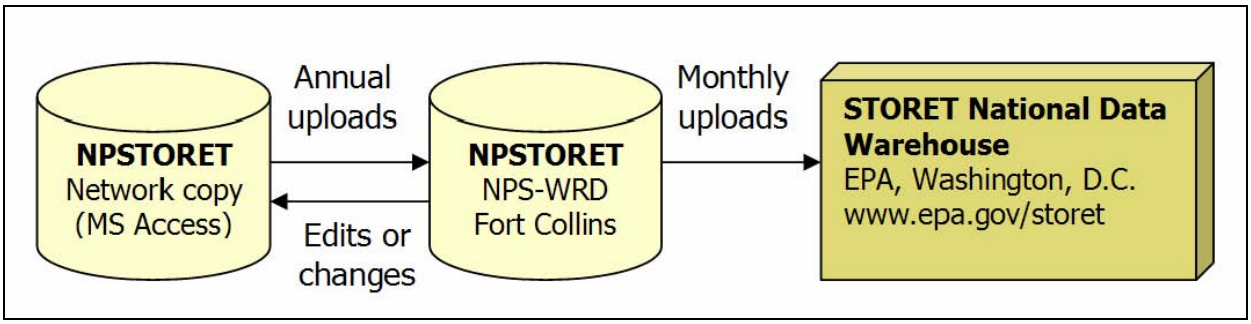

Figure 3.3. Data flow diagram for water quality data.

Chapter 4. Infrastructure and System Architecture

The I&M Program relies on university, park, regional, and national information technology personnel and resources to design and maintain a functional infrastructure to support I&M database and information system applications. There is currently a lot of interest within the NER, the Natural Resource Program Center (NRPC), the NPS, and the Department of Interior (DOI), in general, to move to enterprise data solutions. The first major movements in this direction have been within the GIS world with the NRPC's announcement that all future application development projects will take a service-oriented architecture (SOA) approach, but it is soon to pay all of us a visit. Adding some teeth to this will be restriction of funds for application development projects related to enforcement of the Clinger-Cohen Act of 1996 and implemented by the NPS Office of the Chief Information Officer (OCIO).

Much of this chapter is specific to the ERMN and their network office location at The Pennsylvania State University (PSU), and I imagine that the following information will quickly become out-of-date, given the speed with which the NPS has taken to enterprise system architectures and imminent ERMN plans to make changes to local network and system architectures.

4.1. Network and System Resources

ERMN computers sit on a local network within the School of Forest Resources at PSU and rely on local administrators to keep the network operational (this has never been a problem). ERMN staff connects to the NPS network through a virtual privacy network (VPN) software connection that's usually reliable, but provides only a temporary NPS network connection. The Network Coordinator and Data Manager's computers have common folders mapped to one another's computers where shared files reside, but this connection is only functional when not also simultaneously logged onto the NPS network, as the VPN seems to block all other outside communications with the local machine. The hope is, that by installing a local server, the ERMN can bypass some of these connection issues and establish a more effective local network.

At the PSU based NPS offices there's limited to no IT support from either PSU staff or the Philadelphia or Boston Support Offices of the NPS (where other administrative support staffs reside). ERMN staff can fend for themselves in keeping their computer hardware, software, and limited network connections operational. The Network Data Manager serves as the "de facto," although not always very competent, IT specialist at the PSU NPS offices.

Infrequent hardware purchases and software updates are forwarded through either the Boston or Philadelphia support offices of the NPS. This includes new computer purchases and basic NPS software installation.

Reliable and efficient computer and network resources are an integral aspect of a functional I&M monitoring network. There's a trade-off of both advantages and disadvantages to being duty-stationed at a remote NPS office location and some added IT and computer infrastructure hurdles are some of the disadvantages. ERMN staff needs to remain diligent to make sure they have all the computer and network resources they need to support a functioning monitoring program. It is anticipated this will remain a stumbling block into the future.

4.2. ERMN Digital File Storage and Maintenance

(The following information represents how the ERMN has organized files for inventory datasets; whereas it is intended that monitoring data acquired in the future will be treated differently).

With the exception of working files found on the local machines of ERMN staff, there are three main locations where ERMN digital files are stored:

1. On a local drive between the ERMN Network Coordinator and Data Manager;
2. On the machines of PSU based cooperating staff; and
3. In a digital archive at North Carolina State University (NCSU).

The ERMN maintains two primary and standardized folder structures for most shared digital resources, in addition to a custom database to track project finances. The first folder is for ongoing projects—its structure was borrowed from "Information Management of a Project for Kenai Fjords National Park," created by Dorothy Mortenson of the Southwest Alaska Network. There is also a working folder for projects that are finished, but not yet posted to the NPS Data Store, and an "Archive" folder for projects that are complete. The latter does not represent the ERMN's digital archive; it is simply a way to keep old project information accessible while not cluttering current project information. Figure 4.1 shows a screen capture of this project folder structure.

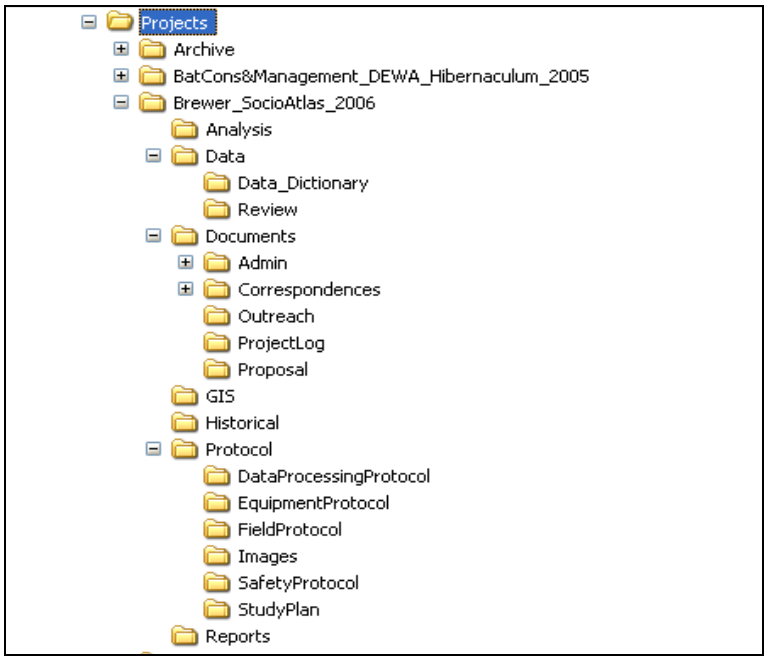

Figure 4.1. ERMN project folder structure.

The second shared folder is a spatial and tabular data folder for completed and certified information organized by park and referred to as the "Park Data" folder. Though presence in this folder doesn't guarantee documentation, most datasets do have some form of metadata. This folder is searchable in the sense that, if the ERMN possesses a dataset on a species at a park, it will be in that park's data folder. These static project specific data folders mirror the aforementioned project folders. This folder structure is also available from the Standards and Policies page of the I&M website: http://science.nature.nps.gov/im/datamgmt/docs/datastr5.doc. A screenshot of an ERMN park data folder is shown in Figure 4.2.

The final main repository that stores ERMN data is a digital and physical archive under long-term cooperative agreement maintained by Dr. Hugh Devine located at NCSU. All short-term project records and information, a working version of the ERMN Website and other data sets and records of "high value" to the network are placed in this archive. Accessibility to this archive is low, but security and stability is considered to be high.

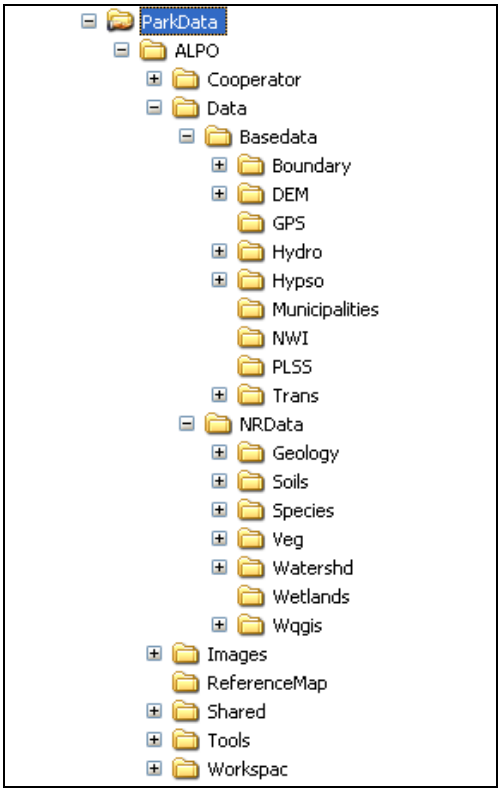

Figure 4.2. ERMN park data folder.

4.3. National Information Management Systems Architecture

The NPS Natural Resource Program Center (NRPC) and the I&M Program actively develop and implement a national-level, program-wide information management framework. NRPC and I&M staff integrate desktop database applications with Internet-based databases to serve both local and national-level data and 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 utilizing natural resource information.

4.3.1. National-level Application Architecture

To achieve an integrated information management system, two of the national-level data management applications (NatureBib and NPSpecies) use distributed application architectures with both desktop and Internet-accessible (master) components, while the other (NPS Data Store) serves as a stand-alone Web application with links to a variety of other NPS (NPSFocus) and government-wide (National Spatial Data Infrastructure node) search engines (Figure 4.3).

4.3.1.1. NatureBib: NatureBib is the master database for bibliographic references that merges a number of previously separate databases, such as Whitetail Deer Management Bibliography (DeerBib), Geologic Resource Bibliography (GRBib), and others. It also contains citation data from independent databases like NPSpecies and the Dataset Catalog and NR-GIS Metadata Database. It currently focuses on natural resource references, but may eventually be linked to references on cultural resources and other park operations. See Appendix K and or the following web link for more information on NatureBib: http://www.nature.nps.gov/nrbib. The NatureBib online version can be accessed at: https://science1.nature.nps.gov/naturebib/nb/simple/clean.

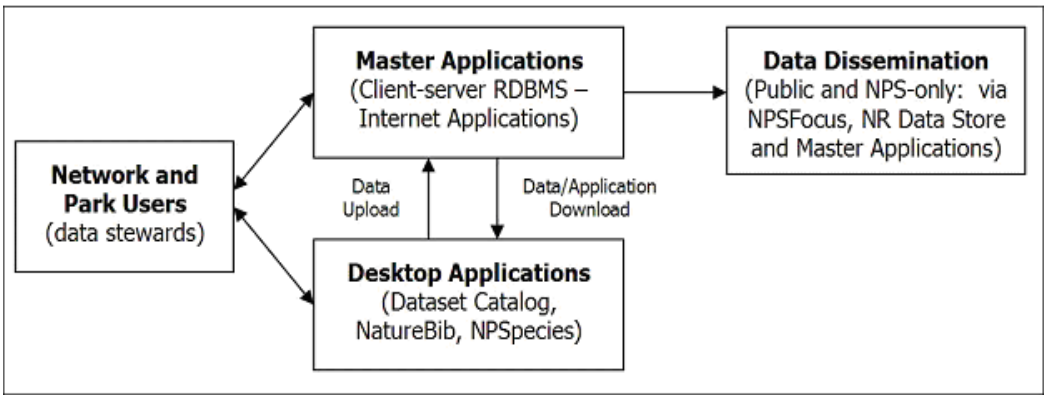

Figure 4.3. Model of national level application architecture.

4.3.1.2. NPSpecies: NPSpecies is the master species database for the NPS. The database lists the species that occur in or near each park and the physical or written evidence for the occurrence of the species (e.g., references, vouchers, and observations). Taxonomy and nomenclature are based on ITIS, the interagency Integrated Taxonomic Information System. The master version of NPSpecies for each Park or Network can be downloaded from the master Website into an MS Access version of NPSpecies. The Internet-based version is the master database, which can be accessed via password-protected logins administered by park, Network, and regional data stewards. The master database requires that species lists are certified by Networks before any data will be available to the public. NPSpecies is linked to NatureBib for bibliographic references that provide written evidence of a species' occurrence in a park and will be linked to NR-GIS Metadata Database to document biological inventory products. The MS Access application and additional details can be found at the NPSpecies Website: http://science.nature.nps.gov/im/apps/npspp/index.htm

NPS Data Store: The NPS Data Store is a key component of the data dissemination strategy employed by the I&M Program. The NPS Data Store is a graphical search interface that links dataset metadata to a searchable data server on which datasets are organized by NPS units, offices, and programs. The interface allows customized public or protected searches of natural, cultural, facilities and fire resource geospatial data, tabular data, map products, data standards, and other products produced by the I&M and other NPS programs. Each park or Network is able to post and curate its own data on the server. Links to NatureBib and NPSpecies are coming soon and will enhance the utility of the NPS Data Store. See the NPS Data Store Website for further information: http://science.nature.nps.gov/nrdata/.

4.3.1.3. Other National-Level Inventory and Monitoring Information Management Applications

4.3.1.3.1. NPSTORET: STORET is an interagency water quality database developed and supported by the Environmental Protection Agency (EPA) to house local, state, and federal water quality data collected in support of managing the nation's water resources under the Clean Water Act. STORET is used by the NPS as a repository of physical, chemical, biological, and other monitoring data collected in and around national park units by park staff, contractors, and cooperators. The NPS operates its own servicewide copy of STORET and makes periodic uploads to the EPA STORET National Data Warehouse so that data collected by and for parks will be accessible to the public. NPS Director's Order 77 indicates that the NPS should archive water quality data in STORET, and the NPS Water Resources Division (WRD) requires that any data collected as part of a WRD funded project get archived in STORET. NPSTORET is the NPS master database designed to facilitate park-level standardized reporting for STORET. More details on NPSTORET can be found at: http://www.nature.nps.gov/water/infoanddata/index.cfm. Additional information on STORET can be found at: http://www.epa.gov/storet.

4.3.1.3.2. Natural Resource Database Template: The Natural Resource Database Template (NRDT) is a flexible, relational database in MS Access for storing inventory and monitoring data (including raw data collected during field studies). This relational database can be used as a standalone database or in conjunction with GIS software (e.g. ArcGIS) to enter, store, retrieve, and otherwise manage natural resource information. The template has a core database structure that can be modified and extended depending on the components and specific sampling protocols in use. The NRDT adds an element of compatibility to databases across networks but does not

ensure it. Field methods must also be coordinated in order to be sure that data within databases can reasonably be rolled-up or compared. The NRDT provides a good core set of tables and relationships on which to build a personal level database; however, as the NPS moves towards a more enterprise level data solution, MS Access will likely be replaced by another database application. Monitoring protocols, including databases that are based on the Database Template, are made available through a Web-based protocol clearinghouse (see below). A description of the NRDT, a data dictionary, and example implementations are located on the NR Database Template Website: http://science.nature.nps.gov/im/apps/template/index.cfm.

4.3.1.3.3. Natural Resource Monitoring Protocols Clearinghouse: The Natural Resource Monitoring Protocol Clearinghouse (i.e., Protocol Database) is a Web-based clearinghouse of sampling protocols. The database provides a summary of, and in many cases allows the user to download a digital copy of, sampling protocols that have been developed by Networks or others and are being applied in national parks. The Protocol Database also makes it possible to download database components (e.g., tables, queries, data entry forms) in MS Access that are consistent with the NRDT that have been developed for a particular protocol. See the Protocol Database Website for available protocols:
http://science.nature.nps.gov/im/monitor/protocoldb.cfm.

4.4. Credits

Material in this chapter was adapted from Rob Daley, author of the Greater Yellowstone Network Data Management Plan (Daley 2005), and Sara Stevens and Gary Entsminger, authors of the Northeast Coastal Barrier Network Data Management Plan (Stevens et al. 2004).

Chapter 5. Data Acquisition and Processing

This chapter describes the general procedures the ERMN follows for acquiring and processing natural resource-related data. These steps will vary depending on the data source, which can be placed into three general categories:

- Eastern Rivers and Mountains Network data: data resulting from projects that are initiated, sponsored, or funded by the ERMN;
- Other National Park Service (NPS) data: data resulting from projects that are initiated, sponsored, or funded by park units or by regional or national NPS programs; and/or
- External data: data produced or managed by agencies, organizations, or individuals other than the NPS.

5.1. Eastern Rivers and Mountains Network Data

ERMN staff have the most control over data produced for or by the Network. These data will be processed, maintained, and archived by the Network. Network data usually fits into one of the following categories:

- Inventory Data;
- Monitoring Data; or
- Administrative Records.

5.1.1. ERMN Inventory Data

Most ERMN-sponsored inventory projects are completed by cooperators working under a contract or agreement that has been written by the network for the project. In these agreements, the ERMN specifies data deliverables and formats that are required from the cooperator, as is outlined in Appendix A. In other cases, inventories are carried out under the supervision of park staff if they have an existing working relationship with the cooperator, or the network is simply adding money in combination with park resources to achieve a mutually desired goal. The data manager works as closely as possible with the principal investigator and the project leader on inventory projects. However, despite everyone's best intentions, data management still seems to be the first thing with inventory projects to fall through the cracks. Whether it is a cooperator's unwillingness to change their work habits, or the network's inability to stay on top of data management with the project, such is the nature of these projects. The network data manager needs to pay special attention to these short-term projects.

Due to the short time frame of inventory projects it is not always worth the effort of ERMN staff to try to dictate how a cooperator will manage their data if they are not fully receptive. However, what you put off today you will have to deal with tomorrow, and it is sometimes the case that several rounds of revisions are necessary before data are delivered in acceptable form. Keep this in mind when writing up deliverable dates and make sure that there is ample time and money planned for data revisions. It is often a good idea to request that data be submitted along with progress reports or at a project mid-point to ensure that it is being managed appropriately.

In other cases, where an inventory project has fallen short of its data management goals, the ERMN data manager may spend extra time converting and processing data sets after the project is complete; this is often the most efficient use of ERMN staff time. This somewhat liaise-fare approach to inventory data projects was born from past experiences where large amounts of time were invested building elegant, efficient, and "user-friendly" databases and tools for cooperators and then training and educating them in their use. The end result was often the same: poorly managed inventory data attributed to the fact that data management is more than just databases. There are few inventory cooperators that have good data management skills and then there are others; keep in mind that they were likely chosen for their expertise in something other than data management and most all have nothing but the best intentions. Patience and diligence is required to ensure that quality data products are delivered by inventory cooperators. See Appendix D for details on inventory project product review.

Once inventory project products are accepted as final by both the Network Data Manager and Project Leader and final invoices are paid (see Administrative Records section below), the data and information are sent to North Carolina State University for additional review and biological metadata creation. NCSU staff will send the final products back to the network, whereby network staff convert the data to NPS standards where appropriate and enter them into I&M program cataloging and indexing databases in the steps generically depicted in Figure 5.1.

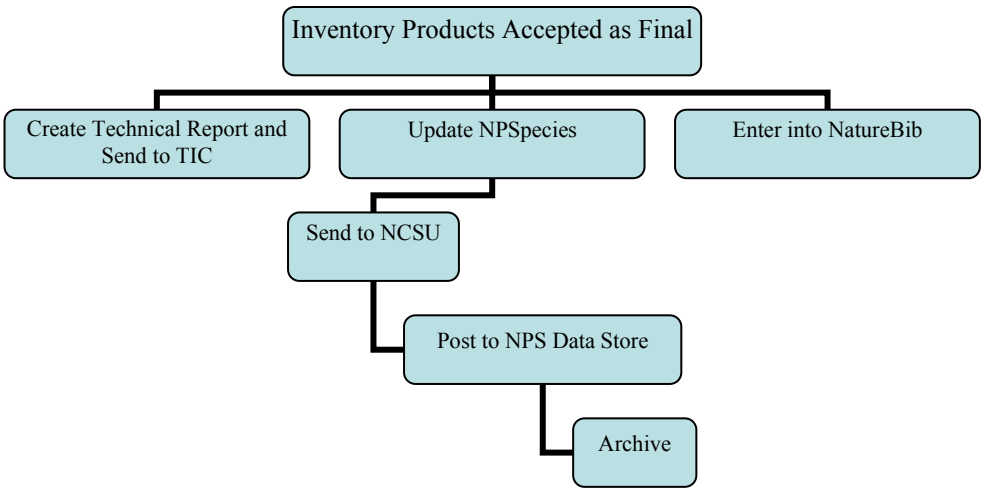

Figure 5.1. Inventory data flow model.

5.1.2. Monitoring Data

Data acquisition for network monitoring projects is guided by established protocols that include step-by-step procedures for all aspects of data collection, entry, quality assessment (QA), quality control (QC), reporting, analysis, distribution, and archiving as depicted in Figure 5.2. Depending on the protocol, data can be acquired by park staff, network staff, cooperators, or other agencies or organizations. As with inventory projects, monitoring projects will have specific data deliverables included in the project contracts and agreements, where not carried out by NPS staff. Monitoring data will be collected using a variety of methods specific to the vital sign in question, including field data collection (paper sheets, field computers, GPS), remote sensing, and data logger deployment.

Monitoring data need to go though steps repeated annually or on a schedule corresponding to sampling frequency. Each monitoring project will have a processing and integration checklist that is specific to the project to ensure that all steps are completed.

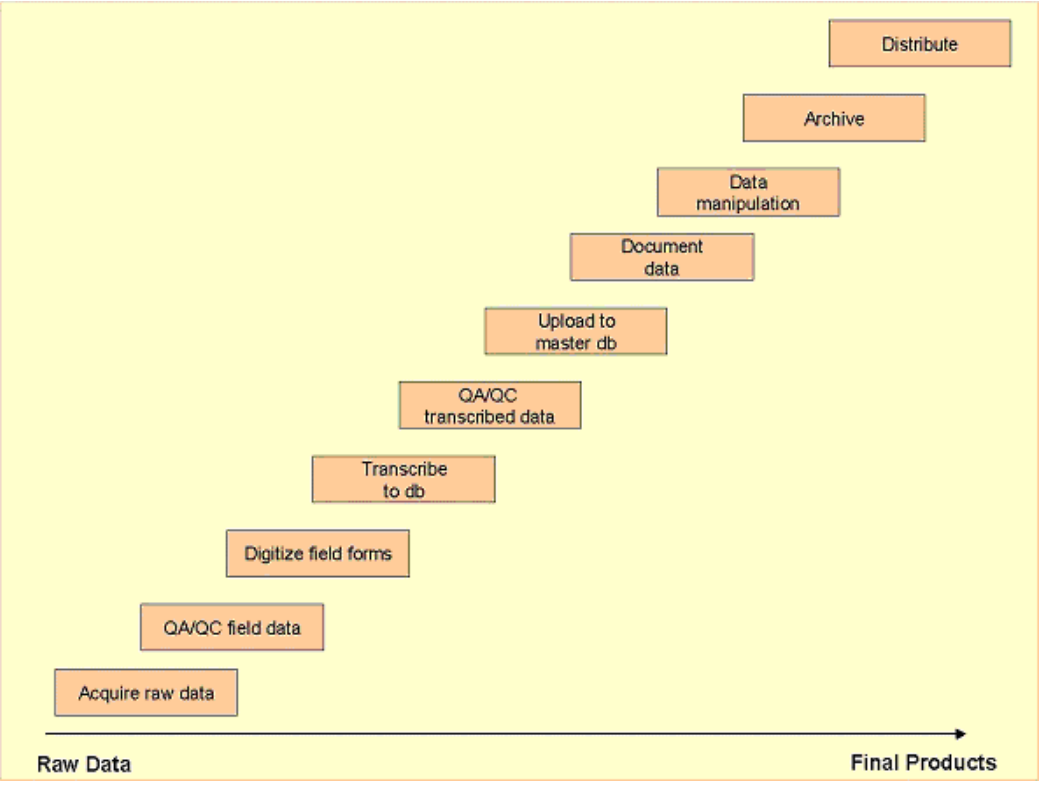

Figure 5.2. Data flow model for monitoring projects. (Note, this diagram was adapted from Approaches to Quality Assurance and Information Management for Regional Ecological Monitoring Programs, Craig Palmer, 2003.)

5.1.3. Administrative Records

Administrative records can include cooperative agreements, task agreements, modifications, and invoices. Many of these records are stored in the ERMN shared "Projects" folder with their associated project specific files (see chapter 3). However, others are stored and processed in an ERMN project tracking database from which hard copies (of invoices paid, for example) are printed and stored in the ERMN's project physical files and eventually in the ERMN library (physical archive) when the project is complete. For task agreements, the ERMN typically withholds one invoice equal to five or ten percent of the total awarded to the project until products are accepted as final by the network. Figure 5.3 shows the project information screen of the ERMN project tracking database which resides on the local computer of the Network Coordinator.

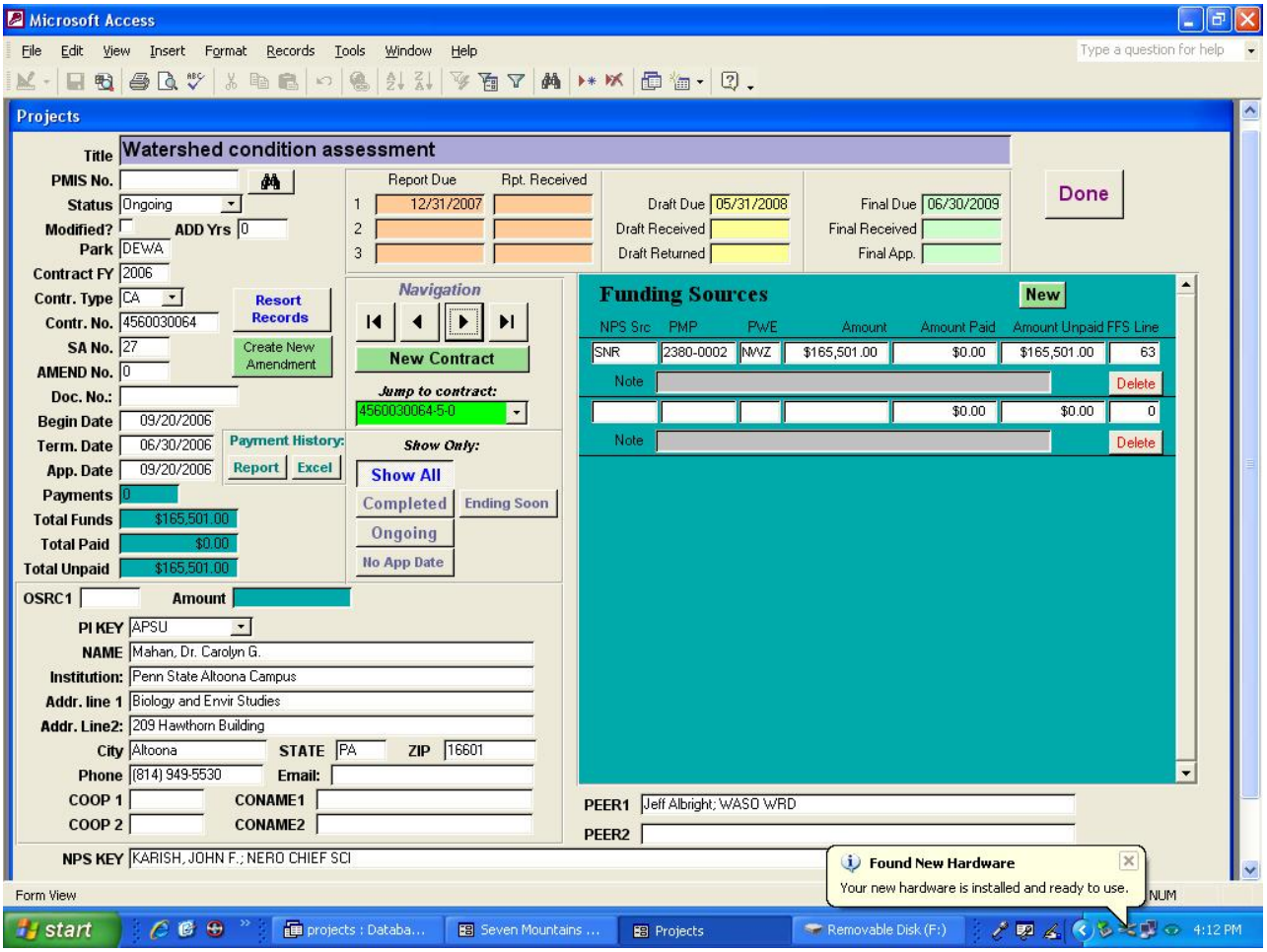

Figure 5.3. ERMN project tracking database.

5.2. Other National Park Service Data

A fundamental step in developing and maintaining the information base for the ERMN is locating, evaluating, and documenting, on an ongoing basis, park-related natural resource information. The term "data mining" refers to this process, which requires regular visits to network parks and establishing data mining contacts and procedures tailored specifically to each park. Unfortunately, this most often requires park staff time (which is very limited), so establishing familiarity and positive rapport with park staff is of utmost importance. The range of park materials that require documenting is broad, ranging from historical inventories to complex databases to remote-sensing data.

Over the course of the last three years, the ERMN has undertaken a large effort to both search for park relevant digital documents and acquire and scan hard-copy only park or natural resource management documents of critical importance. These documents have been safe-guarded in their duplication and digitization and have been made discoverable by being cataloged in NatureBib. Documents that aren't protected by copyright or other publisher agreements are posted as full-text, downloadable digital documents from the NatureBib Data Store. It is increasingly unusual for reports or other documents to be unavailable in digital format; however, in the off-chance that some are encountered, the ERMN expects to continue to serve in this capacity for parks in the future. See Appendix M for more information about scanning project personnel commitment and technical specifications. See Appendix N for proposed digital image metadata standards not yet implemented by the NPS or ERMN.

The goal of park data mining is not to store all park data in the network office, but to inform the network of data sets that exist and are maintained at the park, catalog them using I&M program databases, and, where appropriate, acquire duplicate copies either for network use or general safe-keeping. Some parks have the ability to do this themselves and have adopted effective schemes for doing so, whereas Network staff need to assist other parks in keeping these databases up to date.

One way that the ERMN intends to keep its records up to date is by monitoring the NPS Project Management Information System (PMIS) for new projects and contact the appropriate park staff to inquire about those specific projects. In this way the onus is not placed on a single park staff to transfer records, and park staff, in general, do not need to keep track of what they have passed along to ERMN staff.

5.3. External Data

Institutions and organizations outside of the NPS are valuable data resources which should be mined periodically and their data incorporated into Network and I&M databases where appropriate. This process follows much the same steps as outlined with the above datasets and, in a generic sense, involves the following:

Acquire data→Evaluate→Transform→Catalog→Make Available

Outside organizations that have provided fruitful searches to date include:

- Museums, Herbaria, and other specimen repositories;
- Other federal agencies, including the U.S. Geologic Survey (USGS), U.S. Fish and Wildlife Service (USFWS), and Environmental Protection Agency (EPA), among others;
- State organizations including Departments of Conservation of Natural Resources, GIS programs, Universities, and Natural Heritage programs;
- Private organizations, including NatureServe, The Nature Conservancy (TNC), the Western Pennsylvania Conservancy (WPC), and private universities; and/or
- Other organizations and individuals as they present themselves, usually through word of mouth or as an unexpected product of another search effort.

5.4. Credits

Information in the above chapter was adapted from Margaret Beer, Elizabeth Nance, Aneth Wight, Melissa Powell, and Russ DenBleyker, all of the Northern Colorado Plateau Network and their Data Management Plan (Beer et al. 2005) and Sara Stevens and Gary Entsminger, authors of the Northern Coastal and Barrier Network (Stevens et al. 2004).

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

High quality data and information are mandated by national NPS directives and orders and are vital to the credibility and success of the I&M program. According to Abby Miller (2001):

> *"Data need to meet national-level quality standards and be accessible for wise and defensible decision-making at all levels. Data need to be shared and aggregated with data from other parks and from adjacent lands to support landscape-level and national planning and decision-making."*
> -retired deputy associate director, Natural Resource Stewardship and Science Division, NPS, WASO Office

What's more, the NPS Director's Order #11B, issued in 2002 (see Appendix I), titled "Ensuring Quality of Information Disseminated by the National Park Service," promotes information and data quality. It defines 'quality' as incorporating three key components–objectivity, utility, and integrity.

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

Utility refers to the usefulness of the information to its intended users, from the perspectives of both the Network, perhaps most importantly natural resource managers, and the general public.

Integrity refers to the security of information, which could entail protection from unauthorized access or revision to ensure that the information is not compromised through corruption or falsification.

Director's Order #11B (also in Appendix I) also specifies that information must be based on reliable data sources, which 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.

6.1. Quality Assurance and Quality Control Defined

Quality assurance procedures plan for quality in all stages of the data development process, while quality control procedures monitor or evaluate the resulting data products. Palmer (2003) defines quality assurance as "an integrated system of management activities involving planning, implementation, documentation, assessment, reporting, and quality improvement to ensure that a process, item, or service is of the type and quality needed and expected by the consumer." He defines quality control as "the overall system of technical activities that measures the attributes and performance of a process, item, or service against defined standards to verify that they meet the stated requirements established by the customer."

QA/QC mechanisms are designed to prevent data contamination, which occurs when a process or event other than the one of interest affects the value of a variable and introduces two

fundamental types of errors into a data set. Errors of commission include those caused by data entry and transcription errors or malfunctioning equipment. They are common, fairly easy to identify, and can be effectively reduced upfront 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 of these errors should be revealed by rigorous QC procedures.

QA/QC procedures applied to ecological data include four activities ranging from simple to sophisticated, inexpensive to costly: 1) defining and enforcing standards for electronic formats, locally defined codes, measurement units, and metadata; 2) checking for unusual or unreasonable patterns in data; 3) checking for comparability of values between data sets; and 4) assessing overall data quality. Much QA/QC work is related to the first activity, which begins with data design and continues through acquisition, entry, metadata development, and archiving. The progression from raw data to verified data to validated data implies increasing confidence in the quality of the data through time (Figure 6.1).

6.2. General Operations

There are a number of easy ways to ensure that data quality is maintained through day-to-day Network operations—two of the most important of these are file naming conventions and version control.

6.2.1. File Naming

File naming standards are an important and basic quality assurance procedure. The following guidelines apply to all Network files:

- Do not use spaces or special characters within a file name;
- Use underscores or periods as delimiters instead of spaces, where necessary;
- Include a date, in yyyy.mm.dd format (e.g., 2006.12.05 for December 6, 2006) to help with version control (see section below), and save drafts regularly; and
- Include an indication of the QA/QC status (raw, draft, final, or verified etc.) of the file in the name element. For example: "ERMN_DM_Plan_draft_2006.12.05.doc."

6.2.2. Version Control

Before making major changes to a file, a copy of the file with the appropriate control (in this case the file name) should be made. As indicated above, files are stored with the appropriate eight digit date which serves as a version control. This allows changes to be tracked over time and facilitates efficient collaboration between multiple personnel working on a common file.

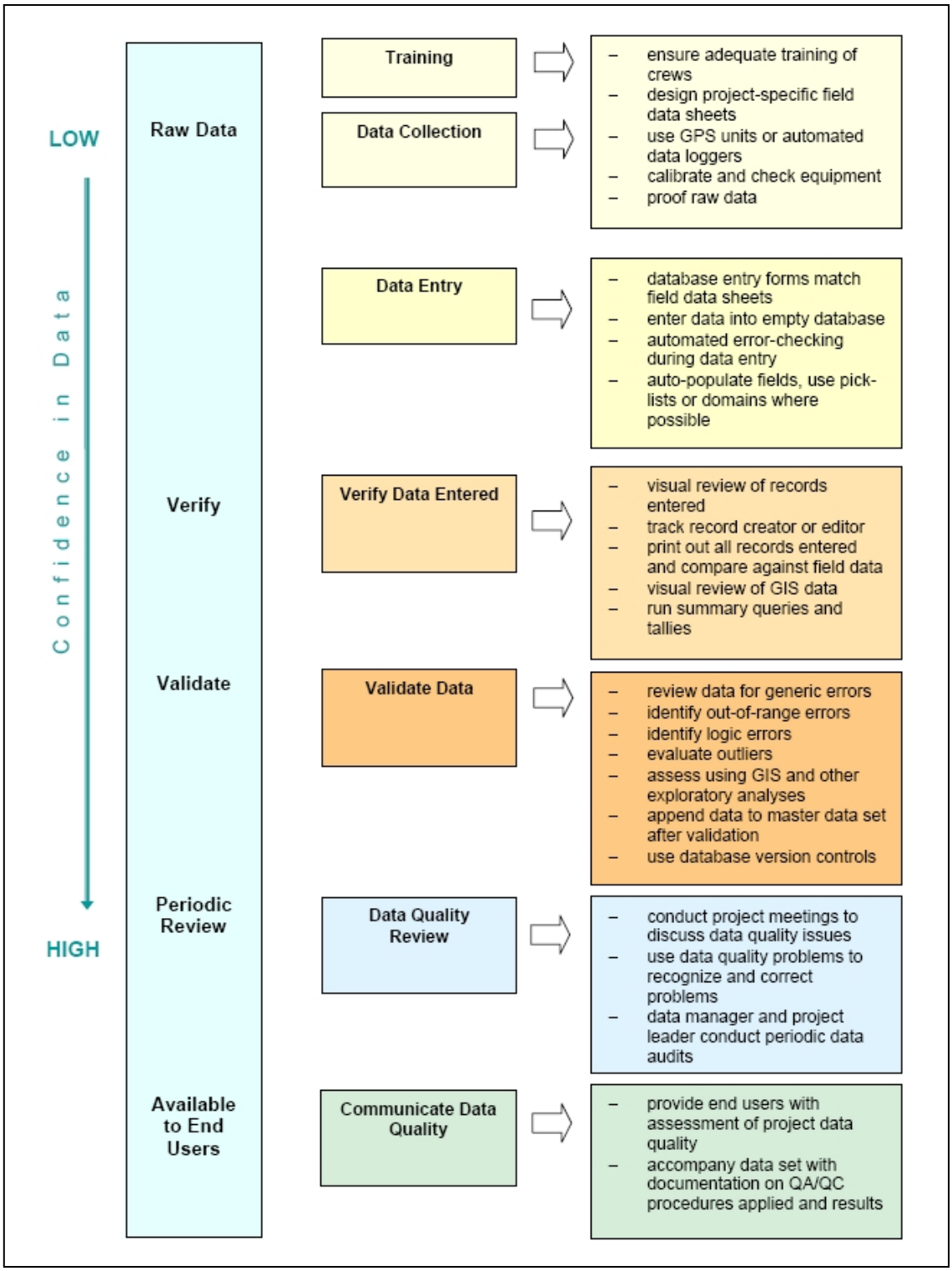

Figure 6.1. QA/QC controls and project workflow.

6.3. Monitoring Protocol Development

The most effective mechanism for ensuring that ERMN monitoring protocols produce data of the desired type, quality, and quantity is to provide detailed procedures and guidelines to assist ERMN staff in accurate data collection, entry, and validation. The network's approach is to develop a comprehensive set of SOPs and data-collecting protocols for quality control. These include clear field methodologies, well-trained field staff, well-organized field forms, and data entry applications with simple built-in validation controls.

Although QA/QC procedures will depend upon the individual vital signs being monitored and must be specified in the protocols for each monitoring vital sign, some general concepts apply to all network projects. These general quality assurance guidelines will ensure that all data collected are checked for integrity before being integrated into monitoring program databases. Refer to SOPs and monitoring protocols for specific QA/QC procedures. The following quality assurance guidelines are applied to all protocols:

- Common lookup tables are created for values of parameters recorded in an identical field for more than one project;
- Each project has SOPs for core data management (collection, entry, verification, etc.);
- Each project uses, at a minimum, standard field sheets for data collection;
- Data entry forms resemble field sheets;
- Automated error checking features are included in database applications;
- Database application design will maximize the use of auto-fill, auto-correct, range limits, pick lists, and other constraints specific to projects;
- Database applications will include a means to track errors reported on the data after dissemination;
- Database maintenance logs will be maintained for each ERMN database and housed in association with database files; and
- Databases will have a record level track log that shows who (what user) created or last edited a given record.

6.4. Data Collection

Attention to detail while collecting information in the field is crucial to achieving desired data quality. The following guidelines will be applied to each ERMN monitoring protocol:

- Each protocol will have a data collection SOP. Data collection SOPs will include step-by-step instruction on field methods and data collection. These SOPs will be reviewed and maintained by project leaders or cooperators on an annual basis. All field methods will be peer reviewed. All field sheets (where mobile computing devices are not being used) and field data recording procedures will be reviewed and approved by the data manager prior to data collection.
- Protocols will include specific instructions on field equipment use, including calibration, maintenance, and minimum timing requirements. Accurate field measurements are dependent upon field equipment regularly being calibrated and maintained. Field crews will be required to record, along with their field data, equipment calibration and failures.

- All Network project leaders will ensure that field crews understand all procedures by providing pre-field season training. Field crews are expected to understand the existing data collection forms or databases, know how to take measurements correctly and accurately, and follow the appropriate SOPs. Prior to routine data collection, training sessions will ensure that field personnel have a clear understanding of data collection procedures described in the SOPs. The ERMN will develop training manuals as part of the protocols that will involve either a large number of field staff or annual field staff turn-over. Network project leaders will also develop a process to certify that field staff understand and can perform the specified data collection procedures.
- All Network project leaders will conduct data collection reviews throughout the field season. Project leaders will periodically review the work of field technicians to ensure that their work does not drift from standards during the course of the field season.
- The following apply only to projects using hard-copy field forms for data collection:

    1. Information added to field sheets must be printed and clearly legible.
    2. If alterations to the data are necessary, the original data/information should be crossed out with a single line and the new information written next to the original entry. Information should never be erased and old information should not be overwritten.
    3. Upon return from the field, copies of all original data sheets should be made and checked for legibility and completeness (i.e., no data cut off at the edges). Copies of the data sheets will be stored as specified in the protocol SOP, and the original data sheets will be used for data entry.
    4. Standardized, project-specific field forms will be used for all projects not using mobile digital field collection options. Acid-free paper will be used for all field forms to prevent fading and subsequent data loss. Some circumstances may warrant the use of paper and writing implements that can withstand moisture, dust, and other extreme environmental conditions.
    5. Field forms will contain as much basic preprinted project information as possible and sufficient space for recording additional information. Pick lists will be provided on field sheets wherever possible to ensure proper data are recorded.

- Finally, all project leaders and field crews will be required to maintain a log of important decisions and events that occurred during the field season. This will allow for proper revisions of SOPs, if necessary.

6.5. Data Entry and Import

'Data entry' is the initial set of operations where raw data are entered or transferred to a database table using a computerized form or automated routine. When data are gathered or stored digitally in the field (e.g. on a data logger), data entry consists of transferring the data (downloading) to a file on an office computer where they can be further manipulated. The goal of data entry is to transcribe field observations into a computer database with 100% accuracy. In other words, that which is recorded in the field should be entered exactly in the database. Subsequent data verification is conducted to ensure that 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 during data entry. The following quality assurance guidelines will be applied by all ERMN monitoring protocol:

- Data will be entered as soon as reasonably possible after collection.
- Data entry will be done by someone familiar with data collection. The project manager (with assistance from the data manager, if needed) must ensure that the data entry staff is familiar with the database software, database structure, and any standard codes for data entry 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. They must learn how to commit both a 'field entry' and a 'complete record' entry and 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 to maximize error control. Whenever possible, data should not be entered into spreadsheets. 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 automated data entry.

6.6. Data Verification (Quality Control Part 1)

Network staff appraise the quality of data by applying verification and validation procedures. *Data verification* determines whether the digitized data match the source data and *data validation* makes sure that the data make sense. It is essential that we validate all data as truthful and do not misrepresent the circumstances and limitations of their collection. Although data entry and data verification can be handled by personnel who are less familiar with the data, validation requires in-depth knowledge of the data.

All ERMN monitoring protocols will adopt the following data verification guidelines.

- Project leaders are responsible for specifying in the monitoring protocol one or more of the data verification methods listed below.
- The project leader is responsible for proper execution of data verification.
- Data verification is carried out by staff who are thoroughly familiar with data collection and entry. All records (100%) will be verified against original source data. A subset of randomly selected records (10% of records) will be reviewed after initial verification by the project leader. If errors are found, the entire data set is verified again.
- A record of each data set's verification process, including number of verification iterations and results, will be prepared by the project leader as part of formal metadata generation (see Chapter 7 for more details on data documentation).

6.6.1. Methods for Data Verification and Quality Control

Each of the following methods has a direct relationship between effectiveness and effort. 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. Recorded values in the database are compared with the original values from the hard copy.

Identified errors are immediately corrected. This method is the least complicated, since it requires no additional personnel or software. Its reliability depends entirely upon the person entering data and, thus, is probably the least reliable data verification method.

- Visual review after data entry: All records are printed upon completion of data entry. The values on the printout are compared with the original values from the hard copy. Errors are marked and corrected in a timely manner. This method's reliability increases if someone other than the individual entering the data performs the review. Alternatively, two technicians can perform this review. One data entry technician reads the original data sheets (the reader) and the other reads the same data on the printout (the checker).
- Duplicate data entry: The data entry technician completes all data entry, as normal. Random records are selected (every nth record) and entered into an empty replica of the permanent database, preferably by someone other than the person entering the permanent data. A database query is then used to compare the duplicate records from the two data sets in an automated fashion and report any apparent errors. Disparities are manually reviewed and corrections applied if necessary. This method adds the overhead of retyping the selected records, as well as the creation of a comparison query. Professional data entry services frequently use this method.

6.6.2. Additional Methods

Data verification can be improved by calculating summary statistics and by identifying duplicate or omitted records. For example, we can view the number of known constant elements, such as the number of sampling sites, plots per site, or dates of samples. Posing questions in different ways and looking for differences in the answers can reveal errors. The more checks we devise to test the completeness of the data, the greater our confidence that we can have in the data.

6.7. Data Validation (Quality Control Part 2)

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.

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 the certification of the data and a required component of any ERMN monitoring protocol. Invalid data commonly consist of slightly misspelled species names or site codes, wrong dates, or out of range errors in parameters with well defined range limits (e.g. pH or elevation). But, 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). We call these types of erroneous data logic errors because using them produces 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 ERMN will adopt the following guidelines regarding data validation:

- Each ERMN monitoring protocol will abide by the data validation guidelines in this document, including selection of at least one of the specific methods presented above as part of a comprehensive data validation process.
- 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 editor's initials.
- Modifications of the field data should be clear and concise while preserving the original data entries or notes (i.e., no erasing!).
- Validation efforts should also include a check for the completeness of a data set since field sheets or other sources of data could easily be overlooked.
- The ERMN will maximize the use of automated routines and/or data summary and visualization such as histograms, line plots, and basic statistics to reveal possible logic and range errors.
- The ERMN will maximize the use of database programming to control data entry. In general, this will be achieved via the use of lookup tables, drop down pick lists, and other ways of controlling data entry.

6.7.1. Methods for Data Validation

Specific procedures for data validation depend upon the vital sign being monitored and will be included in monitoring protocol SOPs. The following general methods can be used to validate data:

1) Data entry application programming: Certain components of data validation are built into data entry forms. This method is essentially part of database design. Not all fields, however, have appropriate ranges that are known in advance, so knowledge of what are reasonable data and a separate, interactive validation stage are important. Caution must be exercised when using lookup tables to constrain variable values. Values occurring outside the range set by a lookup table may not always be invalid. As part of data validation procedures, the project leader is responsible for appropriate use of lookup tables and other automated value range control.
2) 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 set 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 be considered 'unusually' extreme. Data quality assurance 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 a better way to explain this quality assurance goal. If contamination is not detected during data collection, it is usually only detected later if an outlying data value exists. When an outlier is detected, an attempt should be made to determine if some contamination is responsible. Database, graphic, and statistical tools can be used for ad-hoc queries and display of the data to detect outliers. Some of these outlying values may appear unusual but prove to be quite valid after confirmation. Noting correct but unusual values in documentation of the data set saves other users from checking the same unusual values.

3) Other exploratory data analyses: There are countless other statistical and exploratory methods for familiarizing oneself with a data set and these should be planned for and applied, as appropriate, by project leaders.

6.8. Data Quality Review and Communication

The NPS requires QA/QC review prior to communicating or disseminating data and information. Only data and information that adhere to NPS quality standards may be released. Data and information disseminated to the public must be approved by the appropriate reviewing officials and programs. Documentation of the QA/QC standards used in producing the information and that substantiate the quality of the information must be formally certified and distributed with related data and information. Also, mechanisms must be in place for receiving and addressing comments or complaints pertaining to the quality of data. Chapter 9 has a more in-depth discussion of Network data dissemination.

6.9. Conformance Guidelines and Revisions

The material presented in this chapter is meant to be a starting point to provide general guidelines to project leaders and other Network staff. As such, this plan and these methods are only a beginning, likely to be modified, added to, and in other ways changed. It is the responsibility of the Network Data Manager, working in conjunction with project leaders, to revise this document, as well as monitoring protocol and SOPs, as needed. This includes checking to make sure that methods are up to date and appropriate to the subject, being applied correctly by staff, and, most importantly, improving the quality of network data. Things that the Data Manager or project leader could check for include:

- Data collection and reporting requirements are being met in accordance with guidelines;
- Verification and validation procedures are being followed in accordance with guidelines;
- Data file structures and maintenance are clear, accurate, and effective;
- Revision control of program documents and field sheets are adequate and effective;
- Calibration and maintenance procedures of equipment are being followed;
- Seasonal and temporary staff have been trained in Network data management practices;
- Metadata collection and creation proceeds in a timely manner; and
- Data are being archived and catalogued appropriately for long term storage.

6.9.1. Documenting and Communicating Quality

In addition to documents such as this plan, monitoring protocols and SOPs, an overall assessment of data quality, and verification and validation procedures applied to each data set should be created and distributed along with the data. This report will be created by project leaders and will most often be incorporated into standard Federal Geographic Data Committee (FGDC) metadata records (see chapter 7 for a more in-depth discussion of data documentation). Communications of year-to-year changes in data quality and QA/QC procedures can be captured in monitoring project annual reports or project logs.

6.10. Credits

Information and concepts in this chapter were adapted from Mark Hart and Ulf Gavert, authors of the Great Lakes Network Data Management Plan (Hart et al. 2005).

Chapter 7. Data Documentation

Documenting data is the most important step toward ensuring that data sets are usable well into the future. Data longevity is roughly proportional to the comprehensiveness of their documentation (Michener 2000).

7.1. NPS Standards

Direction for documentation for federal agencies and the NPS are as follows:

- Executive Order 12906 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).
- FGDC CSDGM extensions, such as the Biological Data Profile, Remote Sensing Extension, and Shoreline Data Profile are not required, but recommended where appropriate.
- NPS Geographic Information System (GIS) Committee requires all GIS data layers be described with FGDC standards and the NPS Metadata Profile.

7.2. Network Standards

The ERMN has adopted the following data documentation standards for projects:

- Projects will be documented and organized in their appropriate "Projects" folder using standard folder organization as described in Chapter 4;
- Project finances will be tracked in the ERMN project tracking database as described in Chapter 4;
- Spatial data files will be documented in compliance with FGDC standards using the CSDGM; and
- Biological databases will be documented in compliance with FGDC standards using the NBII biological metadata profile.

7.3. Project Documentation

The following has been developed in the ERMN for inventory projects and data. Needs and standards for documenting monitoring projects will vary and will be addressed in the development of monitoring protocols and SOPs.

7.3.1. Project Organization

It is important that project files and communications are well organized and easily accessible. The "Projects" folder mentioned in Chapter 4 provides a guideline for project folder organization; however, there is still a lot of gray area left for users to decide where to put certain files. It is therefore incumbent on those involved to come up with a common approach to organizing these project files in a way that makes sense to those involved. One possibly unique approach that the ERMN has adopted is to export all important emails related to the project as a text document and save them in the "Communications" project folder in addition to any project email folders. In this way the network keeps all important project records and communications in

one place, the project's "Projects" folder. This folder is archived upon project completion and has been valuable in documenting the life of the project.

7.3.2. Project Tracking Database

Although it includes contact information and deliverable deadlines, the primary function of the ERMN project tracking database is to track funds, accounts, and invoices.

7.3.3. Spatial Data Documentation

The following are guidelines for spatial metadata records created in or for the ERMN:

- FGDC compliant metadata is developed for all spatial data files using the CSDGM. The metadata file should parse with no errors; however, this is sometimes more difficult to achieve than is worth the effort. Generally, less than 5 errors, none of which are associated with missing content, is what the network has treated as acceptable.
- In addition to FGDC standards and the NPS Metadata profile (including the standard NPS liability statement), the following NPS specific information must be included in ERMN metadata records:
    1. In the identification Section (Section One), the primary originator should be: "NPS NER Science Office". And the second originator should be the cooperator.
    2. In the Distribution Section (Section Six), the distributor should be the ERMN Data Manager.
    3. In the Metadata Reference Section (Section Seven), the author of the metadata should be identified (usually either the cooperator, ERMN Data Manager, or NCSU Center for Earth Observation).
    4. For sensitive data, the Standard Order Process in the Distribution Section needs to be revised to state that the data may be requested by contacting the ERMN Data Manager.
- Information on GPS settings and estimated accuracies are provided in the Data Quality element if GPS units were used to collect field information. See Appendix O for a helpful GPS Metadata Questionnaire Form that can be sent to cooperators.

7.3.4. Tabular Database Documentation

A required element of any ERMN database is a data dictionary. The data dictionary documents descriptions of each field in the words of its creator. It also includes fully written out abbreviations, acronyms, codes, or classes for the universe of possible entries used in the associated field (even if the code appears in more than one field). The data dictionary may also include ranges of values, if these are known. Data dictionaries are created either in the field description portion of the MS Access database when viewed in Design View, in an Excel table, or sometimes an MS Word document. What's important is that the information is there and is interpretable by anyone who should desire to use the database in the future.

In addition to the above, FGDC compliant biological metadata is developed for all biological ERMN projects using the NBII biological metadata profile. This is what the ERMN uses to document tabular biological project databases. It is also what the ERMN posts to the NPS Data Store and becomes the publicly searchable record for that project. In the case of projects with

sensitive information, this is the only thing that gets posted to the NPS Data Store. For other projects, FGDC metadata records are zipped up along with their associated spatial data and posted as a zip file along with the other project data files and posted for download (see Chapter 3 for more information).

Finally, especially complex tabular databases or those that are long-lived, such as monitoring databases, require additional documentation, such as:

- Revisions to the database;
- Overview of the use and purpose of the data;
- Illustration of the entity diagrams;
- Data dictionary of tables, attributes, and relationships; and
- Explanation of queries, forms, and reports.

7.4. Vital Signs Protocol Documentation

Vital signs protocols (protocol narratives and accompanying SOPs) constitute essential project documentation that must accompany the distribution of monitoring data. Over time there will be instances when the protocol narrative and SOPs will need to be updated or in other ways changed. Narrative and SOP updates may occur independently. That is, a change in one SOP will not necessarily invoke changes in other SOPs, or a narrative update may not require SOP modifications. The ERMN will develop a Change SOP that includes instruction on exactly how to change and document changes in protocols and SOPs. This documentation will, at a minimum, include the following information:

- The version of the protocol or SOP being revised;
- The date of the revision;
- The person doing the revising; and
- A short description of the purpose of the revision.

7.5. Additional ERMN Documentation

Not all Network data documentation needs will fall neatly into the above categories, especially as the ERMN monitoring program gets under way. Complex and multi-step analyses, algorithms, output, and other files may require some documentation to ensure that future users can understand the steps that were taken to create the file. “ReadMe.txt” files stored in a root directory are one way of documenting a file or folder and sub-folders. The file should provide a narrative statement that answers the following questions:

- What is in the file or folders?
- Who created it?
- How was it created?

The last question should be answered in sufficient detail that a similarly skilled individual as the original creator could reasonably recreate the files or files therein.

7.6. Credits

This chapter was adapted from ideas and concept presented by Dorothy Mortenson, author of the Southwest Alaska Network Data Management Plan (Mortenson 2005).

Chapter 8. Data Analysis and Reporting

Timely analysis and reporting is central to the efficacy and continued support of any long-term ecological monitoring program, and efficient data management strategies are an essential first-step in achieving this. A complete discussion of the ERMN approach to data analysis and reporting is presented in the ERMN Monitoring Plan (Marshall 2006). What is presented below is a very abbreviated summary of that Monitoring Plan chapter embellished with some data management points of significance.

8.1. Data Analysis

The ERMN Monitoring Plan (Marshall 2006) states that:

> *For the purposes of this program, we have defined data analysis as the processes by which observations of the environment are turned into meaningful information. We have defined "data analysis" broadly to include all evaluations of data after the data are collected and entered into an electronic file. Thus, data analysis includes quality control checks that occur during summarization and exploratory data analysis and extends through to analytical procedures leading to conclusions and interpretations of the data.*

When interpreted in this way, much of the preliminary data analyses in the ERMN monitoring program will either be performed, or in other ways directed, by the data management strategies adopted by each monitoring protocol. The goal here is to generate meaningful information, and by that is meant precise and accurate information. Data precision is limited by the methods, expertise, instrumentation, and other tools used to make field observations. Data accuracy, however, can be dramatically affected by data management strategies. The QA/QC concepts, methods, and strategies presented in Chapter 6 of this document are intended to do one thing: increase the accuracy, and, thereby, meaning, of information generated by this monitoring program.

8.2. Data Reporting

The ERMN Monitoring Plan continues with:

> *"…for the reporting end of the monitoring program to work well, all other parts of the monitoring program must also be functioning properly as part of an integrated information management system…"*

Indeed, timely reporting is often a function of the quality of a data management plan and its implementation. The ERMN will employ, whenever possible, automated routines for data analysis, summarization, and reporting. Along with those reports will be provided all appropriate documentation, including but not limited to metadata, in order that a user not familiar with our monitoring program and its methods would be able to evaluate our data and reports for their own use (see Chapter 7 for more information on data documentation). Applying sound data management concepts and methods, the ERMN will develop an integrated information management system for its monitoring program that will support efficient and effective reporting.

Chapter 9. Data Dissemination

One of the most important goals of the Inventory and Monitoring Program is to *integrate natural resource inventory and monitoring information into National Park Service planning, management, and decision making*.

To accomplish this goal, procedures must be developed to ensure that relevant natural resource data collected by NPS staff, cooperators, researchers, and the public are entered, quality-checked, analyzed, documented, cataloged, archived, and made available for management decision-making, research and education. Providing well-documented data in a timely manner to park managers is especially important to the success of the Program. The ERMN Program will make certain that:

- Data are easily discoverable and obtainable;
- No data that have not been subjected to full quality control are released;
- Distributed data are accompanied by complete metadata which clearly establishes the data as a product of the NPS I&M Program;
- Sensitive data are identified and protected from unauthorized access and inappropriate use; and
- A complete record of data distribution/dissemination is maintained.

9.1. Data Ownership

The National Park Service defines conditions for the ownership and sharing of collections, data, and results based on research funded by the United States government. All cooperative and interagency agreements, as well as contracts, should include clear provisions for data ownership and sharing as defined by the National Park Service:

- All data and materials collected or generated using National Park Service personnel and funds become the property of the National Park Service.
- 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.

The Office of Management and Budget (OMB) ensures that grants and cooperative agreements are managed properly. Federal funding must be disbursed in accordance with applicable laws and regulations. OMB circulars establish some degree of standardization government-wide to achieve consistency and uniformity in the development and administration of grants and cooperative agreements. Specifically, OMB Circular A-110 establishes property standards within cooperative agreements with higher institutions and non-profit organizations. Section 36 of Circular A-110, “Intangible Property” describes the following administrative requirements pertinent to data and ownership:

(a) The recipient (higher institution or non-profit organization receiving federal monies for natural resource inventory and/or monitoring) may copyright any work that is subject to copyright and was developed, or for which ownership was purchased, under an award. The Federal awarding agency(ies) (in this case the National Park Service) *reserve a royalty-free, nonexclusive and irrevocable right to reproduce, publish, or otherwise use the work for Federal purposes, and to authorize others to do so.*

Section 36 also states:

(c) The Federal Government has the right to:

(1) obtain, reproduce, publish or otherwise use the data first produced under an award

(2) authorize others to receive, reproduce, publish, or otherwise use such data for Federal purposes

(d) (1) *In addition, in response to a Freedom of Information Act (FOIA) request for research data relating to published research findings produced under an award that were used by the Federal Government in developing an agency action that has the force and effect of law, the Federal awarding agency shall request, and the recipient shall provide, within a reasonable time, the research data so that they can be made available to the public through the procedures established under the FOIA (5 U.S.C. 552(a)(4)(A)).*

(2) The following definitions apply for purposes of paragraph (d) of this section:
(i) Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This "recorded" material excludes physical objects (e.g., laboratory samples)…
(ii) Published is defined as either when:
(A) Research findings are published in a peer-reviewed scientific or technical journal; or
(B) A Federal agency publicly and officially cites the research findings in support of an agency action that has the force and effect of law.
(iii) Used by the Federal Government in developing an agency action that has the force and effect of law is defined as when an agency publicly and officially cites the research findings in support of an agency action that has the force and effect of law.

To ensure that proper ownership, format, and development of network products is maintained, all cooperative or interagency work must be conducted as part of a signed collaborative agreement. Every cooperative or interagency agreement or contract involving the Central Alaska 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, <u>institution or organization name</u> *shall follow the procedures and policies set forth in OMB Circular A-110."*

Cooperative or interagency agreements or contracts must include a clearly defined list of deliverables and products. Details on formatting and media types that will be required for final submission must be included. Typical products include, but are not limited to, field notebooks, photographs, specimens, raw data, and reports. See Chapter 3 and Appendixes A, B, and C for more information on NER product specifications.

Cooperative agreements and contracts should also provide a schedule of deliverables that includes sufficient time for NPS review of draft deliverables before scheduled final submissions. The ERMN typically withholds, for new cooperators with whom the network is unfamiliar, 10% of the total worth of a cooperative agreement to be invoiced after the project products are accepted by the ERMN as final, and 5% of total funds for cooperators with whom the network already has an established relationship and are confident that theywill receive acceptable project products.

9.2. Data Sensitivity

All data and information from I&M activities must be assessed to determine their sensitivity. This includes, but is not limited to, reports, metadata, raw and manipulated spatial and non-spatial data, maps, and others. Though it is the responsibility of network staff (as stewards of these data) to initiate the process of identification of data as sensitive or non-sensitive, rarely do they have the expertise and knowledge of current resource and social conditions to make that decision themselves. Network staff should work closely with park natural resource staff to assess data sensitivity. It then becomes the responsibility of network staff to define, document, and manage data as sensitive or non-sensitive and make these data available or protect them, as is appropriate.

9.2.1. Identifying Data as Sensitive

The Freedom of Information Act, 5 U.S.C. § 552, referred to as FOIA, stipulates that the United States Government, including the National Park Service, must provide access to data and information of interest to the public. FOIA, as amended in 1996 to provide guidance for electronic information distribution, applies to records that are owned or controlled by a federal agency, regardless of whether or not the federal government created the records. FOIA is intended to establish a right for any person to access federal agency records that are not protected from disclosure by exemptions. Under the terms of FOIA, agencies must make non-protected records available for inspection and copying in public reading rooms and/or the Internet. Other records, however, are provided in response to specific requests through a specified process. The Department of the Interior's revised FOIA regulations and the Department's Freedom of Information Act Handbook can be accessed at www.doi.gov/foia for further information.

In some cases, public access to data can be restricted. Under one Executive Order, Director's Order #66 (draft), and four resource confidentiality laws, the National Park's Omnibus Management Act (16 U.S.C. 5937), the National Historic Preservation Act (16 U.S.C. 470w-3), the Federal Cave Resources Protection Act (16 U.S.C. 4304), and the Archaeological Resources Protection Act (16 U.S.C. 470hh), the National Park Service is directed to protect information about the nature and location of sensitive park resources. Through these regulations, information

that could result in harm to natural resources can be classified as 'protected' or 'sensitive' and withheld from public release (National Parks Omnibus Management Act (NPOMA)).

The following guidance for determining whether information should be protected is in the draft Director's Order #66 (the final guidance may be contained in the Reference Manual 66):

- Has harm, theft, or destruction occurred to a similar resource on federal, state, or private lands?
- Has harm, theft, or destruction occurred to other types of resources of similar commercial value, cultural importance, rarity, or threatened or endangered status on federal, state, or private lands?
- Is information about locations of the park resource in the park specific enough so that the park resource is likely to be found at these locations at predictable times now or in the future?
- Would information about the nature of the park resource that is otherwise not of concern permit determining locations of the resource if the information were available in conjunction with other specific types or classes of information?
- Even where relatively out-dated, is there information that would reveal locations or characteristics of the park resource such that the information could be used to find the park resource as it exists now or is likely to exist in the future?
- Does the NPS have the capacity to protect the park resource if the public knows its specific location?

Natural Resource information that is sensitive or protected requires the:

- Identification of potentially sensitive resources;
- Compilation of all records relating to those resources;
- Determination of what data must not be released to the public; and
- Management and archival of those records to avoid their unintentional release.

Classification of sensitive I&M data will be the responsibility of the ERMN staff, park staff, and investigators working on individual projects, although it is ultimately park natural resource staff working with park management and their park FOIA officer to determine whether data should be classified as sensitive or not. Data will be classified as sensitive on a case-by-case, project-by-project basis. Network staff will work closely with investigators and park staff for each project to ensure that potentially sensitive park resources are identified, and that information about these resources is tracked throughout the project.

The Network staff is also responsible for identifying all potentially sensitive resources to project leaders working on each project. The project leaders, whether network staff or partners, will develop procedures to flag all potentially sensitive resources in any products that come from the project, including documents, maps, databases, and metadata. When submitting any products or results, investigators should specifically identify all records and other references to potentially sensitive resources. Note that partners should not release any information in a public forum before consulting with network staff to ensure that the information is not classified as sensitive or protected. Though this may sound like a straightforward endeavor, NPS cooperators and

partners often post data to clearinghouses and other public data repositories as part of their project or annual data management routine. Each of these repositories classifies and protects sensitive data in different ways and it is important that cooperators and partners understand that data that result from NPS funded projects are the property of the NPS and must be protected according to NPS and other federal policy, law, and regulation.

The following strategies will be employed by network staff, as appropriate:

- Project leaders, whether network staff or partners, will develop procedures to flag all potentially sensitive resources in any products that come from the project.
    - Park staff are always consulted in this process and given final say in data or record classification; it is ultimately their resource to protect, or chose not to protect.
- Network staff will define and document data as sensitive on a case-by-case, project-by-project basis.

For example, information may be withheld regarding the nature and/or specific locations of the following resources recognized as 'sensitive' by the National Park Service. According to NPOMA, if the NPS determines that disclosure of information would be harmful, information may be withheld concerning the nature and specific location of:

- Endangered, threatened, rare, or commercially valuable National Park System Resources (species and habitats);
- Mineral or paleontological objects;
- Objects of cultural patrimony; and
- Significant caves.

Note that information already in the public domain can, in general, be released to the public domain. For example, the media has reported in detail the return of condors to the Grand Canyon. If an individual requests site-specific information about where the condors have been seen, this information, in general, can be released. However, the locations of specific nest sites cannot be released. It is also often the case that other groups and agencies have similar sets of information and are subject to different regulations and policy. Some state departments of conservation and natural resources for example, are subject to state specific "Sunshine Laws" that require them to release information under certain circumstances upon requests from the public. Groups interested in accessing information on sensitive resources will often submit requests to multiple agencies, and in these cases a FOIA request submitted to the NPS is sometimes rendered inconsequential because another group or agency has already released the data.

9.3. Access Restrictions to Sensitive Data

ERMN staff is responsible for managing access to sensitive data handled by the program. All potentially sensitive park resources will be identified and investigators working on network projects will be informed that:

- All data and associated information must be made available for review by network staff prior to release in any format.

- Any information classified as protected should not be released in any format except as approved in advance by the National Park Service.

The Network Coordinator, NPS park staff liaison, or Data Manager identifies all potentially sensitive park resources to the project leader for each project. Reciprocally, the project leader must identify any known references to potentially sensitive park resources to others.

When preparing or uploading any information to a program or other database or repository, ERMN staff ensures that all protected information is properly identified and marked as such. All references to protected information are removed or obscured in any reports, publications, maps, or other public forum. This can sometimes be an issue, since park staff, being the most familiar with threats to resources, are often the best people to do this editing. However, asking them to spend significant amounts of time editing I&M reports is often not a feasible approach. ERMN staff will work closely with park staff (when they are the ones doing the obscuring) to edit materials for public release.

Network staff will remove any sensitive information from public versions of documents or other media. They will isolate sensitive from non-sensitive data and determine the appropriate measures for withholding sensitive data. The main distribution applications and repositories developed by the I&M Program are maintained on both secure and public servers and all records marked 'sensitive' during uploading will only become available on the secure servers. Procedures for assigning a sensitivity level to specific records when uploading to both the NPSpecies and NatureBib databases are given on the following websites and in Appendixes F and G.

All references to protected information are removed or obscured in any reports, publications, maps, or other information to be released to the public.

http://science.nature.nps.gov/im/apps/npspp/index.htm
http://www.nature.nps.gov/nrbib/index.htm

Thus, access to data on sensitive park resources can be limited to network staff or research partners with access to NPS records. However, limits to how these data are subsequently released must also be clearly defined. It is crucial that the person uploading records to the NPS online applications (repositories) is familiar with the procedures for identifying and entering protected information.

The topic of how to handle sensitive data and records is a complex one that's evolving quickly in the NPS, and federal government, in general. In cases where tools do not yet exist (e.g. with the NPS Data Store), or where clear guidance is not yet developed, ERMN staff will work closely with park staff and the NER regional natural resource science office to develop conservative and robust solutions to adopt for the near-term future.

9.4. Data Distribution Mechanisms

The Network's main mechanism for distribution of inventory and monitoring data will be the Internet. Use of the Internet will allow the dissemination of data and information to reach a broad community of users. As part of the NPS I&M Program, Web-based applications and repositories have been developed to store a variety of park natural resource information. The Network will use the following applications and repositories to distribute data developed by the program (see Chapter 4 for Web addresses for each of these applications as well as more information):

- NatureBib–is a master Web-based database housing natural resource bibliographic data for I&M Program parks including the ability to upload and download full-text pdf documents.
  - The ERMN converts all reports to meet NER Technical Reporting Guidelines (see Appendixes B and C), with slight modifications where it meets ERMN specific needs. Full-text versions of reports are posted to NatureBib and hard and digital copies are mailed to the Technical Information Center (TIC) in Denver, CO.
  - Documents found as a result of Network data mining are posted to NatureBib where copyright and other publisher agreements allow for this distribution.
  - Sensitive data: the ERMN has produced duplicate reports for projects with large amounts of sensitive information. The public version of these reports can be posted to NatureBib, labeled under "Sensitivity Level" as public for both the Citation and Abstract and made available as full-text downloads. A separate version of the report will have a NatureBib record labeled under "Sensitivity Level" as "NPS Only" and will not be searchable or viewable by the public. It's also possible to identify a citation as publicly viewable and its abstract as NPS only.
- NPSpecies–is a master Web-based database to store, manage, and disseminate scientific information on the biodiversity of all organisms in all I&M National Park units.
  - Sensitive Data: there are a number of different ways to identify NPSpecies records as sensitive. Each reference, observation, and voucher has a sensitivity level associated with it, either "NPS Only" or "Public." Each species under "Management Information" also has a "Management Priority" associated with it that can be manipulated to reflect an appropriate level of sensitivity. The latter provides users with a way to search species by park and by sensitivity level.
- NPS Data Store–is an agency wide online repository for metadata and associated products that spans and is supported by many NPS programs. The NPS Data Store is the primary repository for static ERMN products.
  - ERMN project products are posted to the NPS Data Store with data products and their associated documentation zipped together in one file and the project report and associated documentation and products in another file. Biological metadata records are posted as the searchable metadata for the project record.
  - Sensitive Data: At the moment, the NPS Data Store does not have any built in tools or functionality to handle sensitive data products. The ERMN approach to this has been to post only metadata for projects that have sensitive data associated with them. In these cases, the Standard Order Process section of the FGDC Distribution Information metadata element must be filled out with the ERMN Data Manager's contact information and how someone would request a copy of these data.

    - Note: The ERMN process to post any project information to the NPS Data Store involves the following:
        1. If there is no known sensitive information associated with the project, perform the NPSpecies search for "Management Priority.".
        2. Email the park contact for that project with the results of the NPSpecies search and ask for permission to post to the NPS Data Store.
        3. Document park staff's response giving permission to post in the "Projects" folder.
        4. Post project products to NPS Data Store.
- NPSFocus–is an agency wide search engine and image management server being developed by the NPS Information and Telecommunications Center (ITC) that's linked to a number of other NPS repositories including the NPS Data Store and ETIC (see next mechanism below). Individuals can steward their own data on this server and it has the capability to handle very large datasets.
    - Sensitive Data:
- Electronic Technical Information Center (ETIC)–The NPS Technical Information Center in Denver, CO maintains an online repository of a subset of their information.
    - Sensitive Data: TIC sometimes post full-text downloadable copies of reports so it's important that they're notified of sensitivity levels. They also have the ability to post the InsideNPS Only.
- Northeast Region (NER) Philadelphia Science Office (PHSO)–the PHSO posts online full-text downloadable versions of reports produced in the NER Science program to their Web site (see Chapter 4 for address). Betsy Bloomberg (a Pennsylvania State University cooperator) maintains this site under the direction of the Chief Scientist of the Region. At the moment, this site is not searchable in any way.
    - Sensitive Data: The Chief Scientist is usually involved enough in ERMN projects that he's well aware of project sensitivity.
- USGS NPS Vegetation Mapping Website–NCSU staff are working with the USGS vegetation mapping program to post final vegetation mapping data to a USGS Website in USGS approved formats.
- Eastern Rivers and Mountains Network Website–provides detailed information about the Network and the I&M Program. Originally envisioned as a place to post Network data, the ERMN website is no longer being developed in this capacity. The ERMN Website has information on ERMN projects, cooperators, and partners, the Board of Directors (BOD) and Science Advisory Committee (SAC), photos of staff and Network activities, and downloadable Network and programmatic administrative documents.
    - Sensitive Data: the ERMN does not anticipate distributing sensitive data though its Website. Keep in mind, however, restrictions on posting photographs of people to a federally maintained Website.

At the moment, there is significant duplicability and a versioning issue in how ERMN reports are posted to the Internet and made searchable and available to the public. The ERMN will coordinate an effort to post each Network technical report only once, and redirect all remaining addresses with links to the one actual downloadable version maintained by the Network.

9.4.1. Other Online Tools for Data Distribution

The ERMN has undertaken two other exploratory projects to develop online tools for viewing, manipulating, and, in some cases, analyzing and downloading network data. These projects could fill a gap left by the other tools in this section, in that all of the above tools are meant for static and completed products, whereas an active monitoring program will produce dynamic data on a much shorter time-scale than could be served by the above tools. The ERMN would like to develop an enterprise level, Web-accessible database that makes some of its monitoring data available on a real-time or near real-time basis.

The first project includes a custom Adobe flash-based map with mySQL database running in the background that is currently being used to display ERMN weather station metadata and is housed on a PSU server maintained by the PA State Climatologist's office. The Network has also contributed funds to a NPS NRPC ESRI ArcIMS development project that seeks to customize the ArcIMS interface to add functionality and meet more of Networks' Internet mapping needs.

Finally, there appears to be great promise in ESRI's newest ArcGIS Server 9.2 that may also meet many of the Network's dynamic data needs.

9.5. Feedback Mechanisms

The ERMN will develop standard procedures for park staff, cooperators, partners, and the public to report errors or other anomalies in Network data and information, as well as errors, glitches, and other problems with the Network's Web-based tools and presence.

9.6. Credits

Material in this chapter was adapted from both Doug Wilder, author of the Central Alaska Data Management Plan (Wilder 2005), as well as from Sara Stevens and Gary Entsminger's North Coastal Barrier Network's Data Management Plan (Stevens et al. 2004).

Chapter 10. Data Maintenance, Storage and Archiving

Long-term maintenance and management of digital information are vital to the Inventory and Monitoring Program. This chapter describes procedures for maintaining and managing digital data, documents, and objects that result from Network projects and activities. These procedures will help ensure the continued availability of crucial project information and permit a broad range of users to obtain, share, and properly interpret that information.

Effective long-term data maintenance depends on thoughtful and appropriate data documentation. An essential part of any archive is its accompanying explanatory materials. This chapter will refer to, and in some cases elaborate on, metadata standards and dataset documentation procedures that are more fully explained in Chapter 7–Data Documentation–of the NCBN Data Management Plan.

The ERMN has entered into a cooperative agreement with the North Carolina State University NPS Field Technical Support Center (NCSU FTSC), Center for Earth Observation, for the maintenance, storage, and archiving of datasets produced as part of the Network's inventory, vegetation mapping, and vital signs monitoring program. This is a long-term archive used principally for safe-guarding against the loss of information and objects. It is not, however, a viable day-to-day or week-to-week back-up or accessible archive. The future of this NCSU archive is uncertain, as there are other NPS digital archives in development that in the future will be free and available to the Network. The remainder of this chapter is written in a 'business as usual' manner in regards to the ERMN agreement with NCSU, recognizing that this well may change in the near-term future.

10.1. Digital Archiving and Storage

Digital data maintained in the ERMN data archive at NCSU is expected to result from one of two types of projects:

1. short-term projects, after data collection and modification have been completed (i.e., biological inventory and vegetation mapping projects); and
2. long-term monitoring projects, where data acquisition and entry will continue indefinitely. ERMN will remain current and compatible with National Park Service and National Inventory and Monitoring version standards for these software programs.

Technological obsolescence is a significant cause of information loss, and data can quickly become inaccessible to users if they are stored in out-of-date software programs or on outmoded media. Effective maintenance of digital files depends on the proper management of a continuously changing infrastructure of hardware, software, file formats, and storage media. Major changes in hardware can be expected to occur every 1-2 years and in software every 1-5 years. As software and hardware evolve, datasets must be consistently migrated to new platforms, or they must be saved in formats that are independent of specific platforms or software (e.g., ASCII delimited files). Thus, ERMN archiving procedures include saving datasets in both their native format (typically MS Access database or Excel spreadsheet format) and as sets of ASCII text files. As a platform- and software-independent format, ASCII text files ensure

future usability of the data in a wide range of applications and platforms. In addition, datasets will periodically be converted to upgraded versions of their native formats.

10.1.1. Short-term Projects

Short-term projects include biological inventory studies, vegetation and fire fuel mapping projects, and other studies where data are collected during a relatively short, finite period of time. Digital data products from a completed short-term project typically include tabular data such as MS Access databases or Excel spreadsheets and documentation; spatial data such as shapefiles, coverages, georeferenced images, etc., and documentation; a final report and an NPS technical report. Data maintenance procedures described in this section apply to tabular data. Procedures for spatial data are presented in a separate following section.

When a project is completed (including completion of all QA/QC procedures, see Chapter 6), NCSU FTSC staff prepare the tabular data for archiving by creating:

- a set of ASCII comma-delimited text files for the tabular data files and tables comprising the dataset;
- an XML file that preserves relationships between tables for each MS Access database; and
- a readme.txt file that explains the contents of each ASCII file, file relationships, and field definitions.

Quality control checks are then performed on these ASCII files to ensure that the numbers of records and fields correspond to the source dataset and that conversion has not created errors or data loss. If possible, a second reviewer, preferably a program scientist, checks the ASCII files and documentation to verify that tables, fields, and relations are fully explained and presented in a way that is useful to secondary users.

In addition to creating and archiving ASCII files, archived tabular data in older software versions will be updated periodically, with the goal of having no data more than two versions behind the current version used by ERMN. There is a risk of reducing performance as a result of conversion; for example, complex data entry forms or reports may not function properly in an upgraded software version. To the extent possible, proper functionality of data entry forms and reports will be maintained. However, our priority will be to ensure basic table and relationship integrity.

Folder and file names assigned to converted/upgraded databases will clearly indicate the revision number and date. All previous versions of the dataset will be retained in the archive.

10.1.2. Long-term Projects

Data produced as part of the Network's long-term monitoring projects will have variable archiving requirements that depend upon the data types collected. Modifications to protocols and their SOP's will require complete datasets to be archived before modifications are implemented. Depending on the monitoring project, a copy of the dataset will be archived on an annual basis. Whenever possible, archived datasets will be saved in their native formats as well as ASCII text

files. Data archiving requirements for ongoing projects will be spelled out in the data management SOPs for each monitoring project.

Long-term monitoring datasets will include both tabular data format as well as spatial data depending on the protocol. Network staff are responsible for preparing tabular data from the Network long-term monitoring projects for archival following completion of standard QA/QC procedures. The procedures listed earlier under "Short-term Datasets" will also be used for long-term datasets.

Long-term monitoring datasets will also require periodic conversion to current database formats. The Network data manager and project leaders are responsible for determining when to convert a particular project dataset to a newer software version. This decision will depend on specific project needs and data collection schedules. However, our goal is to prevent any dataset from falling more than two versions behind the current version in use by the Network.

If the Network decides to convert tabular data from one software version to an upgraded version, Network staff are responsible for performing the conversion. Network staff are also responsible for:

- performing quality control checks to insure that all forms, queries, reports, and data entry function as intended, particularly if the database is used for data entry or analysis.
- preparing metadata for the converted/upgraded database, including revision information and history in tables within each database file.

File names assigned to converted/upgraded databases should clearly indicate the revision number and date. All previous versions of datasets will be retained in the archive.

10.1.3. Spatial Data

Digital spatial data associated with short- or long-term project data described above will also be archived at the NCSU FTSC. These data include ESRI shapefiles, coverages and geodatabases; global positioning system (GPS) data; and digital aerial imagery. Procedures for maintaining spatial data differ from those that apply to tabular data.

Spatial data in ESRI formats (shapefiles, coverages, and geo-databases) will be maintained in formats that are accessible by the current ESRI software version. ESRI software has maintained compatibility with previous data formats, and while shapefiles have retained all functionality in current format, coverages may require conversion to a newer ArcGIS format if they are no longer supported.

In addition to archiving ESRI spatial data in its native format, shapefiles, coverages, and geodatabases will be converted to sets of ASCII text files for archiving. Two ASCII files are required for each spatial data file. One file stores coordinate data with a unique identification code for each feature, and the second file stores attribute information that can be linked to the corresponding coordinate data by means of a common identification code.

Only corrected or downloaded point GPS data will be archived. However, raw unsaved active track log data for track or polygon files will be saved as ASCII text files in addition to their final and processed ESRI file formats.

Digital aerial imagery is archived in its original native format–e.g., tagged image file (TIF) format. Derived products, such as ortho-rectified image files, are archived in their software-dependent format–e.g. Imagine (IMG) format–and in TIF format with a corresponding world file. The formats in which photo mosaics are archived depend on the file size:

- Mosaics smaller than 2 gigabytes will be archived in their software-dependent format (e.g., IMG files), as TIF files with corresponding world files, and, if provided, in Mr. SID (compressed) format.
- Mosaics 2 gigabytes or larger will be archived in their software-dependent format (e.g., IMG files), in Band-Interleaved-by-Line (BIL) format with corresponding header files, and, if provided, in Mr. SID (compressed) format.

All TIF files must be full resolution with no compression. TIF files with corresponding world files are preferred over GeoTif format.

10.2. Digital Storage and Archiving Procedures

The archiving procedures described below are designed to ensure both the security of and ready access to ERMN data in perpetuity. These procedures apply to both short-term and long-term monitoring project datasets (tabular and spatial data) and to digital aerial imagery acquired by the ERMN.

When a project dataset or a set of digital aerial imagery is ready to be archived, Network staff will send the data to NCSU FTSC accompanied by a transmittal document in the form of an ASCII text file that includes specific information about the data or imagery.

For short-term or long-term monitoring project datasets, this information includes:

- name and address of the sender;
- brief project description, including topic and park name(s);
- type and number of digital media (e.g., 3 CDs, 1 DVD); and
- name of each data file that comprises the dataset (Federal Geographic Data Committee (FGDC) compliant metadata or a data dictionary must be present for each data file).

For digital aerial photography:

- name and address of the sender;
- type and number of digital media (e.g., 4 CDs, 2 DVDs, 1 firewire drive);
- park name(s); and
- number of image data files (per park).

Upon receipt, NCSU FTSC staff will:

- check the digital media for physical damage;
- count and record the number of media received;
- count and record the number of data/image files received; and
- check data file integrity (i.e. makes sure that each file opens and displays appropriate data).

ERMN staff must be notified if any problems or discrepancies are found.

NCSU staff will enter the information about short-term and long-term monitoring project data into a tracking database. Information about sets of digital aerial photography should be entered into a similar spreadsheet.

Copies of the two tracking databases will be sent to the Network once a year, as requested, or will be posted on the Internet. In future, these databases will be posted on the Internet for further accessibility to Network staff. These databases will be backed up following the procedures used for backing up archived data described below.

Once all of the above procedures have been completed, NCSU FTSC staff will archive the original data media in a fireproof cabinet, in a dark, temperature-controlled environment. All procedures meet the NPS Museum Management Handbook standard for archiving.

10.2.1. Storage Facilities for Electronic Archives

Currently, ERMN data are archived at the NC State enterprise storage system, which meets State of North Carolina certified critical data standards. These servers have uninterruptible power supply backup and are housed in a temperature-controlled room with secure access. The servers are backed up externally by the NC State University High Performance Computing Center.

10.2.2. Directory Structure for Electronic Archives

The ERMN electronic data archives directory structure at NCSU is organized by park, and for each park, by project. Each park folder contains one subfolder for each project.

The organizing and naming of project folders and files should be intuitive enough that users unfamiliar with the project can still easily navigate them. Since each project has its own variations and idiosyncrasies, a standardized structure is not realistic. However, all project archives will include several to most of the following elements:

- programmatic documents including protocols, procedures, and supporting documents;
- interim datasets or milestones;
- datasets submitted by contractors;
- datasets–in ASCII format;
- conceptual or statistical models used for data interpretation;
- final and NPS technical reports; and
- README files–including an explanation of directory contents, project metadata (including a dataset catalog report), and version documentation.

10.2.3. Backup Procedures for Digital Data

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 security breaches and vandalism. Performing regular backups of data and arranging for off-site storage of backup sets are the most important safeguards against data loss.

The following represents the temporary backup plan implemented by the network until a more permanent and efficient plan can be developed. Currently all ERMN network staff (and associated cooperators) have external hard drives with the capacity to back up all user data on their machines on a weekly basis, replacing the entire prior backup file each time. In addition to this, the network has one larger external hard drive onto which network staff back up their machines on a monthly basis. This larger hard drive is taken home by the Network Data Manager for a temporary off site storage back-up. Finally, all network staff are encouraged to backup their software files occasionally and create start-up floppy disk to restore their operating system if they should suffer a windows failure. All of the above backups are performed with Microsoft's system tools backup utility.

10.2.4. Data and Network Security

As of January 2005, ERMN data archives maintained by NCSU are stored on servers maintained by the Information Technology Division (ITD) located at 2620 Hillsborough Street, Raleigh, North Carolina, 27695. The ITD is open from 7:00 AM – 6:00 PM Monday through Friday. The building may be accessed only from the main entrance; all other entrances are locked to the general public. A university identification card is required to enter the building and must be displayed to a security guard. Non-university employees are required to provide picture identification and sign in/out of the building. There are ITD employees at the facility 24 hours a day, seven day a week to ensure operability. Additional information may be obtained from: http://sysnews.ncsu.edu.

Access to ERMN data on the NCSU ITD servers is restricted to a limited number of designated, authenticated employees.

10.3. ERMN Hardcopy and Digital Reports

ERMN reports are archived in two locations:

- the ERMN physical archive; and
- the Technical Information Center (TIC) Denver, CO.

The ERMN maintains an onsite physical archive for projects that contains the following, some of which are physical media containing digital copies of information:

- A CD with all final products from NCSU (the same as those in the NCSU digital archive);
- One bound published hardcopy final technical report;

- All hardcopy administrative records associated with the project including signed Cooperative and Task agreements, printouts of invoices from the project tracking database (see Chapter 8);
- A CD with the "Projects Folder" that contained all digital administrative records, draft data, interim products, progress reports, and correspondences; and
- Copies of all field forms, notes, and other hand written documents associated with the project.

TIC provides temperature and humidity-controlled facilities, a professional archival staff, and meets all museum standards set by the NPS. This repository is used for all reports published under the NPS technical reporting series; however, the ERMN has chosen not to publish some reports to the technical reporting series because of sensitivity concerns in some cases, and the reports' quality or value not being worth the effort in others.

10.4. Specimens

The Network will provide specimens collected under the auspices of the ERMN to the Network Park in which they were collected for inclusion in a park museum or to a repository approved by the Park. Specimens at non-NPS repositories are considered on loan and remain property of the NPS. ERMN will provide Park curators with associated data required for cataloging each specimen. These data are exported from NPSpecies where they're cataloged for Network purposes (see Appendix P for NPSpecies Museum Voucher Fields). Data will be provided to non-NPS curators in Excel format.

See Appendix Q for proposed additions to the NPS Museum handbook on how to handle multi-park projects (as is commonly the case with Network projects).

10.5. Photographs

Archivists have been reluctant to fully embrace digital photography and some have expressed concern that, with the acceleration of technological change, documentary heritage is in danger of being lost in the information age.

The ERMN prefers all images (photographs and scanned documents) are delivered as TIFF files as is specified in the NPS Focus manual (see the NPSFocus Website for more information).

Though the ERMN doesn't currently have an image management system in place, it will eventually develop and adopt one that preferably allows search ability of image subject and project and tracks the location of image files on local machines or servers.

10.5.1. Aerial Photographs

Analog aerial photographs produced by the Network and designated for archiving will be sent to the NCSU FTSC for archiving along with the following documentation:

- The name and address of the sender;
- The park name(s); and
- The number of photos transmitted (for each park).

Then NCSU staff will proceed as follows:

Upon receipt, NCSU FTSC staff count the aerial photos received and check each photo for physical damage. If the number of photos does not match the transmitted number, or if any photos are physically damaged (e.g. torn, water stained, etc.), NCSU will notify the Network.

NCSU enters information about the set of aerial photos (as appropriate and as it becomes known) into a tracking database. NCSU makes this database available upon request, sends to the Network on a regular schedule, or posts on the Internet.

Aerial photographs are then stored in plastic sleeves, lying flat, in a fireproof cabinet, in a dark, cool, temperature-controlled environment.

Table 10.1 summarizes the much of the above information into a table of the principle repositories for ERMN project digital and hard copy products.

10.6. Credits

Much of the information from this chapter was adapted the North Coastal Barrier Island Network Data Management Plan written by Sara Stevens and Gary Entsminger (Stevens et al. 2004).

Table 10.1. Principal ERMN repositories for project products.

Project Products	Repository
Tabular Data and Documentation	
Final Products	NCSU Digital Archive
	ERMN Physical Archive (on media)
	NPS Data Store
	NPSpecies
Draft Products	ERMN Physical Archive (Projects Folder on media)
Spatial Data Final Products	 NCSU Digital Archive ERMN Physical Archive (on media)
	NPS Data Store
Draft Products	ERMN Physical Archive (Projects Folder on media)
Technical Reports digital	 NPS Data Store, NPS Focus, NERO Science Website, E-TIC
hard copy	ERMN library
	Park library
	ERMN physical archive TIC
bibliography Photographs Aerial hard copy Aerial digital Project Related	NatureBib NCSU physical archive NCSU digital archive NCSU digital archive ERMN physical archive (on media)

Chapter 11. Literature Cited

Beer, M., E. Nance, A. Wight, M. Powell, and R. DenBleyker. 2005. Northern Colorado Plateau Network, data management plan. National Park Service. Inventory and Monitoring Program. Northern Colorado Plateau Network. Moab, UT. 86 p. plus appendices.

Boetsch, J. R., B. Christoe, and R. E. Holmes. 2005. Data management plan for the North Coast and Cascades Network Inventory and Monitoring Program. USDI National Park Service. Port Angeles, WA. 88 pp.

Daley R. 2005. Data and Information Management Plan, Greater Yellowstone Inventory and Monitoring Network. Bozeman, MT. National Park Service. Greater Yellowstone Network. 82 p. plus appendices.

Edwards, Don. 2000. Data quality assurance. Pp. 70-91 *in* Michener, William K., and James W. Brunt (eds). 2000. Ecological Data: Design, Management and Processing, Methods in Ecology Series. Malden, MA. Blackwell Scientific, Ltd.

Marshall, M. R., and N. B. Piekielek. 2006. Eastern Rivers and Mountains Network Ecological Monitoring Plan. Technical Report NPS/NER/NRTR—XXXX/XXX. National Park Service. Philadelphia, PA.

Michener, William K., and James W. Brunt. 2000. Ecological Data: Design, Management, and Processing. London, England: Blackwell Scientific, Ltd.

Miller, Abigail B. 2001. Managing data to bridge boundaries. Pp. 316-320 *in* Harmon, David (ed). Crossing Boundaries in Park Management. Proceedings of the 11th Conference on Research and Resource Management in Parks and on Public Lands. Hancock, MI. The George Wright Society.

Southwest Alaska Network. 2005. Data management plan for the Inventory and Monitoring Program, Southwest Alaska Network. National Park Service. Anchorage, AK. 234 pp.

Palmer, Craig J. 2003. Approaches to quality assurance and information management for regional ecological monitoring programs. Pp. 211-225 *in* Busch, David E., and Joel C. Trexler (eds). 2003. Monitoring Ecosystems: Interdisciplinary Approaches for Evaluating Ecoregional Initiatives. Washington, DC. Island Press.

Stevens, Sara M., and Gary Entsminger. 2004. North Coastal Barrier Network Inventory and Monitoring Program Information Management Plan. National Park Service. Boston, MA.

Wilder, D.,editor. 2005. Data Management Plan, Central Alaska Network. National Park Service. Inventory and Monitoring Program. Fairbanks, AK. 86 pg, 12 appendices.

As the nation's primary conservation agency, the Department of the Interior has responsibility for most of our nationally owned public land and natural resources. This includes fostering sound use of our land and water resources; protecting our fish, wildlife, and biological diversity; preserving the environmental and cultural values of our national parks and historical places; and providing for the enjoyment of life through outdoor recreation. The department assesses our energy and mineral resources and works to ensure that their development is in the best interests of all our people by encouraging stewardship and citizen participation in their care. The department also has a major responsibility for American Indian reservation communities and for people who live in island territories under U.S. administration.

NPS 962/100731 December 2006 (Revised December 2009)

National Park Service
U.S. Department of the Interior

Northeast Region
Natural Resource Stewardship and Science
200 Chestnut Street
Philadelphia, Pennsylvania 19106-2878

http://www.nps.gov/nero/science/

EXPERIENCE YOUR AMERICA ™