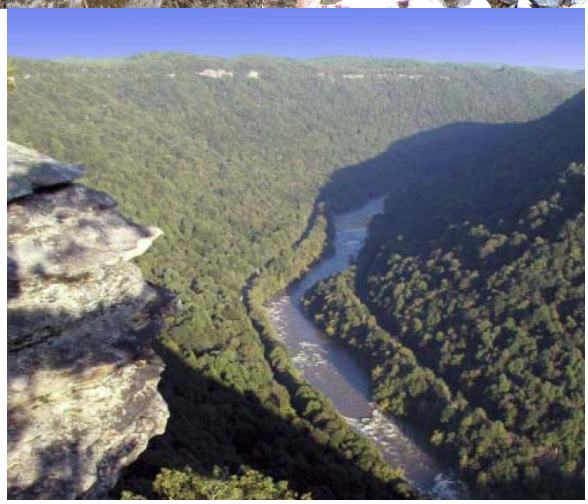
National Park Service
U.S. Department of the Interior

Northeast Region
Philadelphia, Pennsylvania

Eastern Rivers and Mountains Network
Data Management Plan

Technical Report NPS/NER/NRTR— (DRAFT) XXXX/XXX



ON THE COVER
A US 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

Technical Report NPS/NER/NRTR— (DRAFT) XXXX/XXX

Nathan Piekielek

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

December 2006

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 D-XXX.

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. December 2006. Eastern Rivers and Mountains Network Data Management Plan. Technical Report NPS/NER/NRTR—XXXX/XXX. National Park Service. Philadelphia, PA.

NPS D-XXX December 2006

Table of Contents

Page

Figures vii

Tables viii

Appendixes ix

Executive Summary xi

Acknowledgments xxii

Chapter 1 Introduction 25

1.1. Eastern Rivers and Mountains Network 26

1.2. ERMN Data Management Goals and Objectives 26

1.3. Data Defined 27

1.4. Sources of Natural Resource Data 28

1.5. ERMN Data Management Plan Scope and Organization 29

1.6. Intended Audience 30

1.7. Revisions 30

1.8. Credits 30

Chapter 2 Data Management Roles and Responsibilities 31

2.1. Definition of Roles versus Responsibilities 31

2.2. Roles and Responsibilities of Network Staff 31

2.3. Coordination 37

2.4. Documentation 39

2.5. Awareness and Accountability 39

2.6. Communication 40

2.7. Credits 40

Chapter 3 Data Life Cycle and Work Flow 41

3.1. Project Work Flow 41

3.2. Data Life Cycle 44

3.3. Integrating and Sharing Products 48

3.4. Credits 50

Chapter 4 Infrastructure and System Architecture 51

4.1. Network and System Resources 51

4.2. ERMN Digital File Storage and Maintenance 52

4.3. National Information Management System Architecture 53

4.4. Credits 56

Chapter 5 Data Acquisition and Processing 57

5.1. Eastern Rivers and Mountains Network Data 57

5.2. Other National Park Service Data 60

5.3. External Data 61

5.4. Credits 61

Chapter 6 Quality Assurance and Quality Control (QA/QC) 63

6.1. Quality Assurance and Quality Control Defined 63

6.2. General Operations 64

6.3. Monitoring Protocol Development 66

6.4. Data Collection 66

6.5. Data Entry and Import 68

6.6. Data Verification (Quality Control Part 1) 69

6.7. Data Validation (Quality Control Part 2) 70

6.8. Data Quality Review and Communication 71

6.9. Conformance Guidelines and Revisions 72

6.10. Credits 72
Chapter 7 Data Documentation 73
7.1. NPS Standards 73
7.2. Network Standards 73
7.3. Project Documentation 73
7.4. Vital Signs Protocol Documentation 75
7.5. Additional ERMN Documentation 76
7.6. Credits 76
Chapter 8 Data Analysis and Reporting 77
8.1. Data Analysis 77
8.2. Data Reporting 77
Chapter 9 Data Dissemination 79
9.1. Data Ownership 79
9.2. Data Sensitivity 81
9.3. Access Restrictions to Sensitive Data 84
9.4. Data Distribution Mechanisms 85
9.5. Feedback Mechanisms 87
9.6. Credits 87
Chapter 10 Data Maintenance, Storage and Archiving 89
10.1. Digital Archiving and Storage 89
10.2. Digital Archiving and Storage Procedures 92
10.3. ERMN Hardcopy and Digital Reports 94
10.4. Specimens 95
10.5. Photographs 95
10.6. Credits 96

Chapter 11 Literature Cited 97

Figures

Page

Figure 1.1. Map of ERMN Parks 26

Figure 2.1. Intersection of Data Management Responsibilities 37

Figure 3.1. Conceptual Model of Project Work Flow 42

Figure 3.2. Diagram of Typical Project Data Life Cycle 47

Figure 3.3. Data Flow Diagram for Water Quality Data 51

Figure 4.1. ERMN Project Folder Structure 52

Figure 4.2. ERMN Park Data Folder 53

Figure 4.3. Model of National Level Application Architecture 54

Figure 5.1. Inventory Data Flow Model 58

Figure 5.2. Data Flow Model for Monitoring Projects 59

Figure 5.3. ERMN Project Tracking Database 60

Figure 6.1. QA/QC Controls and Project Work Flow 65

Tables

Page

Table 1.1. Categories of Data Products 27

Table 2.1. Categories of Data Stewardship 32

Table 2.2. Summary of Roles and Responsibilities by Position 33

Table 3.1. Repositories for ERMN Products 48

Table 10.1. Principal ERMN Repositories for Project Products 96

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. NER NPSpecies Certifier's Guide 10

Appendix J. NER NatureBib Data Management Plan 5

Appendix K. NER NPSpecies Data Management Plan 6

Appendix L. Scanning Specifications and Personnel Commitment 7

Appendix M. Proposed NPS Image Metadata Standard 23

Appendix N. NER GPS Metadata Questionnaire Form 3

Appendix O. ERMN NPSpecies Museum Voucher Fields 2

Appendix P. 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 Work Flow

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

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

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.

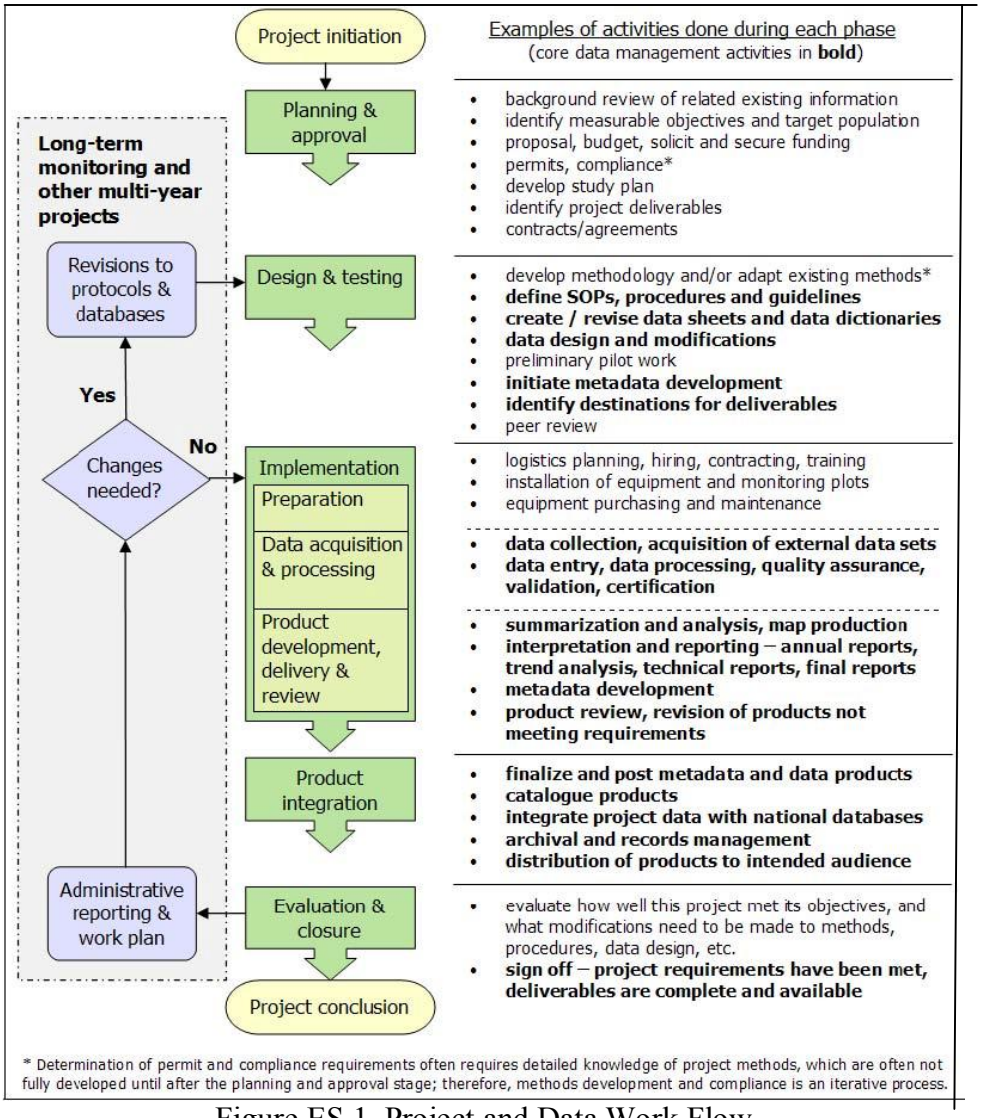


Figure ES.1. Project and Data Work Flow

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

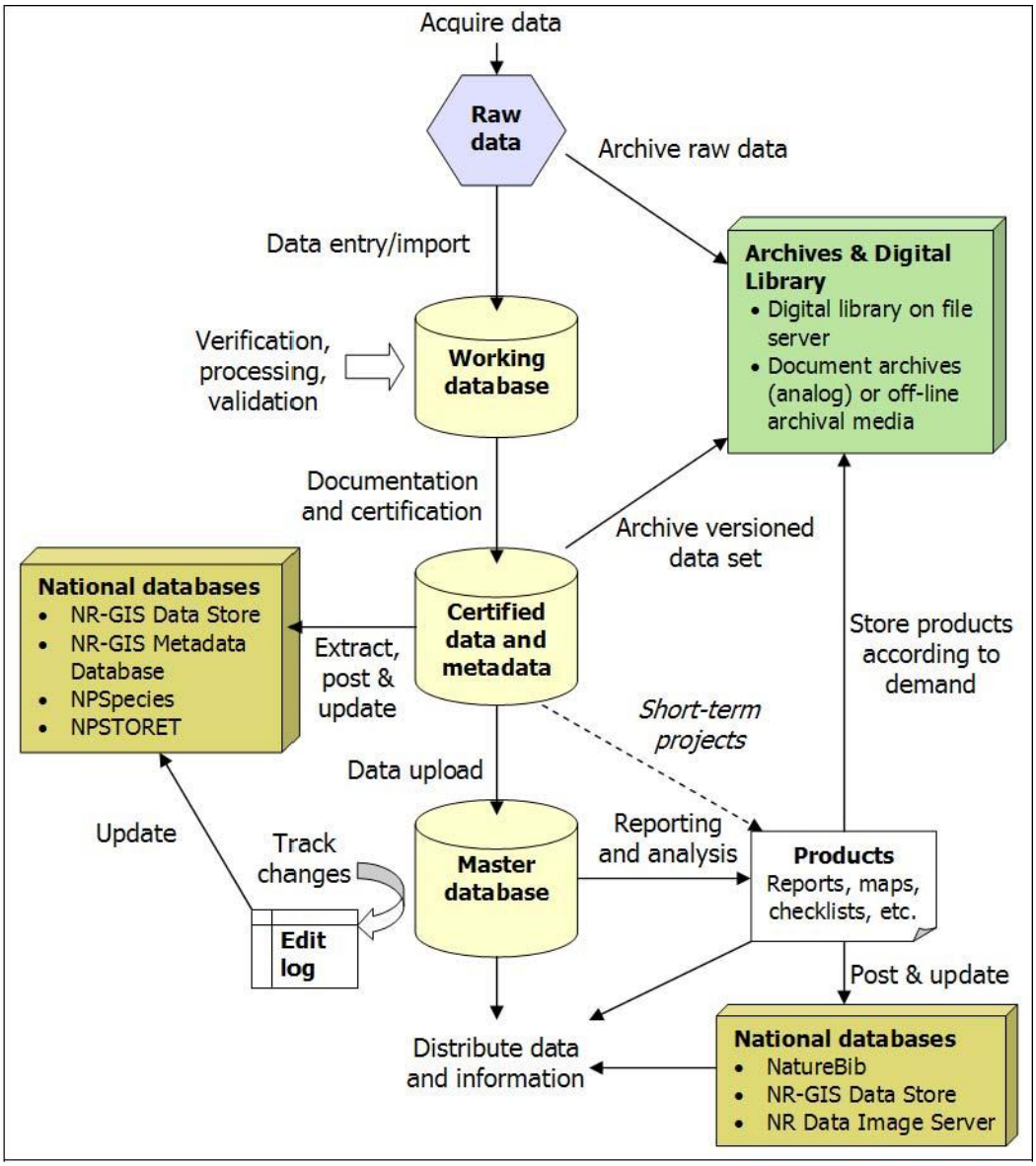


Figure ES.2. Diagram of Typical Data Life Cycle

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.

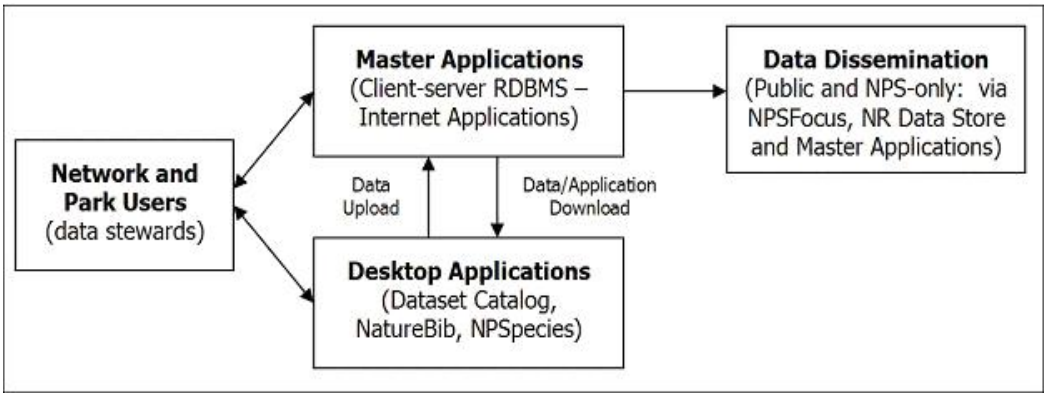


Figure ES.3. Sample Conceptual Model of I&M Program System Architecture

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.

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.

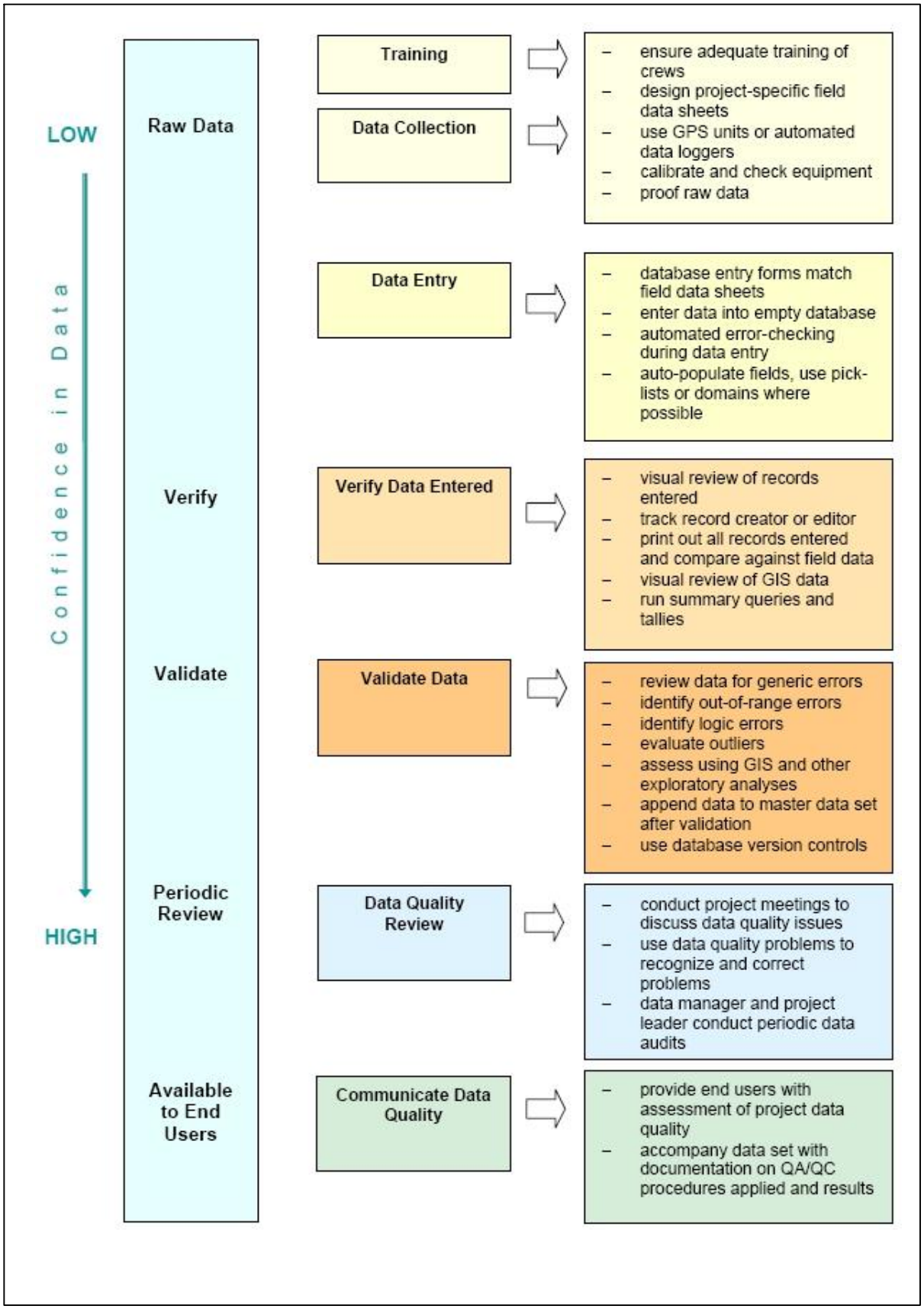


Figure ES.4. QA/QC Controls and Project Work Flow

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.

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)

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.

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.