



# Northeast Coastal and Barrier Network Inventory and Monitoring Program

## Information Management Plan

December 15, 2004  
Draft



Developed by:

*Sara M. Stevens*  
&  
*Gary Entsminger*

## **Mission Statement**

This information management plan has been developed to support NPS Inventory and Monitoring Program goals and objectives to ensure that Northeast Coastal and Barrier Network (NCBN) Inventory and Monitoring Program data and information are accessible and useful long beyond the initial field data collection phase.

## **Executive Summary**

As the basic and most important products of scientific research, data and information represent a valuable and an often irreplaceable resource (Michener and Brunt, 2000). Because field experiments and associated data collection are often time consuming and expensive, management of data and information products plays an important role in any scientific program. For long-term ecological monitoring programs, such as the NCBN Inventory and Monitoring Program, retention and documentation of high quality data are the foundation upon which the success of the overall program rests.

To develop and retain high quality data, the Network has developed the following draft information management plan that describes the Network's information management infrastructure (e.g., staffing, hardware, software) and architecture (databases, procedures, archives). This includes 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.

In addition to the Network's Information Management Plan, NCBN staff are developing specification and guidance documents to share with park, Network, regional and national staff ([http://www1.nature.nps.gov/im/units/ncbn/d\\_guidelines.htm](http://www1.nature.nps.gov/im/units/ncbn/d_guidelines.htm)). These guidelines describe methods that will be used by NCBN for managing natural resource information from hard copy reports to digital photos. Standard Operating Procedures (SOP) that describe in detail how to create FGDC compliant metadata or conduct quality control procedures on data collected by the Network are also being developed and included each of the Network monitoring protocols.

In order for the NCBN Information Management Plan to be successful, the Network must:

- provide up-to-date technical guidance for the preparation and management of data (guidance, specification and Standard operating procedure documents)
- maintain efficient standards for data processing, from acquisition to distribution

Success will be based on the resulting:

- production and maintenance of high quality information products that fulfill a wide variety of user needs

The NCBN Information Management plan outlines the approach that the Northeast Coastal and Barrier Network will take to implement and maintain a system that will serve the data and information management needs of the NCBN Inventory and Monitoring Program. The plan reflects the commitment to the establishment, maintenance, description, accessibility, and long-term availability of high-quality data and information.

### ***Roles and Responsibilities***

For the NCBN Inventory and Monitoring program to work effectively, everyone within the Network will have stewardship responsibilities in the production, analysis, management, and/or end use of data produced by the program. To meet the new data management goals and standards developed by the National Park Service and its constituents, Network staff must understand how data and information flow, and what their roles and responsibilities are in this process.

There are four main categories of data stewardship roles to be handled by Network personnel. These are:

1. production
2. analysis
3. management
4. end use

Each of these broad categories has principal, or ‘must-do’ responsibilities as well as many potential ancillary tasks. As coordinator of these tasks, the fundamental role of the Network data manager is to understand and determine program and project requirements, to create and maintain data management infrastructure and standards, and to communicate and work with all responsible individuals.

### ***Project Workflow***

To better understand the information management needs of the NCBN program, it is useful to understand the general work flow of project development and the information management tasks associated with each stage. There are two main types of projects handled by the Network:

- short-term projects, which may include individual park research projects, inventories, or pilot work done in preparation for long-term monitoring
- long-term projects, mainly the Network vital-signs 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

For information management the primary difference between short- and long-term projects is an increased need to adhere to and maintain standards for long-term projects. Maintaining standardization from year-to-year will be necessary when comparing data over an extended period of time (years for vital signs monitoring).

Within this information management plan, both short and long-term projects are divided into four primary stages:

1. initiation and approval
2. planning-design and testing
3. implementation
4. finalization-product integration and evaluation

Each of these four stages are associated with multiple information management tasks. During the *initiation and approval stage*, preliminary decisions are made regarding the scope of the project and its objectives. A scope of work may or may not be written for each project, but a proposal is developed and funding sources, permits, and compliance are addressed. A cooperative agreement or contract is developed and finalized. Although many of these responsibilities rest with the project leader and/or program administrators, data management staff must be involved in identifying project deliverables and assuring that each contract or agreement includes a clear list of these deliverables, with reference to either national, regional or Network information and data standards.

At the *planning-design and testing stage* of a project either an inventory study plan or monitoring protocol is developed that details how data will be acquired, processed, analyzed, and reported. Information management tasks associated with this stage include data design, development and maintenance of guidelines and specifications, and dissemination of this information. This stage is one of the most important as it initiates the development of sound, quality data products.

Once the design and testing and the administrative tasks associated with project information management have been completed, the project *is implemented*. At this stage the technical information management staff is critical to the success of the project. Tasks include acquiring, processing, and documenting data. During this phase, products such as reports, maps, and GIS themes are developed. All raw data undergo QA/QC measures and final manipulated products are reviewed. Although many of these tasks may be completed by Network cooperators, the information management staff must be closely involved in the training, development, and review of all draft and final project products.

Once all products have been developed and gone through extensive review, *product integration and evaluation* occurs. Records are either finalized permanently for short-term projects or finalized for the project year for long-term projects. Records are finalized or closed out for the year in the Network project tracking database to reflect status and deliverables. Information management tasks include the review, dissemination, and archiving of all products.

Although all Network projects vary in terms of the final products they produce, all follow these four basic stages. The differences between projects occur within the stage itself and are dependent

on the category and type of data being collected or compiled. This dictates how data are acquired and processed within each project stage.

### ***Data Acquisition and Processing***

There are two categories of data acquired and managed by the Network:

1. **Network-based data**—those data collected by Network staff and/or cooperators working with the Network
2. **Network-integrated data**—those data collected by other entities (parks, universities, other agencies, other NPS programs), but identified as important natural resource data, necessary for the Network to manage

*Network-based data* are those data originating within the Network or are currently being collected by NCBN staff. These include three of the twelve basic I&M Program biological inventory datasets:

- vegetation maps
- species occurrence inventories
- species distribution inventories

Along with vegetation maps and species inventories, the Network manages long-term monitoring data as part of its Vital Signs program. NCBN is currently developing protocols to monitor:

- salt marsh vegetation
- estuarine nekton
- geomorphologic change
- estuarine nutrient enrichment
- landscape change
- visitor impacts

*Network-Integrated data* can be divided into two more data type categories:

1. *Current or ongoing datasets*—These datasets are pre-determined for acquisition by the Network and follow very specific acquisition and processing steps. These can either be park-based or from external NPS sources. They are protocol dependent datasets that the Network acquires on a regular basis. These datasets are either used for data analyses and comparison purposes with Network-based monitoring data, or they are baseline datasets essential for the completion of a vital sign monitoring protocol.
2. *Legacy datasets*—are those data found and compiled through the data mining process. These include vertebrate and vascular plant species data, other important natural resource inventory data, specimen or voucher data, bibliographic data, and existing monitoring datasets.

Network-based data and Network-integrated data follow slightly different acquisition and processing steps. These are described in detail in the NCBN Information Management Plan (Chapter 5). Field data (Network-based data), data acquired from external sources and legacy data (Network-integrated data), are handled through a variety of steps to reach their final archiving stage. Although some of the steps differ from one data type to another such as the acquisition and dissemination steps, there are four main steps that all NCBN data undergo once acquired:

1. quality assurance/quality control
2. documentation
3. transcription to master databases
4. archiving

Data and information are stored, maintained and disseminated through Network and nationally based database management systems. Details are available in the Network plan (Chapter 5). NCBN vital signs data are stored in the NCBN monitoring database template. Network water quality data are housed in the national water quality database STORET. Species data are managed in the NPS NPSpecies database, and bibliographic data are compiled by the Network in the NPS NatureBib database.

### ***Quality Assurance, Quality Control***

When developing a long term ecological monitoring program, it is imperative that information and data developed as part of the program be of high quality and adequate for its intended use (US EPA, 2001). In order to develop quality products, a plan for quality assurance as well as methods for quality control must be developed at all levels of the program. Network staff and cooperators conducting ecological monitoring must be aware of both the need for and the mechanisms to achieve excellence at all levels of product development. To accomplish this, NCBN is developing a quality management system. The Network's quality management system will include the organizational structure, responsibilities, procedures, processes, and resources for implementing QA/QC for its ecological monitoring program.

NCBN will establish guidelines for the identification and reduction of error at all stages in the data lifecycle, including project planning, data collection, data entry, verification and validation, processing, and archiving. This approach requires that the Network:

- develop a plan for quality assurance that will include the identification of roles and responsibilities of Network, park, and cooperative staff for maintaining quality standards at all levels of the program—from field and laboratory data collection to overall data management procedures
- ensure that the process of achieving quality is not only documented, but maintained through routine review by Network staff
- develop protocols and SOPs to ensure data quality
- evaluate the quality of all data and information based on NPS standards before data are distributed

- perform periodic data audits and quality control checks to monitor and improve the Network's data quality program.

Much QA/QC work involves defining and enforcing standards for electronic formats, locally defined codes, measurement units, and metadata. This process 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. Documentation of the dataset's quality review process will be added to the project metadata.

### ***Data Documentation***

Another critical step following quality assurance and control is data documentation. Data documentation is another step towards ensuring that datasets are useable for their intended purposes well into the future. Data documentation involves the creation of metadata. Metadata describe how, when, and by whom a particular set of data was collected, and how the data are formatted. Metadata include information about the quality, condition, and characteristics of a dataset. Metadata help us create and maintain a framework for cataloging datasets, which in turn helps make datasets more readily available to a broad range of users.

A significant amount of guidance has become available on proper data documentation (see NCBN Information Plan Appendices). As mandated by the National Park Service, all NCBN metadata associated with geospatial data will conform to Federal Geographic Data Committee (FGDC) standards. There are a variety of software tools available for creating and maintaining FGDC compliant metadata.

For biological datasets, NCBN has adopted the Biological Data Profile Metadata standards developed by the National Biological Information Infrastructure (NBII). All Network-based datasets will be accompanied by the Biological Data Profile when distributed. Northeast Region cooperators have developed helpful guidelines on tools used for creating Biological Data Profile metadata (see Plan Appendices).

NCBN data management staff will provide training and support to project leaders to facilitate metadata development. Upon completion, metadata will be posted and made available and searchable in conjunction with related data and reports via the NCBN website, as well as the national NR-GIS Data Store.

### ***Data Distribution***

Access to NCBN data products will be facilitated via a variety of information systems that allow users to browse, search, and acquire Network data and supporting documents. These systems include the NCBN website, and national applications with internet interfaces (NatureBib, NPSpecies, NR-GIS Data Store, etc.). The following table provides a list of repositories and types of data that will or can be housed there (Table A).

**Table A. Information management systems that facilitate dissemination of NCBN information**

<b>Web Application</b>	<b>Data types available at site</b>	<b>Web Address</b>
NPSpecies	Data on park biodiversity (species information)	<a href="http://science.nature.nps.gov/im/apps/npspp/index.htm">http://science.nature.nps.gov/im/apps/npspp/index.htm</a>
NatureBib	Scientific citations related to park resources	<a href="http://www.nature.nps.gov/nrbib/index.htm">http://www.nature.nps.gov/nrbib/index.htm</a>
NR-GIS Metadata and Data Store	Metadata, spatial and non-spatial data products	<a href="http://science.nature.nps.gov/nrdata">http://science.nature.nps.gov/nrdata</a>
Biodiversity Data Store	The raw or manipulated data and products associated with I&M data that have been entered into NPSpecies.	<a href="http://science.nature.nps.gov/im/inventory/biology/index.htm">http://science.nature.nps.gov/im/inventory/biology/index.htm</a>
NCBN Website	Reports and metadata for all Network projects	<a href="http://www1.nature.nps.gov/im/units/ncbn/index.htm">http://www1.nature.nps.gov/im/units/ncbn/index.htm</a>

Because Network data will reside in the repositories listed above, these data will automatically be searchable via the integrated metadata and image management system and search gateway called “NPS Focus.” This system is being built with Blue Angel Enterprise software for metadata management and the LizardTech Express Server for image management. Currently, ten NPS and two non-NPS databases have been integrated into the NPS Focus prototype in either full or test bed form for one stop searching. NPS Focus has been released as an Intranet version only (<http://focus.nps.gov/>) – release of a public version is projected in the near future.

### **Archiving**

The final information management step, and one of the most vital to the Network’s Inventory and Monitoring Program, is the long-term maintenance and management of digital and analog information. 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 (Vogt-O'Connor 2000).

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, NCBN archiving procedures include saving datasets in both their native format (typically MS-Access 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.

Chapter 10 of the NCBN Information Management Plan 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.

# Table of Contents

## ***Chapter 1 Purpose and Scope of the Northeast Coastal and Barrier Network (NCBN) Information Management Plan***

- 1.1 Data Management Overview
- 1.2 The NCBN Inventory and Monitoring Program's Long-Term Goals
- 1.3 Overview of the NCBN Information Plan
- 1.4 Sources of Natural Resource Data
- 1.5 NCBN Information Management Plan Principles and Scope
- 1.6 Revisions
- 1.7 Intended Audience and Layout for the NCBN Data Management Plan

## ***Chapter 2 Information Management Staff Coordination of Roles and Responsibilities***

- 2.1 Information Management Staff
  - 2.1.1 National Level Staff and Coordination
  - 2.1.2 Northeast Region Data Management Staff
  - 2.1.3 NCBN Information Management Staff
  - 2.1.4 Park Level Data Management Staff
- 2.2 Coordination of Roles and Responsibilities
  - 2.2.1 NCBN Data Management Coordination
  - 2.2.2 Data Management Staff-Roles and Responsibilities
  - 2.2.3 Project Leaders
  - 2.2.4 Project Leaders-Roles and Responsibilities
  - 2.2.5 Sharing Responsibilities
  - 2.2.6 Communication
- 2.3 Acknowledgements

## ***Chapter 3 Information Management Resources: Infrastructure and Systems Architecture***

- 3.1 Computer Resources Infrastructure
  - 3.1.1 NCBN Digital File Maintenance and Storage
- 3.2 NCBN Systems Architecture
  - 3.2.1 Project Database Standards
  - 3.2.2 Core tables and fields based on network and national standards
  - 3.2.3 Protocol-specific Field Data Tables
  - 3.2.4 Common Lookup Tables
  - 3.2.5 NCBN NRDT Template User Interface
- 3.3 Project Tracking Application
- 3.4 National Information Management Systems Architecture
- 3.5 Acknowledgements

## ***Chapter 4 Data Management Integration and Project Work Flow***

- 4.1 Project Work Flow
- 4.2 Project Initiation and Approval
- 4.3 Planning: Design and Testing
- 4.4 Implementation
- 4.5 Project Finalization
- 4.6 Acknowledgements

## ***Chapter 5 Data Acquisition and Processing***

- 5.1 Categories of Network Data
  - 5.1.1 Network-based Data
- 5.2 Acquiring and Processing Field Data (Biological Inventories and Monitoring)
  - 5.2.1 Data Acquisition and Processing, Steps
- 5.3 Network Water Quality Data Integration
- 5.4 Network-Integrated Data
  - 5.4.1 Acquiring and Processing Current or Ongoing Datasets
  - 5.4.2 Acquiring and Processing Legacy Datasets
    - 5.4.2.1 Species Data
    - 5.4.2.2 Bibliographic Data
    - 5.4.2.3 Specimen Collection Data
    - 5.4.2.4 Long-term Monitoring Datasets

## ***Chapter 6 Quality Assurance, Quality Control***

- 6.1 National Park Service Mandate for Quality
- 6.2 Quality Assurance (QA) and Quality Control (QC)
  - 6.2.1 Data Verification and Validation
- 6.3 NCBN QA/QC Plan
- 6.4 NCBN QA/QC Plan for Data Collection
  - 6.4.1 QA Measures for Data Collection
  - 6.4.2 NCBN Data Collection Quality Control Measures
- 6.5 NCBN QA/QC Plan for Data Entry
  - 6.5.1 QA Measures for Data Entry
  - 6.5.2 QC Techniques for Data Entry
  - 6.5.3 NCBN Data Verification Methods
  - 6.5.4 NCBN Data Validation Methods
- 6.6 Monitoring Conformance to Plans and Standards
- 6.7 Data Quality Review and Communication
  - 6.7.1 Feedback from QA/QC Measures to Improve Data Quality
  - 6.7.2 NCBN Quality Review of Completed I&M Products
- 6.8 Roles and Responsibilities
- 6.9 Creating a QA/QC Audit Trail

6.10 Acknowledgements

***Chapter 7 Data Documentation***

- 7.1 Metadata
- 7.2 Sources for Data Documentation Guidance
- 7.3 NPS Integrated Metadata System Plan and Tools
- 7.4 Network Data Documentation Strategy

***Chapter 8 Summaries, Analyses, and Reports***

- 8.1 Data Users
- 8.2 Data Delivery
- 8.3 Preparation of Data for Exchange and Distribution
- 8.4 Monitoring Data Summaries and Analyses
- 8.5 NCBN Monitoring Project Reports
- 8.6 Species Inventory Summaries, Analyses and Reports
- 8.7 Other Network-generated Reports

***Chapter 9 Data Ownership***

- 9.1.1 National Park Service Policy on Data Ownership
- 9.1.2 Establishing Data Ownership Guidelines
- 9.2 Data Distribution
  - 9.2.1 Data Distribution Mechanisms
  - 9.2.2 Data Classification: Protected vs. Public
  - 9.2.3 Access Restrictions on Sensitive Data
  - 9.2.4 Public Access to Network Inventory and Monitoring Data
  - 9.2.5 Data Availability
  - 9.2.6 Data Acquisition Policy
- 9.3 Data Feedback Mechanisms
  - 9.3.1 Data Error Feedback Procedures
- 9.4 Acknowledgements

***Chapter 10 Data Maintenance, Storage, and Archiving***

- 10.1 Digital Data Maintenance
  - 10.1.1 Short-term Datasets
  - 10.1.2 Long-term Monitoring of Datasets
  - 10.1.3 Spatial Data
- 10.2 Storage and Archiving Procedures—Digital Data
  - 10.2.1 Storage Facilities for Electronic Archives
  - 10.2.2 Directory Structure for Electronic Archives
  - 10.2.3 Backup Procedures for Digital Data
  - 10.2.4 Data and Network Security

- 10.3 NCBN Storage and Archiving of Official and Non-official Records
  - 10.3.1 Hard Copy and Digital Reports Produced by the Network
- 10.4 Specimens
- 10.5 Photographs
  - 10.5.1 Aerial Photographs
- 10.6 Role of Curators in Storage and Archiving Procedures
- 10.7 Acknowledgements

## **NCBN Data Management Guidance Documents and Appendices Associated with the NCBN Information Plan**

### *Appendix: Document Title:*

1. Northeast Region GPS Field Collection Standards for Cooperators (*to be developed*)
2. [Northeast Coastal and Barrier Network Standard Operating Procedure \(SOP\) for the Garmin V GPS Unit](#)
3. Northeast Coastal and Barrier Network File Management Plan (*to be developed*)
4. Northeast Region I&M Program Guidelines for developing FGDC compliant metadata for both spatial and biological data (*to be developed*)
5. [Northeast Region Metadata tools used in the creation of the FGDC biological data profile](#)
6. NCBN I&M Program Standards for File Revision Control (*to be developed*)
7. NCBN I&M Program Data Management Standards for Voucher Collection and Cataloging (*to be developed*)
8. Northeast Region I&M Program Database Specifications (*to be developed*)
9. [Northeast Region I&M Program NPSpecies Certifier's Guide](#)
10. [Northeast Region Inventory and Monitoring Program NPSpecies Data Management Plan](#)
11. [Northeast Region Inventory and Monitoring Program NPSpecies Data Entry and Data Management Standards](#)
12. [Northeast Region I&M Program NatureBib Data Management Plan](#)
13. [Northeast Region I&M Program NatureBib Data Management and Data Entry Standards Manual](#)
14. [Northeast Region I&M Program Product Specifications](#)
15. [NCBN Standard Operating Procedure Converting the NCBN Database Template](#)
16. [NCBN Saltmarsh Monitoring Database User's Guide](#)
17. Northeast Region I&M Program Data Mining Specifications (*to be developed*)
18. NCBN I&M Program Project Management Guidelines (based on use of Project KickStart Software) (*to be developed*)
19. NCBN Data Distribution Plan (*to be developed*)

20. Northeast Region I&M Program Web-Page Development and Management Guidelines (*to be developed*)
21. [NCBN GIS Specifications and Guidelines](#)
22. [Cape Cod National Seashore Digital Image Management Guidelines \(Draft\)](#)
23. NCBN Quality Assurance Plan (*to be developed*)
24. [Northeast Coastal and Barrier Network FOIA and Sensitive Data Information](#)
25. [NCBN Data Stewardship-Staff Roles and Responsibilities](#)
26. [Northeast Region I&M Program QA/QC Procedures and Metadata Development for Final Biological Inventory Products](#)
27. [NCBN Data Dictionary Example](#)
28. [NCBN Example Language Describing Deliverables for use in Cooperative Agreements and Contracts](#)
29. Northeast Region Natural and Social Science Study Proposal and Deliverable Guidelines (*download this document at: [www.nps.gov/nero/science/](http://www.nps.gov/nero/science/)*)
30. [Northeast Region GPS Metadata Form](#)
31. [Northeast Region Vegetation Mapping Aerial Photo Specification Example for scale 1:6000](#)
32. [Northeast Region Vegetation Mapping Aerial Photo Specification Example for scale 1:12000](#)
33. NCBN Lidar Data Acquisition and Processing Example (*to be developed*)
34. [Northeast Region Inventory and Monitoring Program Vegetation Map Development Example](#)
35. [Northeast Region Example of Spatial Data Review for Vegetation Mapping Products](#)
36. [Northeast Region Inventory and Monitoring Program Example Vegetation Mapping Deliverables and Mapping Report Guidelines](#)
37. NCBN Metadata Questionnaire (*to be developed*)

## **Chapter 1**

### **Purpose and Scope of the Northeast Coastal and Barrier Network (NCBN) Information Management Plan**

#### ***1.1 Data Management Overview***

The Inventory and Monitoring (I&M) Program 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 the framework for the I&M Program to fully integrate natural resource monitoring and other scientific activities into the management processes of the National Park System.

The Act 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 National Park Service initiated a service-wide, natural resource Inventory and Monitoring Program encompassing 270 parks with significant natural resources. Ecologically similar parks among the 270 chosen were grouped into 32 Networks. Each I&M Network has been tasked with documenting existing park vertebrates and vascular plants, developing a management-based, ecological monitoring program with a written plan and protocols, as well as an Information Management Plan that encompasses all aspects of the Network program.

With the primary goal to “improve park management through greater reliance on scientific knowledge” the Northeast Coastal and Barrier Network (NCBN) Inventory and Monitoring Program will identify, catalog, organize, structure, archive, and provide high-quality natural resource data, which institutions and individuals alike can share. Researchers, managers, and administrators will rely on the Network to secure and maintain the quality of these data over a long period of time.

As the basic and most important products of scientific research, data and information represent a valuable and often irreplaceable resource (Michener and Brunt, 2000). Because field experiments and associated data collection are often time and budget consuming, management of data and information products plays an important role in any scientific program. For long-term ecological monitoring programs, such as the NCBN Inventory and Monitoring Program, retention and documentation of high quality data are the foundation upon which the success of the overall I&M Program rests.

## ***1.2 The NCBN Inventory and Monitoring Program's Long-Term Goals***

The I&M Program's success at identifying, cataloging, organizing, structuring, archiving, and providing relevant natural resource information will largely determine the Program's efficacy and image among critics, peers, and advocates.

The I&M Program's long-term goals are to:

- establish natural resource inventory and monitoring standards throughout the National Park system that transcend traditional program, activity, and funding boundaries
- inventory the natural resources and park ecosystems under National Park Service stewardship
- monitor park ecosystems to provide reference points for comparisons with other, altered environments
- integrate natural resource inventory and monitoring information into National Park Service planning, management, and decision making
- share National Park Service accomplishments and information with other natural resource organizations and form partnerships for attaining common goals and objectives

To achieve the last two of these goals, a modern information management infrastructure must be developed (e.g., staffing, hardware, software). This infrastructure will include procedures to ensure that relevant natural resource data collected by NPS staff, cooperators, researchers, and others will be entered, quality-checked, analyzed, reported, archived, documented, cataloged, and made available to others for management decision-making, research, and education.

The National Park Service is a highly decentralized agency with complex data requirements. The primary audience for many of the products from the I&M Program is at the Park level—to provide park managers with the information they need to make better-informed decisions and to work more effectively with other agencies and individuals for the benefit of Park resources. However, certain data are also needed at the regional or national level for a variety of purposes. The National Park Advisory Board (2001) has stated that the findings “must be communicated to the public, for it is the broader public that will decide the fate of these resources.” Therefore, the NCBN Information Management Program cannot simply attend to database tables, and fields, it must also provide a process for developing, preserving, and integrating the context that makes data interpretable and valuable.

**1.3 Overview of the NCBN Information Plan**

All data—whether collected last week or 20 years ago—must be accompanied by enough information to reliably preserve and use it in the future. Future users of this data will need to know as much as possible about how and why these data were collected. Thus, datasets will be accompanied by sufficient documentation and products to answer questions about the data.

Throughout the NCBN Information Management Plan the term “data” is used to encompass all products generated alongside primary tabular and spatial data as part of the I&M Program. These data fall into five general categories: raw data, derived data, documentation, reports, and administrative records (Table 1.1).

**Table 1.1 Categories of data products and project deliverables**

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

To meet I&M Program goals, and to ensure adequate context for primary data products, all of these categories require some level of management to ensure their quality and availability. Thus, the Network will use a more “holistic view” about how natural resource data are generated, processed, finalized, and provided. All phases of data processing will be integrated, and information about each phase and its processes shared through high-quality documentation.

**1.4 Sources of Natural Resource Data**

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

- inventories
- monitoring
- pilot studies associated with protocol development
- special focus studies done by internal staff, contractors or cooperators
- 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
- resource management and restoration work

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

This plan covers four major categories of data sources and types that are coordinated or managed by the Network:

1. Data managed in service-wide databases. NCBN uses three data systems developed by the I&M WASO office. NatureBib is used as a bibliographic tool for cataloging reports, publications, or other documents that relate to natural resources in park units. Dataset Catalog is used to document primarily non-spatial databases or other data assemblages. NPSpecies is used by the Network to develop and maintain lists of vertebrates and vascular plants in Network parks, along with associated supporting evidence.
2. Data developed or acquired directly by the Network as a result of inventory, monitoring, or other projects. This category includes project-related protocols, protocols, vegetation maps, reports, spatial data, and associated materials such as field notes and photographs provided to NCBN by contractors or developed by NCBN staff. Projects can be short-term (one to two years duration) or long-term (ongoing monitoring).
3. Data that, while not developed or maintained by NCBN, are used as data sources or provide context to other datasets. Examples of this category include: GIS data developed by Parks, other agencies or organizations; national or international taxonomic or other classification systems; climate or hydrologic data collected by regional or national entities.
4. Data acquired and maintained by Network Parks that NCBN assists in managing. Because of the lack of data management expertise in many Network parks, NCBN provides data management assistance for high-priority datasets or those that may benefit from standardized procedures. Examples include: a multi-park database for rare plant data; datasets of legacy natural resource monitoring data; and data on exotic invasive plant species.

These above categories can contain one or more of the following data formats:

- hard-copy documents (e.g., reports, field notes, survey forms, maps, references, administrative documents)
- objects (e.g., specimens, samples, photographs, slides)
- electronic files (e.g., Word files, email, websites, digital images)

- electronic tabular data (e.g., databases, spreadsheets, tables, delimited files)
- spatial data (e.g., shapefiles, coverages, remote-sensing data)

Each of these data formats has specific requirements for ongoing management and maintenance, which are address in this plan.

### ***1.5 NCBN Information Management Plan Principles and Scope***

The NCBN approach to data management is to develop an Information Management Plan that is “user friendly” to a varied audience, from park natural resource managers to data managers. This means providing guidance on data management practices at a number of different levels. To facilitate this guidance, the NCBN plan has been developed along three basic principles:

1. keep the plan simple, flexible, and evolving
2. make it useful to all—to park GIS and data management staff, regional technical staff, resource management staff, and cooperating scientists
3. include the data users in the decision making process whenever possible

The NCBN Information Management Plan outlines how the Network intends to implement and maintain a system that will serve the data and information management needs of its Inventory and Monitoring Program. This plan reflects commitment to the establishment, maintenance, description, accessibility, and long-term availability of high-quality data and information.

The NCBN Data Management Plan describes how the Network will:

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

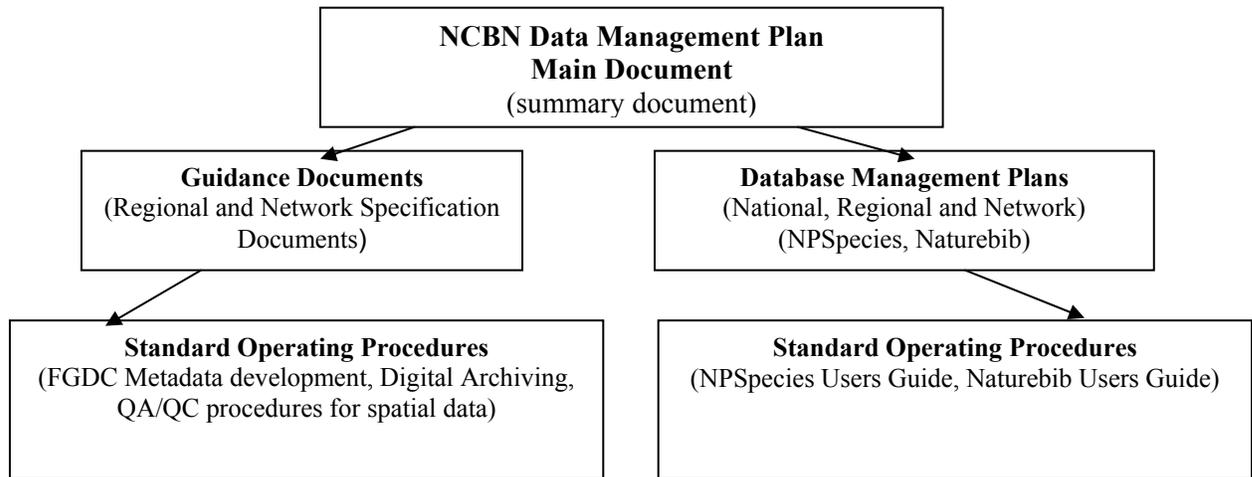
### ***1.6 Revisions***

The NCBN data management plan covers I&M Program needs based on the most current information systems technology relevant through 2004. Revisions to this plan and associated data management documents (guidelines and SOPs) will be made as needed.

### ***1.7 Intended Audience and Layout for the NCBN Data Management Plan***

The conceptual model in Figure 1.1 shows the layout of the NCBN Information Management Plan. There will be three tiers of document types with multiple modules within tiers.

**Figure 1.1 Layout of the NCBN Information Management Plan**



Four document types are being developed as part of this plan:

1. The NCBN Information Management Plan (Summary Document)
2. Subject Specific Guidance Documents (GIS specs) (Some drafts available)
3. Specific Database Management Plans (NPSpecies, NatureBib)
4. Standard Operating Procedures (SOP) (Some drafts available)

NCBN developed this information management plan in modular format with the hope of increasing both its readability and usability among a wider audience. This format allows information plan users to easily locate and access subdocuments pertaining to a particular element of information management guidance or process. Individual procedural documents from 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 complete and technical review. Technical experts can be provided with applicable portions of the plan based on their expertise (GIS, Database, Metadata, GPS, etc.).

Standard Operating Procedures (SOP) developed as part of this plan detail work processes that are to be conducted or followed by Network staff and cooperators. SOPs are useful in that they document the way activities are to be performed to facilitate consistent conformance to technical and quality system requirements and to support data quality. The development and use of SOPs promotes quality through consistent implementation of a process or procedure within the organization, even if there are temporary or permanent personnel changes.

## Chapter 2

### Information Management Staff Coordination of Roles and Responsibilities

An increasing demand for more detailed, high quality data and information about natural resources and ecosystem functions requires a group of people working together to steward data and information assets. Knowledgeable individuals from many disciplines must come together to ensure that data are collected using appropriate methods, and that resulting datasets, reports, maps, models, and other derived products are well managed. Datasets and the presentations of these data must be credible, representative, and available for current and future needs.

For the NCBN Inventory and Monitoring program to work effectively, everyone within the Network must have stewardship responsibilities in the production, analysis, management, and/or end use of data produced by the program. In order to meet the new data management goals and standards developed by the National Park Service and its constituents, Network staff must understand how data and information flow and what their roles and responsibilities are in this process.

Table 2.1 summarizes the data stewardship roles of various Network personnel. Each of these broad categories has primary, or ‘must-do’, responsibilities as well as many potential ancillary tasks. As coordinator of these tasks, the fundamental role of the *Network data manager* will be to understand and determine program and project requirements, to create and maintain data management infrastructure and standards, and to communicate and work with all responsible individuals. This new and crucial emphasis on data management, analysis, and the reporting of results will require a large investment of personnel, time, and money. The NCBN Inventory and Monitoring Program expects to invest at least thirty percent of available resources in developing a high-quality data management system.

**Table 2.1 Categories of data stewardship involving all Network personnel**

Stewardship Category	Related Activities	Principal jobs or positions
<p><u>Note:</u> Each position is listed in only one category according to overriding responsibilities. However, most positions contribute in each category.</p>		
Production	<p>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.</p>	<p>Project Crew Member Project Crew Leader Data/GIS Specialist or Technician</p>

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.	Network Ecologist Resource Specialist
Management	Preparing and executing policies, procedures, and activities that keep data and information resources organized, available, useful, compliant, and safe.	Network Data Manager Project Leader GIS Manager Information Technology Specialist Database Manager National Level I&M Data Manager
End Use	Obtaining and applying available information to develop knowledge that contributes to understanding and managing park resources.	Network Coordinator Park managers and superintendents Others

This chapter describes the Network’s current Information management staff, as well as regional and Network park staff. As the Network’s Long-term Monitoring Program becomes fully implemented, the staffing configuration will change. These changes will include a need to better define the data management positions within the Network in relation to the implementation of the monitoring protocols. At this time, a number of Network long-term staffing plan scenarios are being reviewed by the Network’s Board of Directors and Technical Steering Committee. These scenarios include information management staff similar to the current configuration described below, with data management assistance contracted through one of the Regional Field Technical Support Centers (FTSC).

**2.1 Information Management Staff**

Management of information will be coordinated at national and regional levels.

**2.1.1 National Level Staff and Coordination**

At the national level, 32 Network data managers have been hired as part of the Inventory and Monitoring Program. I&M Program data managers regularly coordinate with each other and national program staff via annual meetings, conference calls, workgroups, a listserv, web sites, and informal communication. Data managers from these Networks share the workload by working closely with each other while developing their respective Network Data Management Plans. This model of cooperation and communication has been very effective, and should be applied to resource information management issues and implemented by administrators.

The Northeast Coastal and Barrier Network maintains an active role in promoting practical consistency among protocols and datasets involving other Networks and organizations. GIS and data managers from all Network parks are encouraged to participate in program development and activities. Park data managers are encouraged to use Inventory and Monitoring Program resources, and to communicate with Network and program staff to share knowledge, expertise and information.

### ***2.1.2 Northeast Region Data Management Staff***

The Northeast Region Inventory and Monitoring Program falls under the Natural Resource and Stewardship division of the Northeast Region Support Office (NESO), and directly under one of two Northeast Region Chief Scientists. Natural Resource and Stewardship staff at the regional level will assist regional Inventory and Monitoring staff in:

- identifying and prioritizing regional and individual park needs
- identifying qualified researchers and potential cooperators
- administering research contracts and cooperative agreements
- providing scientific consultation as Network board of directors and science committee members

At the regional office level no over-arching natural resource data management positions existed prior to the development of the four regional I&M Networks. By 2005, the region will be staffed with four I&M Network data managers. Although some coordination has begun among the region's information managers, it is expected that more coordination will occur with the inception of the NCBN data management plan, which will be the first I&M data management plan available in the region. With the use of this plan as a starting point, regional IT staff, GIS Coordinators, curators and Network data managers can better coordinate on future data management initiatives and ideas within the region to improve on current data management practices.

NCBN data management staff have worked closely with park GMP planners in providing initial vertebrate inventory data collected as part of the Network's I&M Program. Product specifications developed by the Network have been adopted by regional science staff for use with cooperative agreements and contracts funded outside I&M. The NCBN data manager has acted as a reviewer of contract deliverables at the regional level, assuring that data management standards are clearly and completely written. Close coordination has begun between I&M data management staff and regional science staff to develop region-wide standard operating procedures addressing various data management practices.

### ***2.1.3 NCBN Information Management Staff***

*NCBN information management is coordinated by one permanent NPS data manager. The Network Data Manager supervises and coordinates one term NPS database manager/programmer and two part-time research associates at the University of Rhode Island through cooperative agreement.*

Currently, most of the Network's data management is conducted through cooperative agreements with two Regional NPS Field Technical Support Centers (FTSC):

- Environmental Data Center at the University of Rhode Island
- Center for Earth Observation at North Carolina State University

These FTSCs provide technical assistance to all of the parks in the Northeast as well as the four I&M Networks in the region.

*North Carolina State University FTSC (NCSU FTSC)*

The Center for Earth Observation at North Carolina State University has worked with the Northeast Region of the National Park Service (NPS) for over 10 years in the development of geographic information systems (GIS) for resource management. This activity has led to major advances in the planning and application of GIS technology in the NPS and has placed the Northeast Region among the leaders within the NPS in this regard.

In 1996, the Philadelphia Support Office of the Northeast Region designated NCSU as a Field Technical Support Center for GIS support. In this capacity NC State University has been involved in a variety of activities including: developing and reviewing park GIS plans; developing hardware, software, and data acquisition strategies; providing telephone and on-site technical support; programming for GIS enhancement projects; teaching short courses in GIS operations; performing executive reviews; delivering national presentations; providing in-park assistance with GIS implementation and applications; and assisting with region-wide GIS strategic planning.

NCSU's experience in GIS spans 20 years and includes a long association with public resource management agencies. The Center for Earth Observation is part of the College of Natural Resources and our GIS research program has focused on the development of appropriate sized GIS programs for non-centralized resource management (e.g. smaller NPS units, US Forest Service ranger districts, state park regions, etc.) The NCSU facilities include a full complement of GIS and related software packages, a separate remote sensing research program, a large array of computer hardware configurations within two laboratories, and over 10 professional staff.

*University of Rhode Island FTSC (URI FTSC)*

Similar to the NCSU FTSC, the URI FTSC provides the NPS with access to a broad spectrum of expertise that is brought to bear on critical cluster- and service-wide issues. The URI FTSC was developed to support all aspects of GIS research and practices as they relate to park management, including natural and cultural research and resource management, planning and development, interpretation, visitor services and facilities management. The URI FTSC serves as the focus site for GIS support for the National Parks in the New England Cluster. The NCSU FTSC supports those parks in the Chesapeake-Alleghany Cluster. NCBN parks are divided between the two clusters and therefore the Network relies on both centers for park GIS support.

Because the NCBN Network Coordinator and Data Manager are stationed at the University of Rhode Island the Network Data Manager assists in the supervision of research associates hired through the URI FTSC. At this time one research associate works as the Network's NPSpecies Data Manager and the other assists the Network in database development and GIS. Network working data files are also housed on the FTSC server and backed-up on a regular basis as part of the center's program. Technical and hardware support are provided to Network staff as needed.

### ***2.1.4 Park Level Data Management Staff***

Within the Network, two park data management positions exist, one at Cape Cod National Seashore (CACO) a prototype park and one at Assateague Island National Seashore (ASIS). CACO, ASIS, Gateway National Recreation Area (GATE) and Colonial National Historical Park (COLO) all have permanent park GIS Coordinators as well.

#### *CACO Prototype Park*

In 1996, CACO was identified as a prototype park for long-term ecological monitoring within the Atlantic and Gulf Coast biogeographic region. As a prototype park and in partnership with the U.S. Geological Survey (USGS), CACO was charged with developing and refining long-term monitoring protocols that could be of utility to other Atlantic and Gulf Coast parks, in addition to supporting management of Cape Cod's natural resources. With the advent of the Network approach to inventory and monitoring, CACO's prototype mission expanded to include focused technical support to the Northeast and Coastal Barrier Network.

Along with the development of long-term monitoring protocols, data management initiatives related to shared protocols are being coordinated between Network and CACO's data management staff. Databases for all shared protocols are being developed jointly by both Network and park staff. NCBN and CACO staff share ideas and procedures for all aspects of data management. The CACO prototype park data manager has the same basic duties and responsibilities as the Network data manager, but also has the responsibility for mentoring and training others and developing and testing new approaches to data analysis, synthesis, and reporting of monitoring results.

The NCBN and CACO data management staff plan to update the Network data management plan as well as develop and update a plan for the prototype program at the park. Some data management SOPs and guidelines will be developed jointly and shared by both programs.

### ***2.2 Coordination of Roles and Responsibilities***

*A Role* is a function or position (e.g., *Data Manager*).

*A Responsibility* is a duty or obligation (e.g., *to review data records*).

The Natural Resource Challenge states that collaboration among the National Park Service, other public agencies, universities, and non-governmental organizations is necessary to effectively acquire, apply, and promulgate the scientific knowledge gained in National Parks. The Inventory and Monitoring Program encourages coordination among participants at all levels to help ensure that data collected by NPS staff, cooperators, researchers, and others are entered, quality-checked, documented, analyzed, reported, archived, cataloged, and made available for management decision-making, research, and education.

The following sections provide a general conceptual framework for all aspects of information management at the Network level. Related to the framework are the hypothetical roles and

responsibilities associated with each task. At this stage of development this framework is hypothetical because the Network has not completed its staffing plan. A number of scenarios are currently under review. Once a permanent staffing plan is in place, the framework presented below may change. However, this framework is useful in not only laying out potential staffing responsibilities, but also providing an example of the necessary coordination needed to maintain a quality data management program. Table 2.2 summarizes potential roles and responsibilities of Network and park personnel (See Appendix 25–*Data Stewardship Responsibilities*).

**Table 2.2 Summary of Roles and Responsibilities**

Role	Primary responsibilities related to information management
Project Crew Member	Record and verify measurements and observations based on project objectives and protocols. Document methods, procedures and anomalies.
Project Crew Leader	Supervise crew members to ensure their data collection and management obligations are met, including data verification and documentation.
Data/GIS Specialist or Technician	Perform assigned level of technical data management and/or GIS activities, including data entry, data conversion, and documentation. Work on overall data quality and stewardship with project leaders, resource specialists, and the Network data manager.
Information Technology/Systems Specialist	Provide and maintain an information systems and technology foundation to support data management.
Project Leader	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.  <b>Note:</b> The Project Leader is often a resource specialist, in which case the associated responsibilities for data authority apply (see next role). A Project Leader without the required background to act as an authority for the data will consult with and involve the appropriate resource specialists.
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.  <b>Note:</b> The Resource Specialist serving as a Project Leader is also responsible for the duties listed with that role.
GIS Manager	Support park management objectives. Coordinate and integrate local GIS and resource information management with Network, Regional, and National standards and guidelines.

<b>Role</b>	<b>Primary responsibilities related to information management</b>
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.
Database Manager	Apply particular knowledge and abilities related to database software and associated application(s)
Network Ecologist	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
Network Coordinator	Ensure programmatic data and information management requirements are met as part of overall Network business
I&M Data Manager (National Level)	Provide service-wide database design, support, and services, including receiving and processing to convert, store, and archive data in service-wide 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 the appropriate use and application of data and derived products and for providing feedback for improvements.

### ***2.2.1 NCBN Data Management Coordination***

The Network data manager works with Inventory and Monitoring Program data management staff and regional resource information management personnel to maintain a high-level of involvement in service-wide and regional databases and data management policy. The Network data manager works locally with Network personnel, Park staff, and cooperators to promote and develop workable standards and procedures that result in the integration and availability of datasets.

Key contacts for the Network Data Manager include park GIS and data managers and the project leaders for each monitoring or inventory project. Consistent and productive communication among these personnel leads to common understanding and better synchronization of Network and park data management activities. Park and Network staff work together to manage resource information using a variety of methods. These include personal visits, phone calls, email, joint meetings and training sessions, as well as the meetings and work of the Network’s Technical Steering Committee and Board of Directors. Involvement and input from Park scientists and resource information management staff is essential.

### ***2.2.2 Data Management Staff-Roles and Responsibilities***

The data management staff is responsible for ensuring the compatibility of project data with program standards, for designing the infrastructure for the project data, and for ensuring long-term data integrity, security, and availability.

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 digital data
- establish and implement procedures to protect sensitive data according to project needs
- collaborate with GIS specialists to integrate tabular data with geospatial data in a GIS system in a manner that meets project objectives

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

- define the scope of the project data and create a data structure that meets project needs
- become familiar with how the data are collected, handled, and used
- review quality control 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
- create a user interface that streamlines the process of data entry, review, validation, and summarization that is consistent with the capabilities of the 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
- inform project staff of changes and advances in data management practices

The GIS specialists manage spatial data themes associated with Network inventory and monitoring projects, as well as other spatial data related to the full range of park resources. They incorporate spatial data into the GIS. They also maintain standards for geographic data and are responsible for sharing and disseminating GIS data throughout the Network.

The GIS specialists will:

- determine the GIS data and analysis needs for the project
- develop procedures for field collection of spatial data including the use of GPS and other spatial data collection techniques
- display, analyze, and create maps from spatial data to meet project objectives
- properly document data in compliance with spatial metadata standards

GIS specialists will also work directly with data managers to:

- design databases and other applications for the Network
- create relationships between GIS and non-spatial data
- create database and GIS applications to facilitate the integration and analysis of both spatial and non-spatial data
- establish and implement procedures to protect sensitive spatial data according to project needs
- develop and maintain an infrastructure for metadata creation and maintenance
- ensure that project metadata are created and comply with national and agency standards

### ***2.2.3 Project Leaders***

Since the data management aspects of every inventory or monitoring project normally require the expertise and involvement of several people over a period of months or years, it makes sense that one person is charged with keeping track of the objectives, requirements, and progress for each project. This project leader (or *steward*) is usually a resource management specialist with training and experience in the field of science related to the inventory or monitoring project.

The project leader may have worked in the geographic area where the study occurs. This can be helpful as it gives the project leader a background of information to use when overseeing field work, coordinating with GIS and data managers on information management needs, evaluating data resources, and understanding the project's objectives. A project leader who cannot act as an authority for the data should work with the appropriate resource specialists to account for those aspects of data stewardship.

To ensure the quality of each project, including data requirements, project leaders should be assigned only those projects they can effectively oversee based on workload and other relevant factors. Unless the project is short term (three months or less) the project will have at least one alternate or backup project leader to provide continuity in case the principal project leader becomes temporarily or completely unavailable.

### ***2.2.4 Project Leaders-Roles and Responsibilities***

The project leader is responsible 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. Figure 2.1 illustrates some of the overlap between project and data management responsibilities.

Specifically, a project leader is responsible for:

- *project documentation* that describes the 'who,' 'what,' 'where,' 'when,' 'why,' and 'how' of a project
- documentation and implementation of *standard procedures* for field data collection and data handling
- quality assurance and *quality control measures*, which include the supervision and certification of all field operations, staff training, equipment calibration, species identification, data collection, data entry, verification, and validation
- maintenance of concise explanatory documentation of all deviations from standard procedures
- detailed documentation for each field data collection period
- maintenance of hard copies of data forms and archiving of original data forms
- scheduling of regular project milestones such as data collection periods, data processing target dates, and reporting deadlines
- regular summary reports, periodic trend analysis of data, resulting reports, and their public availability
- acting as the main point of contact concerning data content

The project leader will also work closely with the data manager to:

- develop quality assurance and quality control procedures specific to project operations
- identify training needs for staff related to data management philosophy, database software use, and quality control procedures
- coordinate changes to the field data forms and the user interface for the project database
- 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 the process of how project data will be transformed from raw data into meaningful information
- create data summary procedures to automate and standardize this transformation process
- identify and prioritize legacy data for conversion and convert priority data sets to a modern format
- increase the interpretability and accessibility of existing natural resource information

### ***2.2.5 Sharing Responsibilities***

Keeping track of data from the time of acquisition until they are no longer useful is the shared responsibility of everyone involved with data—whether producer, analyst, manager, or end user. This, in essence, is *data stewardship*. It is a principle of mutual accountability rather than one job for one individual.

Successful data stewardship requires that people involved in NCBN activities learn and understand Network expectations for continuous information management. This is equally important for Network staff, park employees, and contractors or cooperators. All project participants receive training, briefings, materials, and additional regular communication about data stewardship from supervisors, project leaders, and data managers. Thus, everyone learns how their efforts relate to Park and Network management objectives, National Park Service and Department of Interior policies, and other federal government requirements.

Inventory and Monitoring project leaders must recognize and understand resource information management issues and requirements. They must accept the limitations of data collecting and processing, yet welcome the challenges. Well-trained field crews, and an ongoing plan for educating field crews in appropriate technology and skills (e.g., GPS, data Q/A, Q/C) will significantly improve all aspects of data processing.

Careful documentation of data sets, the data source(s), and the methodology by which the data were collected or acquired is one of the most important shared responsibilities. Documentation establishes the basis for the appropriate use of the data in resulting analysis and products, both in the short-term and long-term. Network monitoring protocols contain key elements of data documentation. Network data records collected according to these protocols will include the name, date, and version of the associated protocol. (See Chapter 7, Data Documentation, for important guidance and reference for documentation and metadata.)

### ***2.2.6 Communication***

Project leaders, data managers, and GIS specialists comprise the central data management team for inventory and monitoring projects. Each is responsible for certain aspects of project data, and all share responsibility for some overlapping tasks. Because of the collaborative nature of project information management, good communication among project leaders, data managers, and GIS specialists is essential to meeting program goals.

## ***2.3 Acknowledgements***

This chapter was adapted from concepts and material developed by Rob Daley (GRYN).

## **Chapter 3**

### **Information Management Resources: Infrastructure and Systems Architecture**

#### ***3.1 Computer Resources Infrastructure***

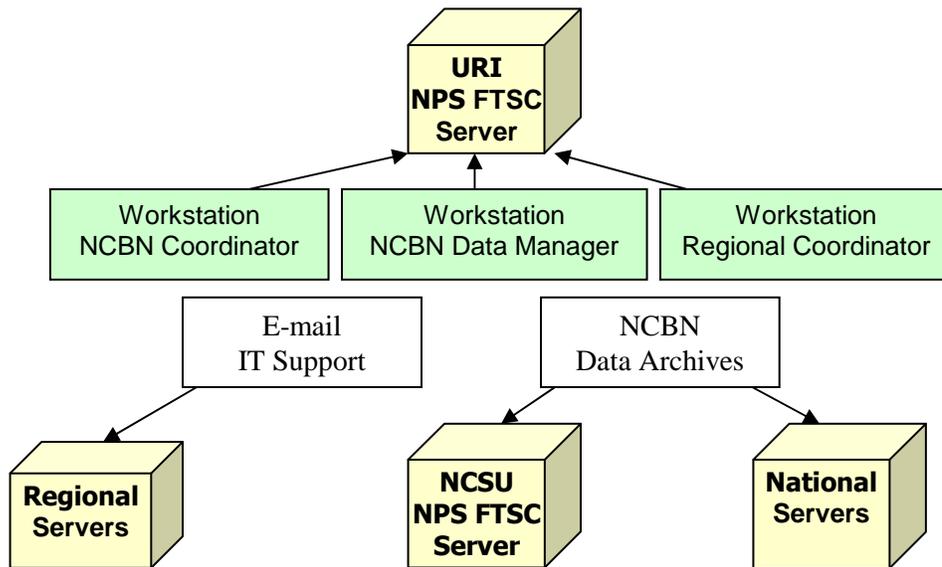
Core NCBN staff and the Regional I&M Program Coordinator are stationed at the University of Rhode Island (URI) in Kingston, Rhode Island. Network staff are part of the NPS Field Technical Support Center and are managed as part of the Environmental Data Center, which is within the department of Natural Resource Sciences at the University. NCBN staff rely on the computer resource infrastructure of the URI FTSC for daily staff usage.

The NCBN I&M Program and the Regional I&M coordinator have developed a cooperative agreement with the University of Rhode Island to work closely with the URI NPS FTSC. This agreement designates NCBN staff computers to be maintained as part of the center's computer Network. The benefits of this inclusion are automated backups and archiving of working data and information on a weekly basis, technical assistance, and complete maintenance of the computer resource infrastructure necessary to maintain the Network's information management program. This agreement also includes, but is not limited to hardware replacement, software installation and support, security updates, virus-protection, and telecommunications Networking.

NPS NCBN staff stationed at URI will also rely on IT specialists of the Northeast Region Boston Support Office for technical support with NPS licensed software. NPS staff e-mail is maintained on servers at the Washington level, but supported by the Boston Office technical support staff.

An important element of a data management program is a reliable, secure Network of computers and servers (Figure 3.1). The NCBN digital infrastructure has four main components: URI FTSC Network that supports the daily working files of NCBN staff stationed at URI, NCSU FTSC server that maintains a digital archive of Network generated data and products, Regional office servers that maintain Network staff e-mail, and servers maintained at the national level that also maintain Network generated data. This infrastructure is maintained by IT specialists, who administer all aspects of system security and backups for all four servers.

**Figure 3.1 Server and computer resources used by NCBN**



Each component depicted in figure 3.1 hosts different aspects of the Network’s information management program.

*National servers*

- Master applications–integrated client-server versions of NatureBib, NPSpecies, NR-GIS Metadata Database
- Centralized repositories–NR-GIS Data Store, Protocol Clearinghouse
- Public access sites–portals to NatureBib, NPSpecies, NPSFocus, and websites for monitoring Networks
- NCBN website

*FTSC servers*

- Working and finalized Network administrative records–proposals, study plans, cooperative agreements, contracts
- Working copies of the Network website
- Local applications–desktop versions of national applications such as NPSpecies and Dataset Catalog
- Working project databases–annually compiled datasets for both long-term and short-term projects.
- Master project databases–compiled datasets for monitoring projects and other multi-year efforts that have been certified for data quality
- Park GIS files–base spatial data, imagery, and project-specific themes
- Project tracking application–used to track project status, contact information, product due dates

## NCBN Information Management Plan

- Network digital library–Network repository and archive for finished versions of project deliverables for Network projects (e.g., reports, methods documentation, data files, metadata, etc.)

### *Regional Servers*

- Network Administrative files–cooperative agreements, contracts, budgets, personnel records
- NCBN staff e-mail

### **3.1.1 NCBN Digital File Maintenance and Storage**

The Network is in the process of developing a new agreement with the URI FTSC that will enhance the organization, maintenance, and storage of NCBN staff digital files, both working and permanent. This will include digital files of temporary staff working on I&M projects hired through the FTSC, as well as the Network and Regional staff stationed at the University. Currently, all files are maintained only as part of each individual's workstation, and no standards for file structures or file maintenance are in place.

An SOP will be developed to describe in detail how all Network digital files will be maintained, stored, and backed-up at the FTSC. This document will also include a standardized file directory structure that will be used by all Network and temporary I&M staff. This document will provide standardized naming conventions for all digital files. Table 3.1 provides some examples of directory structures and file types that could be developed through the URI FTSC agreement.

**Table 3.1 Example directory structures and associated digital files**

<b>Directory</b>	<b>Types of files maintained</b>
Administration	Only working files included, all permanent files are maintained by the Boston Support Office Administrative Branch and maintained on the regional server
Databases	Local copies of national databases (e.g., NPSpecies), working monitoring databases both local copies and the master database
Libraries	Read-only, finished project deliverables
Working	Storage of staff working files and draft material
GIS	Base spatial data, imagery, and project-specific themes

The key aspects of any file management strategy are as follows:

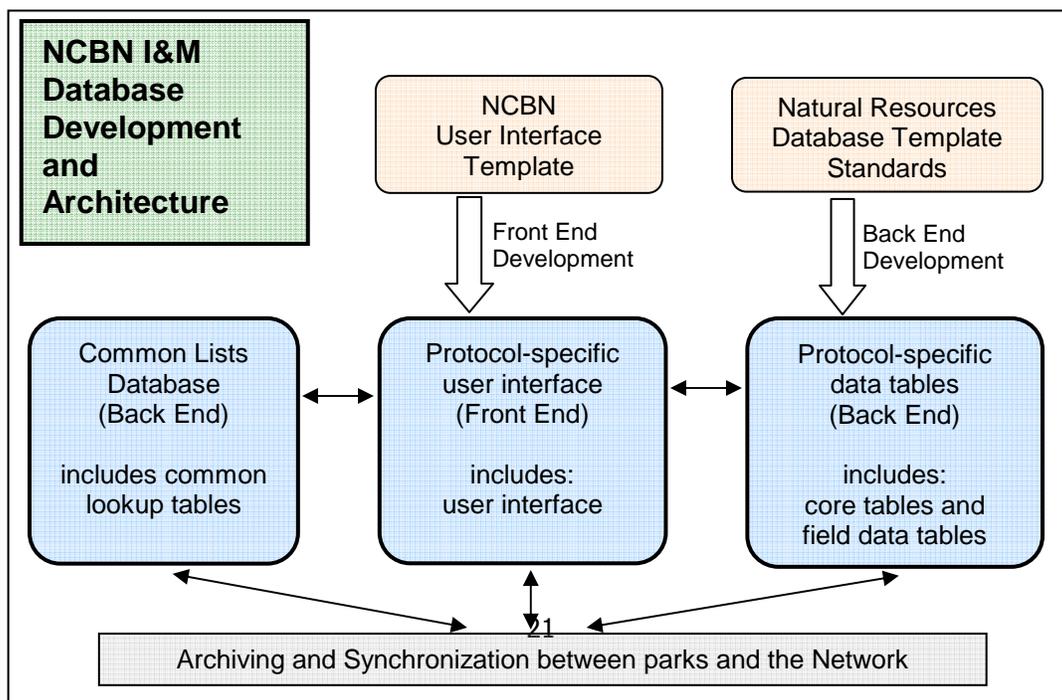
1. Working files are kept separate from finished products.
2. Finished products are typically read-only, except for "inbox" folders where users can drop things off to be cataloged and filed.
3. Standards such as naming conventions and hierarchical filing are enforced within the Libraries, Database, and GIS sections. Although less stringent in other sections, these conventions are encouraged as good practice.

### 3.2 NCBN Systems Architecture

The NCBN database design relies upon modular, standalone project databases that share design standards and link to both protocol-specific and common data tables (Figure 3.2). Individual project databases are developed, maintained, and archived separately. There are numerous advantages to this strategy:

- Databases are modular, allowing greater flexibility in accommodating the needs of each project area. Individual project databases and protocols can be developed at different rates without a significant cost to data integration. In addition, one project database can be modified without affecting the functionality of other project databases.
- The databases are used by Network and regional staff and parks on the NPS intranet and by researchers who are not on the NPS intranet. The most appropriate solution to a centralized database for all to share would be a Web interface to an Access database or databases. Currently, most NPS data managers are not proficient in developing or modifying these Web interfaces, and the web servers are accessible to only a limited number of park staff. It is not currently feasible to develop these Web interfaces for centralized databases. Creating functional modular databases now, does not preclude developing a Web interface to them in the future.
- By working up from modular data sets, a large initial investment is avoided in a centralized database and the difficulties of integrating among project areas with very different—and often unforeseen—structural requirements. Furthermore, the payoff for this initial investment may not be realized down the road by greater efficiency for interdisciplinary use. Meanwhile the databases are available for use as the monitoring projects begin, rather than several years after.

**Figure 3.2 NCBN database development and architecture**

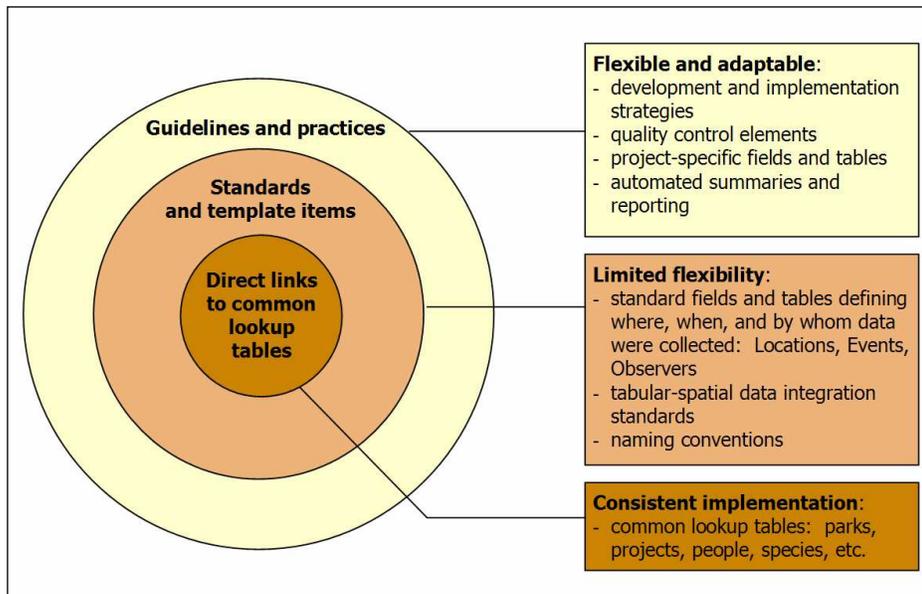


### 3.2.1 Project Database Standards

Project database standards are necessary for ensuring compatibility among datasets and across Networks, allowing managers and researchers to compare data and analyze trends across parks or across protocols. Standards also help to encourage sound database design and facilitate interpretability of datasets. The NCBN-developed databases for Park and Network projects will contain the following main components: core tables, protocol-specific field data tables, common lookup tables, and a user interface.

The three types of database tables correspond to three levels of data standards. Because common lookup tables are stored in one place and are referred to by multiple databases, they represent the highest level of data standard because they are implemented identically among data sets. The second level of standards is implied by the core template fields and tables, which are standardized where possible, but project-specific objectives and needs could lead to varied implementations among projects. The third level of standards is applied most flexibly to accommodate the range of needs and possibilities for each project, yet always with compatibility and integrity in mind. Figure 3.3 presents the resulting variation in implementation of these differing levels as a “bull’s eye,” with the common lookup tables providing the most consistent implementation and hence the smallest range of variation.

**Figure 3.3 Various levels of database design standards**



### 3.2.2 Core tables and fields based on Network and national standards

Core tables manage the information describing the “who, where, and when” of project data and of the database and are required for all inventory and monitoring databases (Figure 3.4). Core tables reside in each individual project database and are populated locally. These core tables are

outlined in the Natural Resources Database Template (NRDT) standards that specify data tables, relationships, data types, field names, and domain ranges.

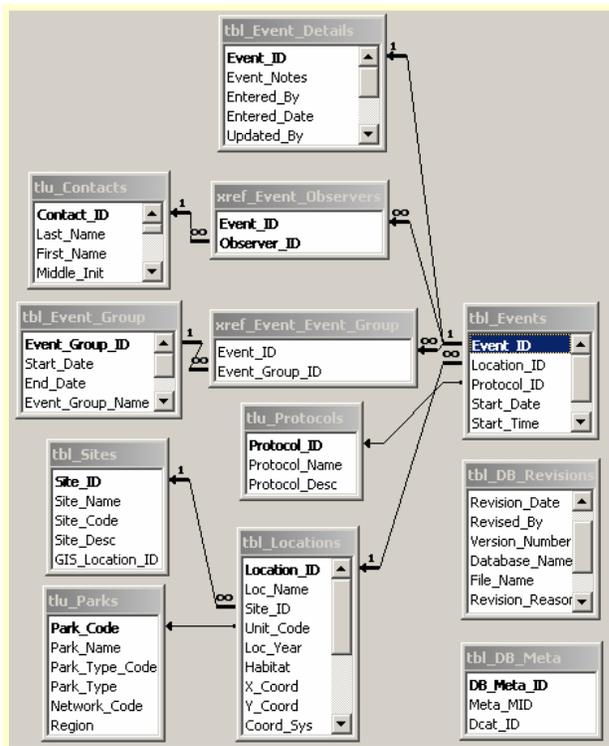
Recommended tables provide additional organization to the core tables. These store supplemental information about the monitoring events, including tables to facilitate aggregating event data in time and space, as well as information about the personnel involved in field work and data entry.

Core fields are the required fields in the core tables. Additional fields may be added to the core tables. Recommended tables will include fields to be used, if the recommended tables are used. The NCBN databases will follow the NRDT recommended fields and tables to the greatest extent possible. The use of these fields and tables will vary by database.

Refer to the NRDT Standards for a complete presentation of the core tables, fields, and standards. In brief, the core tables describe the monitoring events: what protocol was used, where and when the data were collected, and who performed the fieldwork. Groupings of events in time and space are optional tables that will be used in most databases. Two additional tables describing database revision history and linking to the database metadata are also core tables.

If NRDT standards change, all new databases will conform to any new standards revisions. The NCBN will convert existing databases to the updated standards on an as-needed basis.

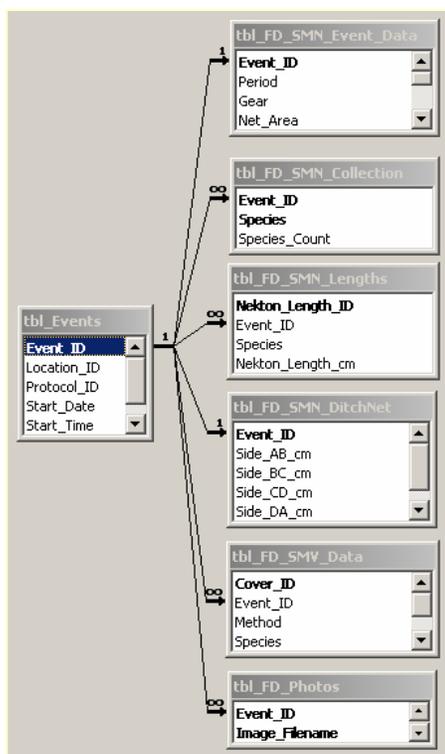
**Figure 3.4 NCBN NRDT core and recommended tables**



### 3.2.3 Protocol-specific Field Data Tables

Each database will have tables specific to that protocol (Figure 3.5). These tables and their fields cannot be entirely specified at a national level, since each protocol has different requirements. Within protocols that are common across the Service, standards for tables and fields may be developed in the future. As much as possible, the Network will comply with these standards when developing new databases, to keep Network databases as compatible with those maintained by other Networks and cooperators managing similar data sets as possible—especially if integration with other data sets is important for meeting project objectives.

**Figure 3.5 NCBN NRDT example project data tables**



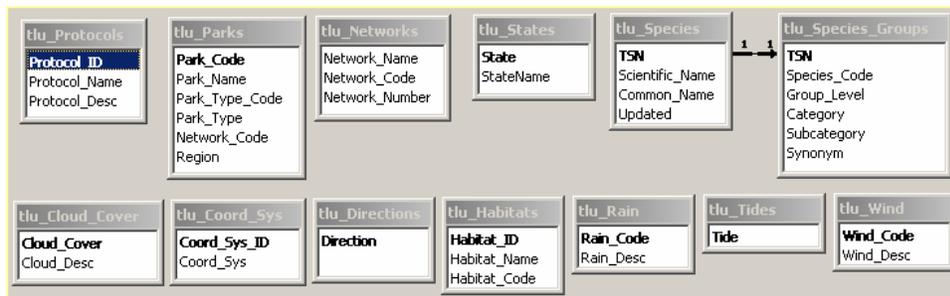
### 3.2.4 Common Lookup Tables

Certain key information is not only common to multiple data sets, but to the organization as a whole (i.e., lists of contacts, projects, parks, species, etc.). By linking the tables, this common data will be stored in a common lists database that serves as a back end to the various protocol-specific I&M databases. Using one common database serves multiple purposes. It requires only one development, and thus it reduces start-up time. Network applications—for project tracking, administrative reporting, or budget management—can also link to the same tables. Thus, all users in the Network have instantaneous access to edits made by other users.

Not all users will be on one Network, however, so the common lists database will be synchronized frequently between the Network, parks, and collaborators. Sharing directly over the Network is preferable where possible, as it avoids redundancy and versioning issues among multiple copies. Many of the tables will not change, such as lists of parks, states, wind directions, cloud cover, etc., and for those tables, versioning will not be a primary concern. Other tables, particularly species and contacts will change a lot in the early years and then less so in later years. These tables are the ones that will require synchronization at least annually.

The NCBN has already developed many of these lookup tables (Figure 3.6). If national standards are developed for some of these lookup tables, the NCBN will adapt existing lookup tables to the new standards as appropriate.

**Figure 3.6 NCBN NRDT common look-up table examples**



### 3.2.5 NCBN NRDT Template User Interface

The NCBN has developed a user-interface template to facilitate the use of the monitoring databases, for data entry, tracking, analysis, and exporting (Figure 3.7). This interface is part of the NCBN I&M database structure and will be used for all new databases. The interface template complies with all current NRDT standards and includes conversion tools for adapting the interface to new databases.

The advantage of the interface template is that it includes front-end design elements that are common across protocol databases. These include NPS format and design, main forms development, menus for navigation, pick lists for display and export, data backups, etc.. These menus and utilities can easily be incorporated into any new monitoring database by data managers who do not have the programming background to develop these utilities independently.

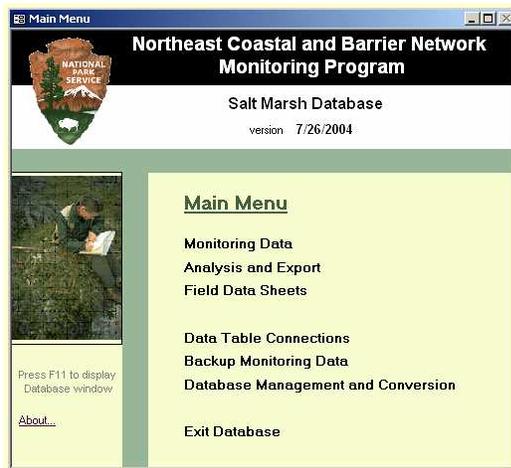
The advantages of using a common interface template include:

- common look and feel across databases
- minimal front end development time
- common utilities across databases
- utilities for data managers without specific programming skills
- NRDT core table standards incorporated automatically

## NCBN Information Management Plan

- complete database model for emulation and reference

**Figure 3.7 NCBN NRDT user interface example**



### ***3.3 Project Tracking Application***

To support program coordination and annual reporting, and to improve accountability for the products of the Network's natural resource inventory and monitoring efforts, the Network has developed and implemented a project tracking database as part of its data management program. Currently, the Network uses project management software called "KickStart." The primary function of this application is to track:

- cooperative agreements and associated projects (cooperative agreement beginning and end date)
- project funding schedules
- project deliverables
- deliverable schedule (draft deliverables received, draft deliverable review completed by Network staff, review returned to cooperator for final draft completion and final products received)
- contact information

This application will eventually be made available on the web so that it is accessible to all NCBN staff. Although primarily maintained by NCBN data managers, the database will be available to project leaders, GIS staff, the Network coordinator, and other Network administrators. Each of these staff will be able to make certain changes to update information about project status, deliverable details, etc..

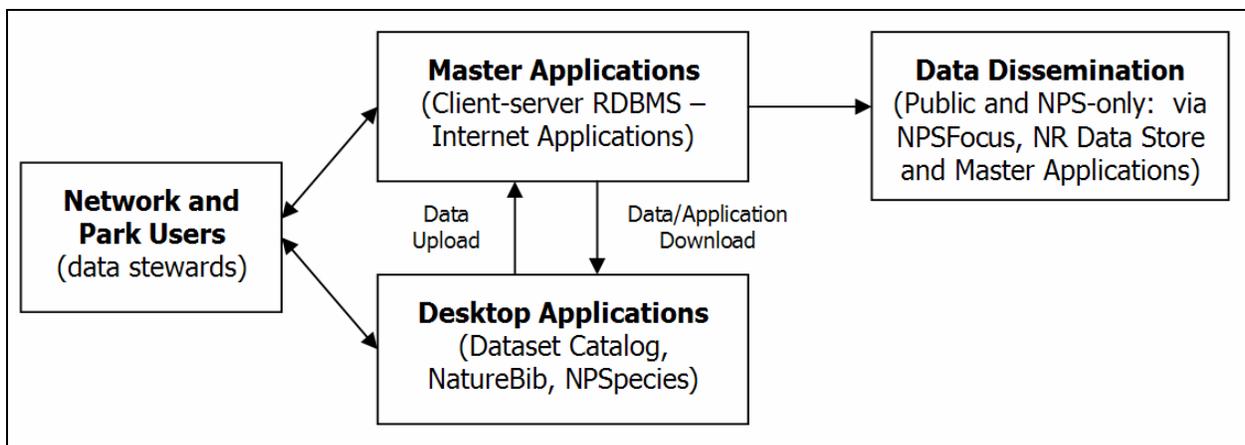
### 3.4 National Information Management Systems Architecture

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.

#### *National-level application architecture*

To achieve an integrated information management system, three of the national-level data management applications (NatureBib, NPSpecies, and NR-GIS Metadata Database) use distributed application architectures with both desktop and internet-accessible (master) components (Figure 3.8).

**Figure 3.8 Model of the national-level application architecture.**



#### **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.

## NCBN Information Management Plan

As with NPSpecies and NR-GIS Metadata Database, it is possible to download data from the master web version into the MS Access desktop version that can be used locally on computers with limited internet connectivity:

<http://www.nature.nps.gov/nrbib>

### **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 assigned for each park and Network. 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>

### **Dataset Catalog and NR-GIS Metadata Database**

Dataset Catalog is a desktop metadata database application developed by the I&M Program to provide a tool that parks, Networks, and cooperators can use to inventory and manage data set holdings. Although not designed as a comprehensive metadata tool, the Dataset Catalog is used for cataloging abbreviated metadata about a variety of digital and non-digital natural resource data sets. The Dataset Catalog helps parks and Networks begin to meet Executive Order 12906 mandating federal agencies to document all data collected after January 1995. It provides brief metadata and a comprehensive list about all resource data sets for use in data management, project planning, and more stringent metadata activities.

As with other service-wide applications, the master metadata database (NR-GIS Metadata Database) is available through a website and will be linked to NPSpecies (the NPS species database) and NatureBib (the bibliographic database). It will be possible to download a version in MS Access format from the master website *Dataset Catalog*:

<http://science.nature.nps.gov/im/apps/datacat/index.htm>

and *NR-GIS Metadata Database*: <http://science.nature.nps.gov/nrdata>).

## **Other National-Level Inventory and Monitoring Information Management and GIS Applications**

### **NPSTORET**

STORET is an interagency water quality database developed and supported by the Environmental Protection Agency's (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 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 service-wide 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 funded WRD project get archived in STORET. NPSTORET (also known as Water Quality Database Templates) the NPS master database designed to facilitate park-level standardized reporting for STORET. The database is still in development, but metadata, protocols, data dictionaries, and reporting capabilities are available through a front-end form. Upon implementation, Network staff and cooperators will be able to use the MS Access version of NPSTORET either as a direct database for data entry and management, or as a means of submitting data for upload to STORET by WRD staff.

The MS Access application and additional details can be found at:

<http://www.nature.nps.gov/water/infodata.htm>

Additional information on STORET can be found at: <http://www.epa.gov/storet>.

### **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 the GIS software (e.g., ArcView or ArcGIS) to enter, store, retrieve, and otherwise manage natural resource information. The template has a core database structure that can be modified and extended by different parks and Networks depending on the components of their inventory and monitoring program and the specific sampling protocols they use. Natural Resource Database Template is a key component of the I&M program's standardized monitoring protocols.

These monitoring protocols include separate modules detailing different aspects of monitoring project implementation, from sampling design to data analysis and reporting, and include data management components that describe database table structure, data entry forms and quality checking routines. Approved monitoring protocols, including the databases that are based on the Database Template, are made available through a web-based protocol clearinghouse (see below).

A description of the Database Template application, a data dictionary, and example implementations are located on the NR Database Template website:

<http://science.nature.nps.gov/im/apps/template/index.htm>

### **Natural Resource Monitoring Protocols Clearinghouse**

The Natural Resource Monitoring Protocol Clearinghouse (i.e., Protocol Database) is a web-based clearinghouse of sampling protocols used in national parks to monitor the condition of selected natural resources. 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 the prototype monitoring parks or other well-established protocols used 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 Natural Resource Database Template 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>

### **NR-GIS Data Store**

The NR-GIS Data Store is a key component of the data dissemination strategy employed by the I&M Program. The NR-GIS 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 resource datasets, inventory products and GIS data produced by the I&M and Natural Resource GIS Programs. Each park or Network is able to post and curate its data on the server. The NR-GIS Data Store will be integrated with the master NR-GIS Metadata Database application to streamline programmatic data documentation and dissemination processes.

The simple browse function of this server can be accessed at: <http://nrdata.nps.gov/>.

See the NR-GIS Data Store website for further information:

<http://science.nature.nps.gov/nrdata>

### **3.5 Acknowledgements**

Thanks to John Boetsch (Olympic National Park), Susan Huse (NCBN) and Lisa Nelson (WASO) for providing material used in this chapter.

## **Chapter 4**

### **Data Management Integration and Project Work Flow**

This chapter describes the general data management work flow throughout project development. It also provides an overview of how natural resource data are generated, processed, finalized and made available. By describing the progressive stages of a project and the life cycle of the resulting data, the overall objectives and specific steps of the data management process can be communicated easily. In addition, by understanding the data management process and flow, staffing resource needs can be better managed to produce, maintain, and deliver quality data and information. Further detail on data acquisition, quality assurance, documentation, dissemination and maintenance can be found in later chapters of this plan.

#### ***4.1 Project Work Flow***

From the perspective of managing workflow, there are two main types of projects:

- *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 mainly will be Network vital-signs 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.

From a data management standpoint, the primary difference between short- and long-term projects is an increased need to adhere to and maintain standards for long-term projects. Maintaining standardization from year-to-year is necessary when comparing data over an extended period of time (years for vital signs monitoring). While the need to follow standards is still present for short-term projects, sometimes the cost of compliance outweighs the benefits due to the scope, budget, and level of NPS influence over project details. Nevertheless, both short-term and long-term projects share many work flow characteristics, and both generate data products that must be managed and made available.

For this plan, projects have been divided into four primary stages: initiation and approval; planning-design and testing; implementation; finalization-product integration and evaluation (Figures 4.1-4.4). In the figures below, main tasks characterize each stage, followed by associated data management tasks. A list of data management guidelines and standard operating procedures that might be used as part of each project phase are listed as well.

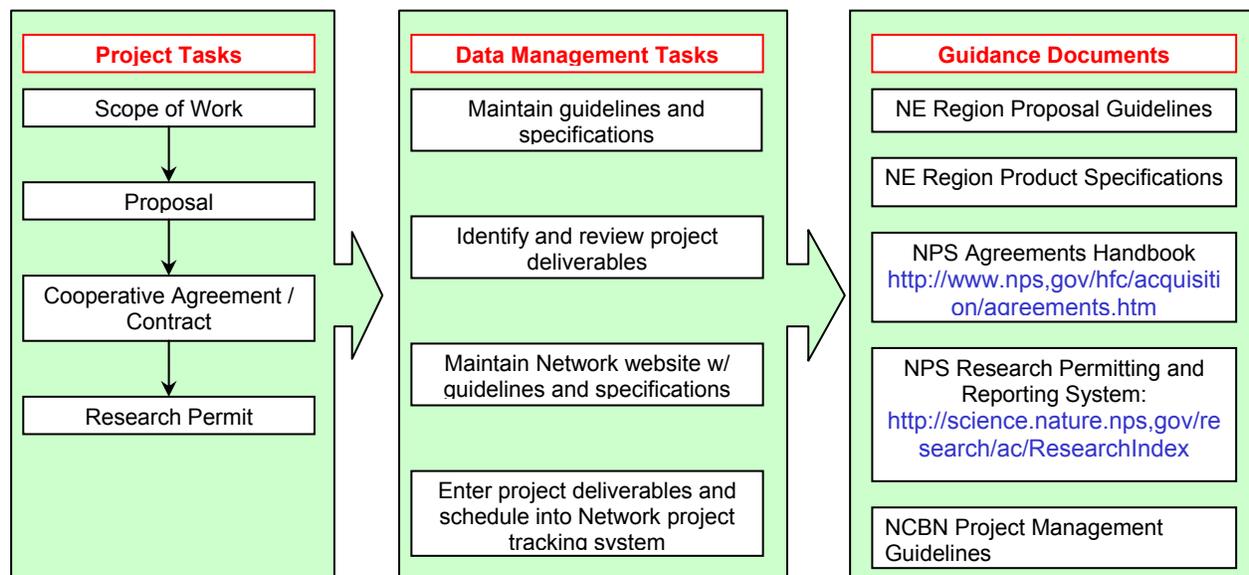
Many of these guidance documents are currently under development by NCBN and regional staff, please refer to the “Table of Contents and Appendices” section at the beginning of this Information Plan for specific status and availability information for each document. Primary responsibility for each of the tasks associated with project phases rests with different individuals, Network staff and contractors. Additional discussion of the different staff roles and responsibilities can be found in Chapter 2 of this Information Plan.

## 4.2 Project Initiation and Approval

During the initiation and approval phase, preliminary decisions are made regarding the scope of the project and its objectives (Figure 4.1). A scope of work may or may not be written for each project, but a proposal will be developed and funding sources, permits, and compliance must be addressed. A cooperative agreement or contract is developed and finalized. Although many of these responsibilities reside with the project leader and/or program administrators, data management staff must be involved in identifying project deliverables and assuring that each contract or agreement include a clear list of these deliverables, with reference to either national, regional or Network information and data standards. All contracts, agreements, and permits should include standard language that describes the formats, specifications, and timelines for project deliverables (for examples, see Appendix 28–*NCBN Example Language Describing Deliverables for use in Cooperative Agreements and Contracts*).

Other data management tasks associated with this project phase include assisting in the development and maintenance of proposal guidelines and study plans, as well as assuring that information will be accessible to the public through a planned distribution mechanism (NCBN website, Regional websites). Development and maintenance of project deliverable guidelines and specifications for reports, spatial and biological data, digital and analog photos, and metadata (FGDC and Biological Data Profile compliant) are important components in assuring quality project products. A final data management task for this phase of a project includes entering project deliverables and schedule into the Network’s project tracking system.

**Figure 4.1 Project initiation tasks and associated Northeast Region, NCBN and National guidance documents**

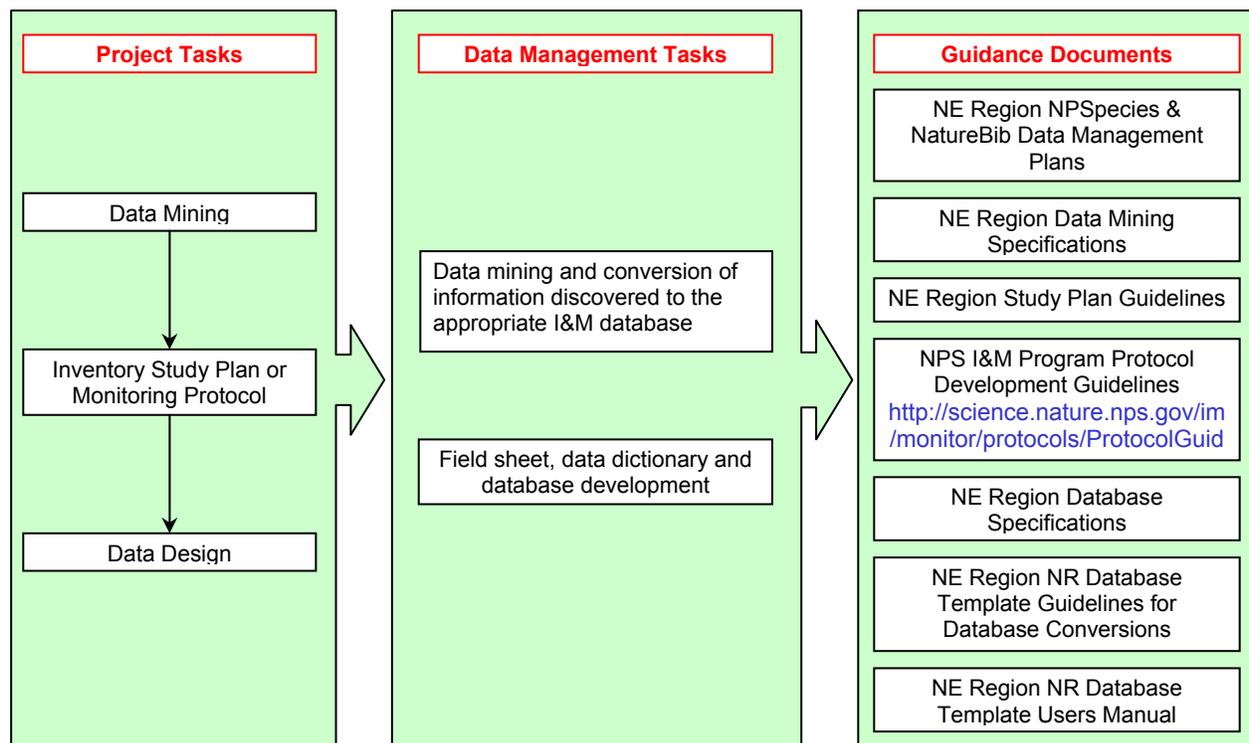


### 4.3 Planning: Design and Testing

During the design and testing phase, data mining specific to the proposed project may occur. Data management staff will convert these results to I&M Program standards and formats and enter this information into the appropriate national I&M or Network database. At this stage of a project either an inventory study plan or monitoring protocol will be developed that details how data will be acquired, processed, analyzed, and reported. It is critical that the project leader and the data manager work together throughout this phase. Close collaboration will help to assure the integrity and quality of the resulting data and information. Data management staff should assist with data design (including field sheet, database, and data dictionary development). Devoting adequate attention to this aspect of project development is critical to assuring the quality, integrity and usability of the resulting data.

Again data management tasks associated with this project stage include data design, development and maintenance of guidelines and specifications, and dissemination of this information (Figure 4.2).

**Figure 4.2 Project planning tasks and associated Northeast Region, NCBN and National guidance documents**



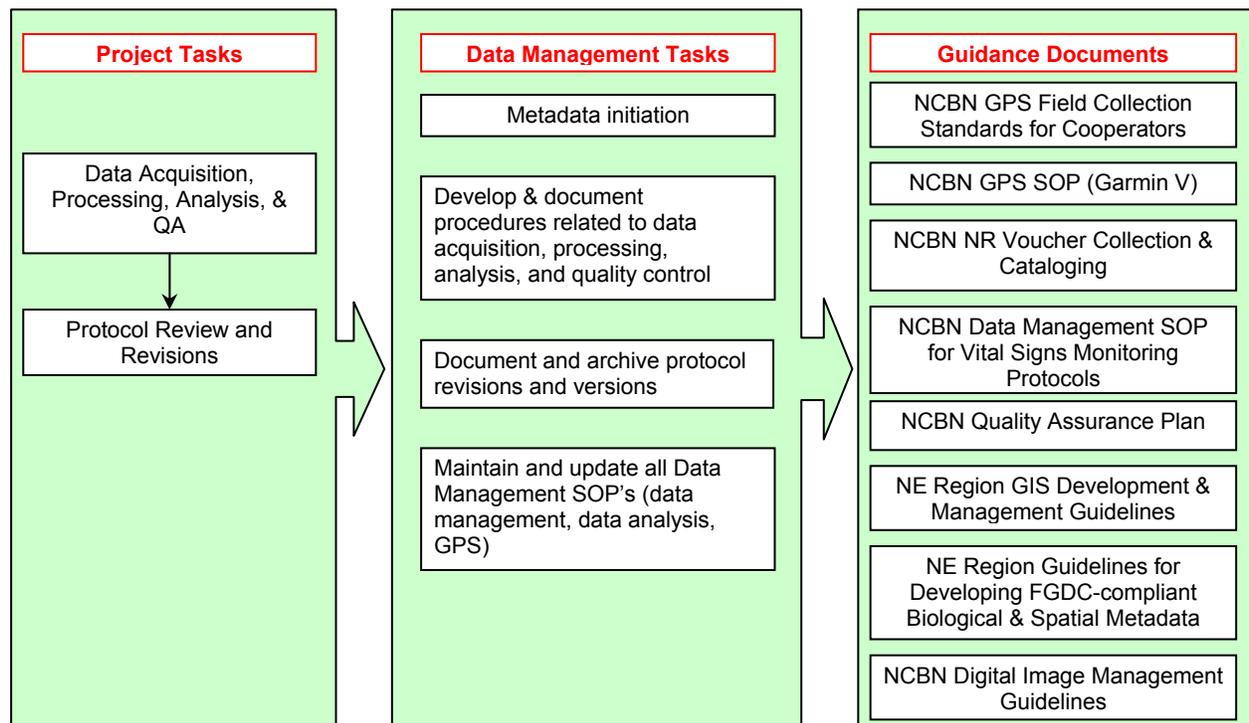
#### 4.4 Implementation

Tasks specific to the implementation phase of a project include acquiring, processing, and documenting data (Figure 4.3). During this phase, products such as reports, maps, and GIS themes, are developed. All raw data undergo QA/QC measures. All manipulated products are reviewed (see chapter 6, QA/QC procedures). The project leader oversees most aspects of implementation—from logistics planning, contracting, training, and equipment procurement to data acquisition, report preparation, and final delivery of data products to Network data management staff. Throughout this phase, data management staff provide training and support for database applications, GIS, GPS and other data processing applications. They provide assistance with data summarization, validation, and analysis, as well as the technical aspects of documentation and product development.

Other data management tasks indirectly associated with this project phase include developing and maintaining guidance documents and standard operating procedures often used during this phase. These include GPS guidelines, metadata creation guidelines, and others.

Toward the end of this phase, project staff members work to develop and finalize project deliverables. All raw and derived data products, metadata, reports and other documentation are delivered to the data steward assigned to the project. Administrative records are delivered to appropriate park, Network and regional staff.

**Figure 4.3. Project implementation tasks and associated Northeast Region, NCBN and National guidance documents**

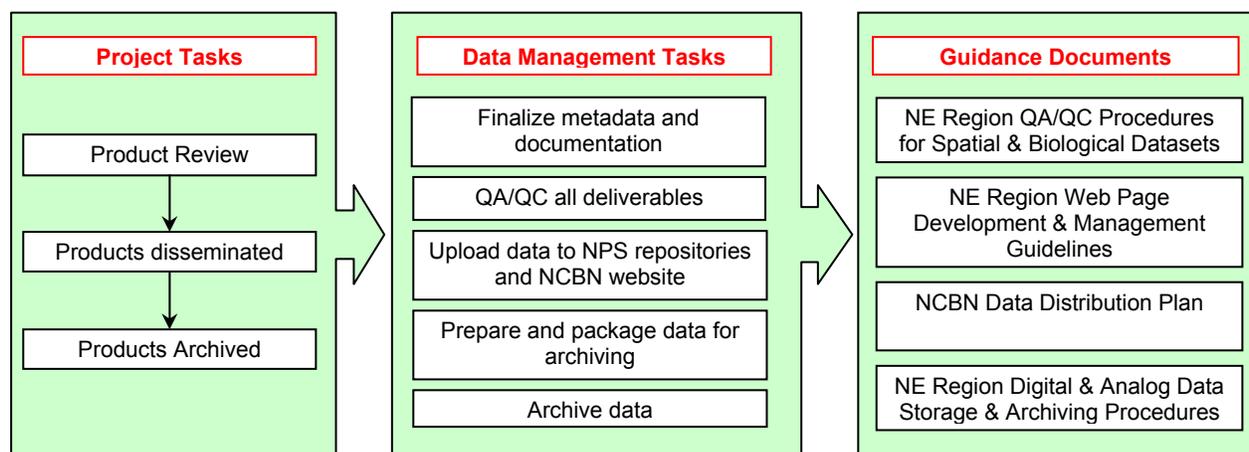


### 4.5 Project Finalization

Tasks associated with the finalization of a project include the review, dissemination, and archiving of these products. See Figure 4.4. The majority of final project phase tasks are data management oriented. An important part to this final phase is the review and QA/QC of all products either developed throughout the lifespan of the project if it is an inventory, or end-of-year products if it is a long-term monitoring project. Metadata must be finalized following the product review and all QA/QC procedures conducted throughout the project be included in the record. A bibliographic record must be created in NatureBib and NPSpecies updated to reflect new species occurrence information. At this stage all products are submitted to the appropriate repository, and project metadata records are posted to the NR-GIS Data Store. All project information is made searchable via service-wide search engines and made available via service-wide and Network specific repositories (See Chapter 9).

Depending upon the type of project this phase may represent the conclusion of a project. This is the case for biological inventories. For those projects, the project record is finalized in the Network project tracking database to reflect the final status of the project and its associated deliverables. For long-term monitoring and other cyclic projects, this phase represents the end of a project year. At this phase, program administrators, project leaders, and data managers assess how well the project met objectives. Improvement of methodology, implementation, and formats of the resulting information are considered. Changes to standard operating procedures are fully documented and tracked, and the database is updated to reflect these changes. Additional peer review is required if major protocol changes are made.

**Figure 4.4 Project finalization tasks and associated Northeast Region, NCBN and National guidance documents**



### 4.6 Acknowledgements

This chapter was adapted from concepts and material developed by John Boetsch (OLYM).

## **Chapter 5**

### **Data Acquisition and Processing**

The NCBN Inventory and Monitoring Program's long-term data management plan will:

- integrate natural resource inventory and monitoring information into National Park Service planning, management, and decision making
- enable the sharing of National Park Service information with other natural resource organizations, while forming partnerships for attaining common goals and objectives

To enable and maintain such an ambitious long-term system of NPS information sharing, NCBN must 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 for management decision-making, research, and education (DMVision\_Framework.doc Aug 2004). This will require standard methods and procedures for processing data, which the Network will develop and use.

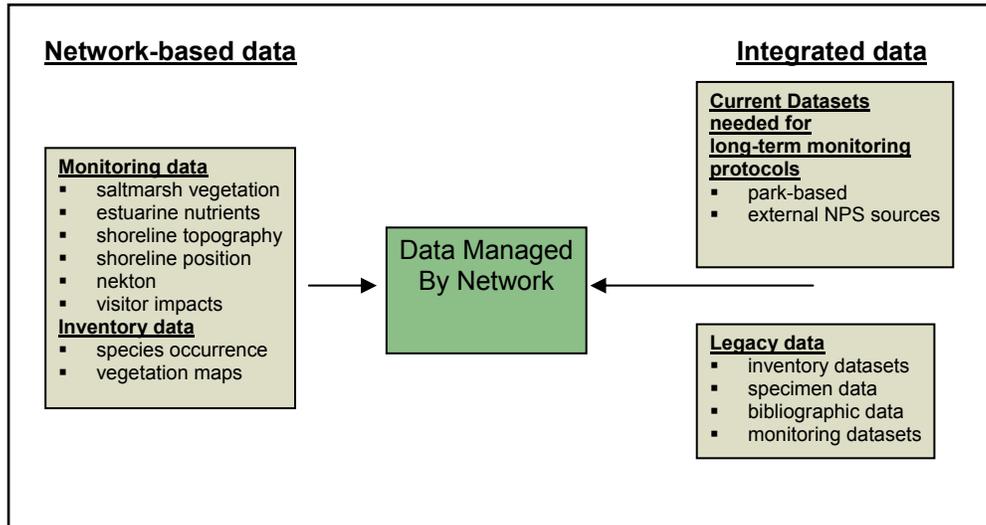
#### ***5.1 Categories of Network Data***

The Network acquires and manages two categories of data:

1. **Network-based data**—those data collected by Network staff and/or cooperators working with the Network
2. **Network-integrated data**—those data collected by other entities (parks, universities, other agencies, other NPS programs), but identified as important natural resource data, necessary for the Network to manage

These categories do not in any way indicate or rank the importance of either type of data source. This distinction is necessary only to facilitate the description and understanding of how data are acquired and processed between the two (see Figure 5.1). The importance or value placed on all datasets is based on the quality and usefulness of the data to assist NCBN park staff in managing and maintaining each park's ecological health. This chapter describes the necessary steps for acquiring and processing these data. Data collected directly by the Network follows a different more lengthy process than those collected by others but managed by the Network.

**Figure 5.1 Sources of data acquired and managed by the Network**



### ***5.1.1 Network-based Data***

Network-based data are those data originating within the Network or currently being collected by NCBN staff. These include three of the twelve basic I&M Program biological inventory datasets:

- vegetation maps
- species occurrence inventories
- species distribution inventories

Along with vegetation maps and species inventories, all of which include multiple datasets, the Network manages long-term monitoring data as part of its Vital Signs program. The Network is currently developing protocols to monitor:

- salt marsh vegetation
- estuarine nekton
- geomorphologic change
- estuarine nutrient enrichment
- land change
- visitor impacts
- submerged aquatic vegetation

Although all of these projects are different, they follow the same basic data acquisition and processing steps. Two main exceptions include the vegetation maps and components of the geomorphologic change protocol data. These both require aerial photography or LIDAR data which are acquired using very specific methods. This data management plan will only cover the very basic methods for acquiring and processing data.

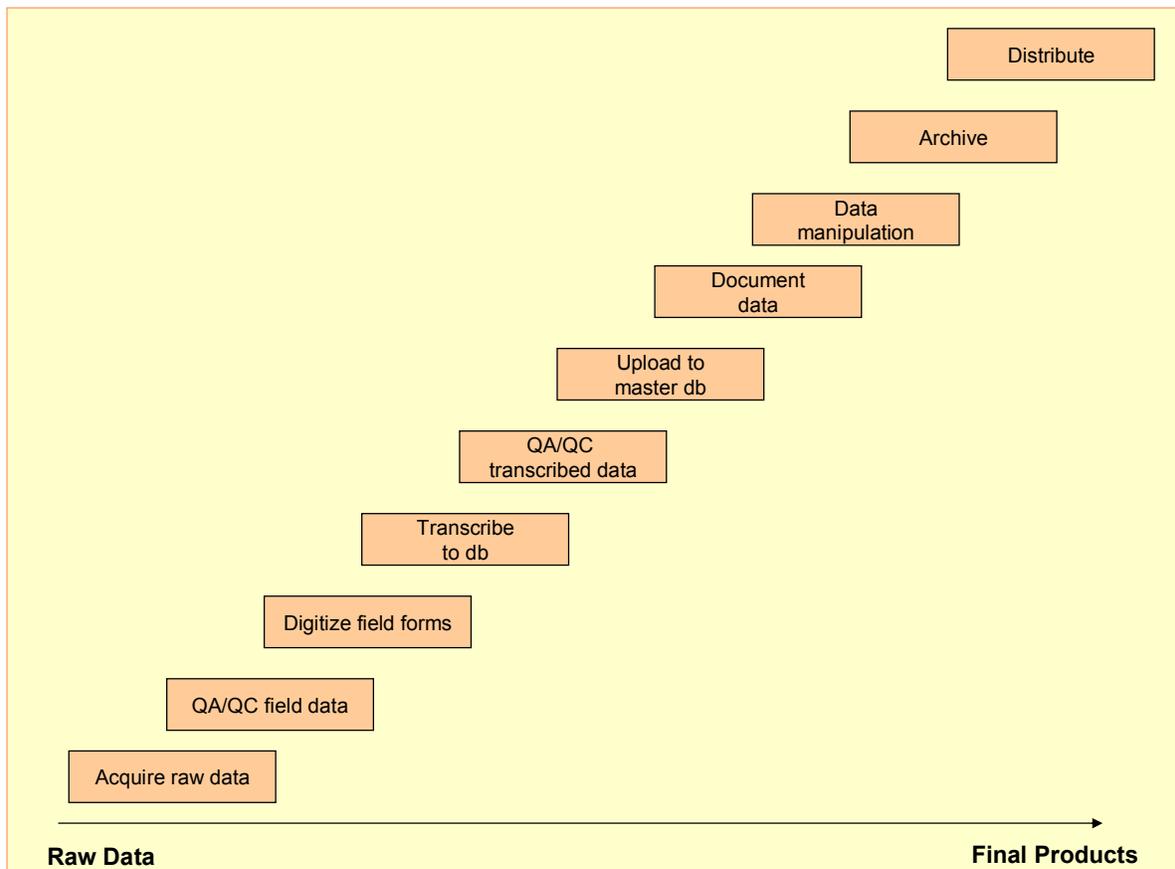
For examples of specifications used for aerial photo acquisition see Appendix 31 (*Northeast Region Aerial Photo Specification Example for scale 1:6000*) and Appendix 32 (*Northeast Region Aerial Photo Specification Example for scale 1:12000*).

For specific information on the acquisition and processing of LIDAR data and vegetation mapping, see Appendix 33 (*Lidar Data Acquisition and Processing Example*), Appendix 34 (*Northeast Region Inventory and Monitoring Program Vegetation Map Development Example*), and Appendix 35 (*Northeast Region Example of Spatial Data Review for Vegetation Mapping Products*).

### 5.2 Acquiring and Processing Field Data (*Biological Inventories and Monitoring*)

The following section describes how Network data are acquired and processed from raw data to the finalized, archived dataset. Although this data flow model is general, most data collected by the Network use a similar model, with a few exceptions mentioned previously. Although data acquisition and processing may vary depending on specific project needs and objectives, the typical stages for NCBN data acquisition are shown in Figure 5.2 and described in Steps 1-10.

**Figure 5.2 Data Flow Model-data acquisition to final products**



*(Note, this diagram was adapted from Approaches to Quality Assurance and Information Management for Regional Ecological Monitoring Programs, Craig Palmer, 2003)*

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

### ***5.2.1 Data Acquisition and Processing, Steps***

Step 1–Data are acquired:

- analog field forms
- GPS units

For both inventory and monitoring projects, all data collected by the Network are currently being recorded on hard-copy field forms. Spatial data are being collected using GPS units that record data in digital format. The Network also uses additional hard copy and GPS field forms (see Appendix 30–*Northeast Region GPS Metadata Form*).

Step 2–Field data QA/QC:

- Field data recorder repeats measures back to data reader as they are given.
- Visual review of all data field forms is completed by field crew same day as data collection.
- GPS data are downloaded at the end of all field sampling periods and compared to hard copy GPS field form and data field form.

Step 3–Digital copies of analog field forms and notebooks are made:

- scanned and converted to .pdf format
- burned to CD and archived

Step 4–Transcription of analog data to digital format:

- Field crew enters data into Access database.

Step 5–QA/QC verification and validation of translated data (see chapter 6 for details):

- Data are printed and compared to digital copy by data entry staff and project leader.
- Spatial data are projected and reviewed with appropriate park spatial datasets (park boundary, water bodies, wetlands, shoreline, etc.).
- Hard copy field forms are archived at this point.

Step 6–Upload verified data to master project database:

- Verified data are uploaded from the working database to the master project database. This step is skipped for short-term projects where there is no need to distinguish working data for the current season from the full set of project data.

Step 7–Documentation (this step is completed once a draft of final report is available):

- FGDC compliant metadata are developed for all spatial data.
- NBII Biological Data Profile is developed for all databases.

Step 8–Data Manipulation:

- Verified data are used to generate data products, analysis, 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. Descriptions of data analyses are described in detail in the Network’s long-term monitoring protocols and inventory study plans.

Step 9–Archive data and information:

- Copies of the final report, verified biological and spatial data and metadata are archived digitally (see Chapter 10).
- Analog copies of final report are archived.

Step 10–Distribute:

- Verified data, metadata and reports are posted to national repositories.
- National databases, NatureBib, NPSpecies, NPSTORET, and NR-GIS Metadata Database are updated.

### ***5.3 Network Water Quality Data Integration***

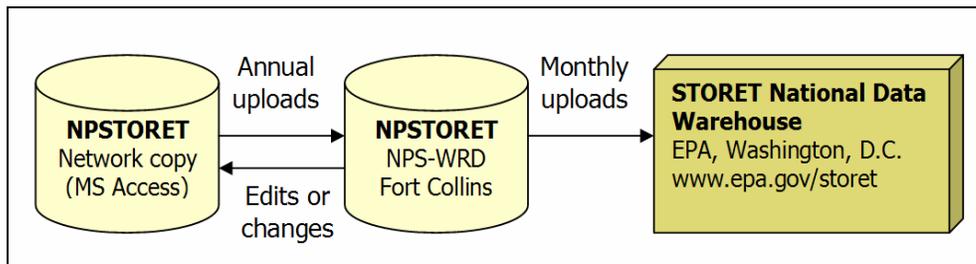
The water quality component of the Natural Resource Challenge (NRC) requires that Vital Signs Networks archive all physical, chemical, and biological water quality data collected with NRC water quality funds in the National Park Service's STORET database maintained by the NPS' Water Resources Division (WRD). Vital Signs Networks are encouraged to also archive any other aquatic-related data collected with base Vital Signs funding in STORET.

To facilitate archiving NPS data in STORET, the WRD has been developing a series of Access-based templates (called NPSTORET), patterned after the Natural Resource Data Templates (<http://science.nature.nps.gov/im/apps/template/>), for all Networks to use to enter their water quality data in a STORET-compatible format. NCBN will send its data from NPSTORET to the WRD on an annual basis for quality assurance and uploading into the NPS' copy of STORET and the Environmental Protection Agency's (EPA) STORET National Data Warehouse.

Although it is the responsibility of Vital Signs Monitoring Networks to ensure that all their NRC funded water quality data are entered in the NPS' copy of STORET, WRD staff will be assigned to assist the Network in this task. Before the initiation of monitoring and/or data entry, WRD staff will advise the Network on how to configure NPSTORET to reflect accurately their water quality monitoring activities. NCBN will perform its own data entry. Annually, NCBN will send the NPSTORET database to the WRD staff contact who will quality assure the database and, in

consultation with the Network, make any edits/modifications necessary to upload the data into the NPS' copy of STORET in Fort Collins using the STORET Import Module (SIM). If necessary, a revised version of a Network's NPSTORET database will be returned to the Network. On a monthly basis, the NPS' copy of STORET will be transmitted to the EPA for inclusion in the STORET National Data Warehouse ([http://www.epa.gov/storet/dw\\_home.html](http://www.epa.gov/storet/dw_home.html)) where the data will be accessible to the states, other entities, and the public on the Internet. Figure 5.3 shows how water quality data goes from the local Network version of STORET to the National STORET database.

**Figure 5.3 Data flow diagram for water quality data**



#### 5.4 Network-Integrated Data

Network-integrated data, as defined by NCBN, are those data collected by entities outside of the Network (parks, universities, other agencies, other NPS programs). These data have been identified as important natural resource information necessary for the Network to manage. These data help to provide park managers with the information they need to make better-informed decisions and to work more effectively with other agencies and individuals for the benefit of park natural resources. The Network has classified these data into two categories:

1. *Current or ongoing datasets*—These datasets are pre-determined for acquisition by the Network and follow very specific acquisition and processing steps. These can either be park-based or from external NPS sources. They are protocol dependent datasets necessary for the Network to acquire at the time of their collection or soon after. These datasets are either used to accompany Network-based monitoring data for data analyses and comparison purposes only, or they are baseline datasets obligatory for the completion of a vital sign monitoring protocol. Examples of these datasets include:
  - weather data-needed for analyses and comparison purposes
  - seagrass distribution data-acquired from states, obligatory for monitoring seagrass in the Network and part of the seagrass vital sign component of the estuarine nitrification protocol
2. *Legacy datasets*—are those data found and compiled through the data mining processes. These include vertebrate and vascular plant species data, other important natural resource

inventory data, specimen or voucher data, bibliographic data, and existing monitoring datasets. Examples of these datasets include:

- species data entered into NPSpecies (1993 Amphibian and reptile inventory of Assateague Island National Seashore)
- bibliographic data entered into NatureBIB (O'Connell, Allan F. 1980—"The relationship of mammals to the major vegetation communities in Gateway National Recreation Area—Jamaica Bay Wildlife Refuge, Breezy Point, and Sandy Hook—including soil analysis of selected areas." Master's Thesis. New York University, New York, NY)
- University of Maryland Collection (400 vascular plant specimens collected by Higgins on Assateague Island National Seashore)

#### ***5.4.1 Acquiring and Processing Current or Ongoing Datasets***

Each NCBN protocol will describe in detail how data are to be acquired and processed from outside sources. Many of these SOPs are still in development, but as the Network monitoring plan moves from draft to final version, all outside data sources will be identified and guidelines for acquiring these data will be put in place. All other datasets will follow the acquisition and processing steps described in the next section.

#### ***5.4.2 Acquiring and Processing Legacy Datasets***

All legacy data and information are discovered by the Network through data mining efforts. Legacy data mining and compilation efforts will be an ongoing component of the Network's data management program. These efforts will be directed toward:

- *species data* for the population of the Network's NPSpecies database
- *natural resource bibliographic* data for the population of the Network's NatureBib database
- *specimen data* for the population of the Network's museum specimen tracking database
- *long-term monitoring data*
- *spatial data*

The four legacy data types follow the data acquisition and processing steps outlined in the next section.

##### ***5.4.2.1 Species Data***

The Network currently has a full time NPSpecies Data Manager. This position handles the data mining and acquisition of species data for NCBN parks, as well as the following additional data processing steps. For further information see the Northeast Region NPSpecies Data Management Plan (Appendix 10—*Northeast Region Inventory and Monitoring Program NPSpecies Data Management Plan*).

Additional documentation can also be found at:

<http://science.nature.nps.gov/im/apps/npspp/index.htm>

Step 1–Data mining:

- The Network has conducted species data mining intensively since the year 2000. Park generated species data, as well as data collected by outside sources such as local universities, national and regional inventory efforts, nonprofit organizations, individual experts, as well as museum collections have been searched or contacted for information. Three types of species evidence are being compiled and added to NPSpecies; these include: observational datasets (very often these exist only in analog format, but can be found digitally as well), specimen or voucher data (often needs to be manually entered unless a collection is digitally available and data can be harvested), and references (analog data or information that often times must be manually entered).

Step 2–Acquisition:

- NCBN parks compiled all species data that had been housed at the park and submitted it to Network staff for entry into NPSpecies. Annually, a request goes out to all Network park resource managers to provide new species data or information submitted to them in the past year. These data are then entered into NPSpecies.
- Data discovered through the data mining process, which occurs on a park-by-park basis annually, are requested from their original sources, preferably as raw data.
- Data from region-wide species monitoring efforts are acquired through web-accessible data sources, or the program’s data coordinator is contacted and a data request made. Some of the sources for these types of data are the Breeding Bird Surveys, Shorebird monitoring initiatives, and natural heritage program surveys.

Step 3–Conversion:

- All digital data are converted to the NPSpecies format by the Network NPSpecies Database Manager; larger datasets are sent to WASO for conversion.
- Analog data are entered manually into the local MS Access version of NPSpecies or entered directly to the online database if the data contain only a small number of records.

Step 4–Uploading data:

- Once data are converted to the Access version of NPSpecies, they are sent WASO for uploading to the master online NPSpecies database. Once uploaded, a final Access version is downloaded by the Network staff, reviewed, and verified. If data are uploaded directly online, review occurs immediately after entry.

Step 5–Certification (see Appendix 9–*Northeast Region NPSpecies Certifiers Guide*):

## NCBN Information Management Plan

- All species data undergo quality assurance and verification through the NPSpecies certification process. Certification is conducted by taxonomic experts familiar with a particular park's species. The certification is a quality control (QC) step to provide documentation that the organism lists and associated attributes for a park and a taxa category have been reviewed for completeness and accuracy. NPSpecies is a dynamic database, so data processors must indicate when data have been reviewed for completeness and accuracy. This will enable users to qualify the use of the data for scientific, management, and interpretive purposes.

### Step 6–Archive:

- Working NPSpecies databases are backed-up on a regular basis on the University of Rhode Island Field Technical Support Center server.
- Those databases sent to WASO for uploading to the master web database are archived on the national server as well.

#### **5.4.2.2 Bibliographic Data**

NCBN shares a NatureBib Database Management position with three other regional Networks. The regional database manager was hired to mine data for natural resource information and data for all Northeast Region I&M Program parks. These data and information are used to populate the NatureBib bibliographic database system for all Network parks. This database houses all scientific citations and bibliographic information on natural resources in the parks. NatureBib is dynamically linked to the NPSpecies database through common citations. For more information see: <http://www.nature.nps.gov/nrbib/index.htm>. Also, see Appendix 12 (*Northeast Region NatureBib Database Management Plan*) and Appendix 13 (*Northeast Region NatureBib Users Guide*).

### Step 1–Data Mining:

- The regional NatureBib database manager has conducted intensive searches for data and information at all Network parks. Park visits have included searches of each park's library and staff files, staff interviews and searches of local universities, web-based library catalogs, and other agency files.
- Bibliographic references on new natural resource information submitted or collected by the park are requested on an annual basis in order to keep the database as current as possible.

### Step 2–Create NatureBib record:

- Bibliographic citation and abstract harvested or created and converted to NatureBib format
- New citation entered into local Access version of NatureBib

## NCBN Information Management Plan

### Step 3–Upload Data:

- Once the search has been completed and a park’s database reviewed and QA’d, the database is submitted to WASO for upload to the master online NatureBib database.

### Step 4–Archive:

- Working NatureBib Access databases are backed-up on a regular basis at the University of Pennsylvania.
- Databases sent to WASO for upload to the master web database are archived on the national server as well.

### ***5.4.2.3 Specimen Collection Data***

Currently, legacy data are only entered into the NPSpecies database. All specimen information collected as part of the Network’s I&M Program are entered into both NPSpecies and the Automated National Catalog System (ANCS+).

#### Step 1–Data mining:

- Museum and herbaria collections are searched for park-based specimens, both manually and digitally.

#### Step 2–Harvest collection information:

- Park museum or herbaria specimen data cataloged in the (ANCS+) (NPS collection management software) are acquired by the Network and submitted to WASO for entry into NPSpecies.
- On an annual basis, a request goes out to all Network park curators or resource managers to provide new natural resource collection data. New ANCS+ data are added to NPSpecies as they become available.
- Collection data discovered through the data mining process are requested from their original sources.

#### Step 3–Enter data:

- All digital data are converted to NPSpecies. If a large dataset (ANCS+), it is sent to WASO for conversion.
- Analog data are entered manually into the local MS Access version of NPSpecies or entered directly to the online database if the data contain only a small number of records.

Note: The Network maintains a museum specimen database to track where large collections (50+ specimens) from the Network parks are maintained throughout the world.

Step 4–Upload:

- Access NPSpecies files are sent to WASO for uploading to the master online NPSpecies database.
- After uploading, a final Access version will be downloaded by the Network staff and reviewed and verified. If data are uploaded directly online, they are reviewed immediately after entry.

Step 5–Archive:

- Working NPSpecies databases are backed-up on a regular basis on the University of Rhode Island Field Technical Support Center server.
- Databases sent to WASO for upload to the master web database are archived on the national server.

#### ***5.4.2.4 Long-term Monitoring Datasets***

Legacy monitoring projects are those projects that took place for an extended period of time, but were discontinued for one reason or another.

Step 1–Data mining:

- web search
- park staff interviews
- library and NatureBib search

Step 2–Acquire and review dataset:

- Data are downloaded or requested from their original source, preferably as raw digital data.
- The dataset is reviewed for readability and validity. The dataset should be accompanied by enough documentation to determine how data were collected and whether the collection methods were valid. If no documentation is available with the dataset, the usefulness of the dataset will depend upon whether or not the original collector is still available for interview.

Step 3–Convert to MS Access format, Excel and/or a GIS format:

- Once a dataset has been reviewed and accepted, depending upon what format the original dataset was acquired in, it may undergo conversion to either MS Access or Excel for databases, and a GIS format for spatial data.

## NCBN Information Management Plan

### Step 4–Document:

- FGDC compliant metadata and the Biological Data Profile will be created for all datasets used as baseline data for the Network’s I&M program.
- Other datasets to be maintained by the Network will be documented using the Dataset Catalog.

### Step 5–Manipulate:

- Depending on the dataset and its application to the Network program, some data are further summarized and/or analyzed and used for current monitoring projects, park general management plan development and other projects.

### Step 6–Archive:

- Data and documentation are backed-up and archived at the URI FTSC.

## Chapter 6 Quality Assurance, Quality Control

### ***6.1 National Park Service Mandate for Quality***

High quality data and information are mandated by directives and orders, and they are vital to the credibility and success of the I&M program. According to Abby Miller (2001) of the Natural Resource Stewardship and Science Division, “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.”

NPS Director’s Order #11B: “Ensuring Quality of Information Disseminated by the National Park Service,” issued in 2002, 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 and the general public.

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

Order #11B 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.

### ***6.2 Quality Assurance (QA) and Quality Control (QC)***

It is imperative that information and data developed as part of a long term ecological monitoring program be of high quality and adequate for its intended use. To develop quality products, a plan for quality assurance, as well as methods for quality control, must be implemented at all levels of the program. Network staff and cooperators conducting ecological monitoring must understand the need for excellence and the mechanism to achieve excellence at all levels of product development. The Northeast Coastal and Barrier Network’s quality-controlled management system will include the organizational structure, responsibilities, procedures, processes, and resources for implementing QA/QC for its ecological monitoring program.

*Quality Assurance* can be defined as “planned and systematic actions necessary to provide confidence that products will satisfy given requirements for quality.” Scientifically sound ecological data and information is crucial to the Network, and QA must include an integrated

system of project planning, implementation, documentation, assessment, reporting, and quality improvement (Palmer 2003).

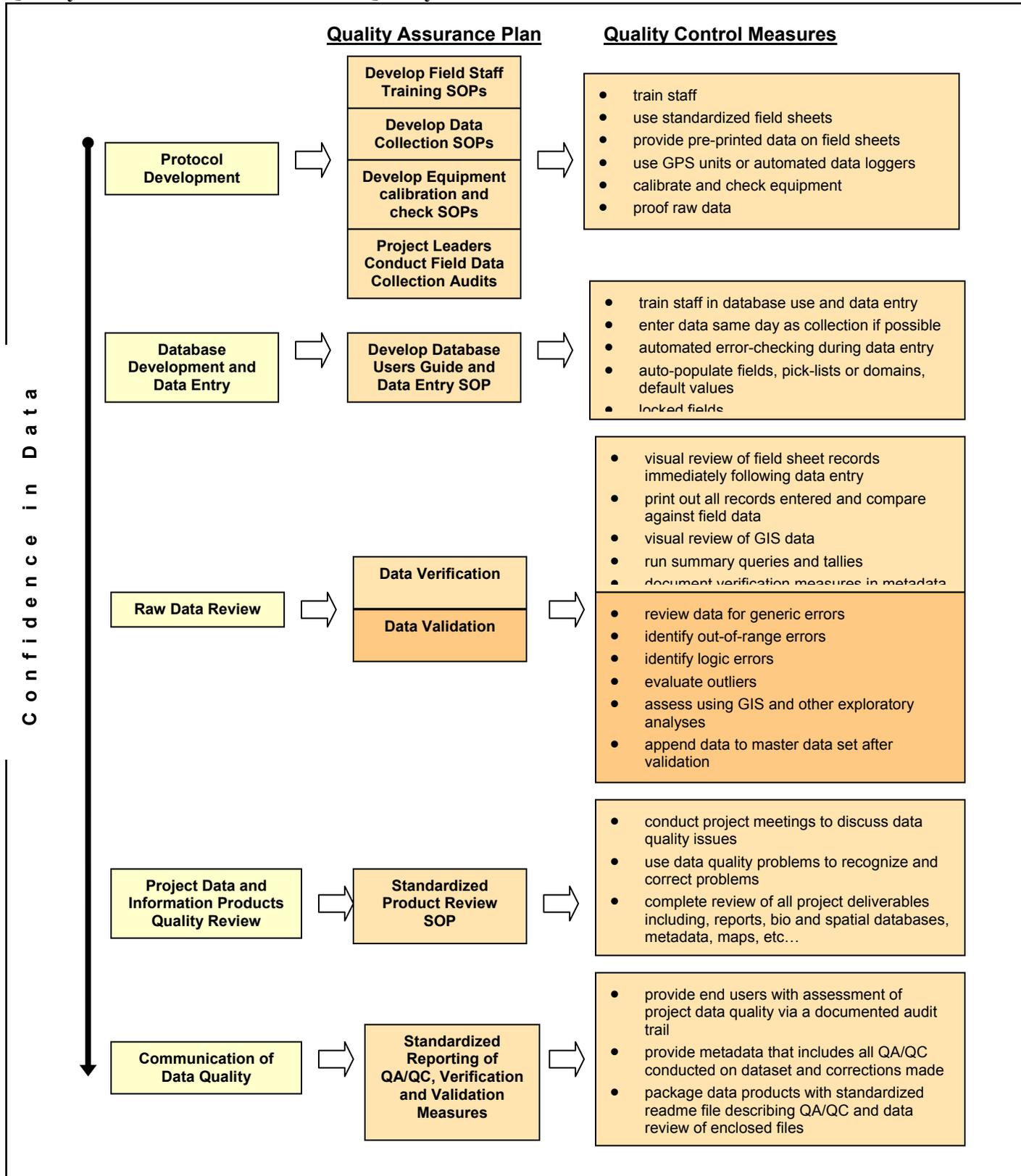
*Quality Control* can be defined as “operational techniques and activities used to fulfill the requirements for quality.” Quality control measures should be integrated into the quality assurance plan. QC provides a set of technical measures that help to maintain the quality of products relative to specified quality assurance standards. Quality Control procedures help us monitor, evaluate, and maintain quality data products and information.

NCBN will establish guidelines for the identification and reduction of errors at all stages in the data lifecycle, including project planning, data collection, data entry, verification and validation, processing, and archiving. This approach requires the Network to:

- develop a plan for quality assurance that will include the identification of roles and responsibilities of Network, park, and cooperative staff for maintaining quality standards at all levels of the program, from field and laboratory data collection to overall data management procedures
- ensure that the processes of achieving quality are not only documented, but maintained through routine review by Network staff
- develop protocols and SOPs to ensure data quality
- evaluate the quality of all data and information based on NPS standards before data are distributed
- perform periodic data audits and quality control checks to monitor and improve the Network’s data quality program

Much QA/QC work involves defining and enforcing standards for electronic formats, locally defined codes, measurement units, and metadata. QA/QC 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 greater confidence in the quality of the data through time. Figure 6.1 illustrates the increase in data quality from project initiation to completion through the use of Quality Assurance and Quality Control measures.

**Figure 6.1 Increase Confidence in Data Quality in Relationship to the Components of a Quality Assurance Plan and related Quality Control Measures**



### **6.2.1 Data Verification and Validation**

Network staff appraise the quality of data by applying verification and validation procedures as part of the overall Quality Assurance process. These procedures will be more successful when preceded by effective QA and QC practices.

*Data verification* determines whether the digitized data match the source data. *Data validation* determines whether the data make sense. Although data entry and data verification can be handled by personnel who are less familiar with the data, validation requires in-depth knowledge about the data. All data must be valid and sensible, and Network staff must not misrepresent the circumstances and limitations of their collection. Failure to follow QA/QC procedures for data entry will render a dataset suspect. All NCBN protocol SOPs pertaining to data management will include guidelines for data verification and validation. Project leaders will ensure that these guidelines are maintained and followed.

After data verification has occurred, all data must be reviewed for illogical or inaccurate entries through the data validation process. Note that validation is a separate operation carried out *after* verification by a project specialist who can identify generic and specific errors in particular data types. Although we may have correctly transcribed the data from original field notes or forms, they might still be inaccurate or illogical. For example, entries of stream pH of 25.0 or a temperature of 95°C in data files raise doubts about their accuracy. Such entries almost certainly are incorrect, whether or not they were properly transcribed from field forms. More often, corrections or deletions of logical or range errors in a dataset require notations in the original paper field records and metadata about how and why the data were changed. 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 dataset since field sheets or other sources of data could easily have been overlooked.

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. Invalid data commonly consist of slightly misspelled species names or site codes, the wrong date, or out-of-range errors in parameters with well defined limits (e.g., 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 as well.

### **6.3 NCBN QA/QC Plan**

The most effective mechanism for ensuring that project data produced is of the right type, quality, and quantity is to provide procedures and guidelines to assist staff in accurate data collection, entry, verification, and validation. NCBN will initiate a comprehensive set of SOPs and data-collecting protocols for quality control, field methodologies, field forms, and data entry

applications with some built-in verification and validation controls. Although specific QA/QC procedures will depend upon the individual vital sign being monitored and must be specified in the more detailed monitoring protocols developed by the Network, some general concepts apply to all Network projects. The general QA/QC plan for Network projects is presented in this chapter, please refer to project specific QA/QC procedures in each Network vital signs monitoring protocol. This chapter of the data management plan provides an overview of QA, QC, Verification and Validation methods and tools currently used by the Network for:

- data collection
- data entry
- data analysis

A Quality Plan for NCBN projects will be developed as a stand-alone document and will include guidance on QA/QC procedures for:

- project development
- project implementation
- product review and integration

#### ***6.4 NCBN QA/QC Plan for Data Collection***

Careful, accurate recording of field observations in the data collection phase of a project helps reduce the incidence of invalid data in the resulting data set. Unlike a typographical error that occurs when a recorded observation is incorrectly transferred from a paper field form to a digital database, an incorrect entry in the field cannot be easily corrected. Therefore, attention to detail during data collection is crucial to overall data quality.

##### ***6.4.1 QA Measures for Data Collection***

The following measures will be administered by the Network to assure quality data collection:

*QA*—All NCBN vital signs monitoring protocols will include a ***data collection SOP***. This SOP or set of 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 and field data recording procedures will be reviewed and approved by the data manager prior to data collection.

*QA*—All NCBN vital signs monitoring 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 equipment calibration and failures that accompany their field data.

*QA*—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, 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. NCBN will develop **training manuals** as part of those 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.

*QA*–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

In addition, the following data collection guidelines will be included in all NCBN protocols and provided to all data collectors:

- Information added to field sheets must be printed and clearly legible.
- 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.
- 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). The 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.

#### **6.4.2 NCBN Data Collection Quality Control Measures**

The following Quality Control techniques will be administered by the Network to assure quality data collection:

*QC*–**Standardized, project-specific field forms** as opposed to field notebooks will be used for all projects. 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.

*QC*–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 where ever possible to ensure proper data are recorded.

**Note:** As an alternative to paper, several options for electronic data collection in the field are now becoming available, including handheld computers, automated data collection instruments, and tape recorders. The Network will be reviewing and testing new data recording devices and may eventually convert applicable protocols from the use of hard-copy field forms to electronic.

*QC*–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 NCBN QA/QC Plan for Data Entry

Transcribing data from field forms to the computer seems a fairly simple task, but without proper preparation and some established guidelines, the quality and integrity of the data can be questionable. Ideally, data entry occurs as soon as possible—immediately after data collection is completed, or as an on-going process during long projects—by a person who is familiar with the data. The primary goal of data entry is *to transcribe the data from paper records into the computer with 100% accuracy*, yet we know that a few transcription errors are unavoidable during data entry. Thus, all data should be checked and corrected during a data verification process.

### 6.5.1 QA Measures for Data Entry

*QA*—All Network monitoring databases will have a complete ***data entry manual***.

*QA*—***Database training*** will be required of all data entry technicians and other users. Network project leaders will make certain that data entry technicians understand how to enter data correctly. Data entry technicians must be familiar with the field data forms and differences in handwriting among field staff. They must also become familiar with the database software, database structure, and any standard codes for data entry used by the Network. At a minimum, data entry technicians must know how to open a data entry form, create a new record, and exit the database properly. They will learn prior to data entry how to commit both a ‘field’ entry and a “complete record” entry and to correct mistakes made while typing.

*QA*—Procedures and schedules for ***entering data in a timely manner*** will be followed for all Network protocols. All data will be entered or downloaded into the project database as soon as possible, preferably on the same day that data are collected.

*QA*—Careful design of database ***data entry forms*** is invaluable to data quality control. The NCBN Monitoring Database Template employs a logical form design to make it easier for data entry personnel to see the data values they are entering, to logically group the data, and to move between entry fields. The Template also makes use of database tools to catch as many data entry errors as possible while the data are being entered.

### 6.5.2 QC Techniques for Data Entry

The following is a list of QC measures developed as part of the NCBN monitoring database template for all monitoring protocols.

- *QC*—***Numeric range limits*** have been set directly on all data fields wherever possible. This is most commonly used for numeric values. For example, the range of valid salinity data for the salt marsh monitoring protocol is 0-75 ppt. If a user enters data that fall outside of this range, an error message will appear and the value will not be stored.
- *QC*—For many data fields, a limited set of values is possible, e.g., park, personnel, or protocol, in these cases, ***text value limits (pick lists)*** have been incorporated into the

database. This restricts the user to only the accepted values. This prevents spelling errors or multiple ways to refer to the same value – such as “Cape Cod”, “CACO” or “Cape Cod National Seashore.” A pick-list will display “CACO”, thereby providing consistency.

- QC–In cases where data in one field is a function of data in other fields, ***auto-filled fields*** are used. Dependent data are automatically entered into the database. This can be for text strings, such as a locationID which is comprised of a park, site and station, or it can be true for numeric fields, such as nekton collection totals which are a function of the number of each species of fish or crustacean that have been collected. Automatically performing the text or numeric calculation reduces the likelihood of data entry errors.
- QC–In many cases it is impossible to have accurate data that are duplications. For example, having more than one simultaneous event using the same protocol is unlikely. The NCBN database template uses ***unique constraints*** to prevent the user from accidentally repeating a record or from entering an incorrect value that duplicates an existing value. If a user enters two events, but accidentally uses the same location for each event, the unique constraint notifies the user of the error.
- QC–***Required fields*** have been established in the database. This means that the user cannot skip data entry for that field. Among other places, this is used in the NCBN database template for events. Before an event record can be saved, it must have a date, location including park, and a protocol.
- QC–***Field data types are standardized*** in the database to keep users from entering approximations for data. The type may be numeric, text, or date. If the data are numeric it can be integer or decimal. This happened commonly in spreadsheets, where the user was unsure of an exact value and would enter “~15” or “between 6 and 12”, when an exact value is needed. If the data value is only known approximately, then the user must coordinate with the principal investigator to determine an appropriate value to enter.
- QC–***Default values*** for fields that have a known value are set in the NCBN database template. One example is the date field data are entered into the database. Obviously the entry date is the current date at the moment that the database opens the field entry date for writing. Entry Date defaults to the current date. The field protocol is also generally known, as only one protocol will be appropriate to a given entry form. The form then enters that protocol as default. In the Salt Marsh Monitoring database, the Nekton data entry form defaults to the Nekton protocol, and the vegetation data entry form default to the vegetation protocol.
- QC–***Automated error checking*** features are used wherever possible in the database. For example, when entering data the user cannot skip the event entry tab. The user must complete the monitoring event before entering any data that was collected in that event.
- QC–***Locked fields*** have been incorporated into the database for some fields once data have been entered. For example, a user might accidentally change the location of an

existing event when he or she meant to select the location for a new event. If the form is displaying a previously entered record, location cannot be edited. If the form is displaying a new record, the location can be selected.

### **6.5.3 NCBN Data Verification Methods**

Data verification immediately follows data entry and involves checking the accuracy of the computerized records against the original source, usually hard copy field records, and identifying and correcting any errors. When the digital data have been verified as accurately reflecting the original field data, Network staff should archive the paper forms. Data are then ready for analysis.

The following methods will be utilized by the NCBN to verify data following data entry.

- *Verification*—All records will be printed from the database upon the completion of data entry and a **visual review and comparison** with the original values from the field forms will be conducted. Errors will be marked and corrected in a timely manner. When one technician performs this review, the method's reliability increases if someone other than the person keying data performs the review. Alternatively, two technicians can perform this review. One technician reads the original data sheets (the reader), and the second reads the same data on the printout (the checker).
- *Verification—Simple summary statistics* will be incorporated into the NCBN database template for simple data verification check. These summary statistics will be protocol dependent, but will help to catch duplicate or omitted entries. For example, by viewing the number of known constant elements, such as the number of sampling sites, plots per site, or dates per sample, these can easily be cross-checked with the field forms.

To minimize transcription errors, the Network's policy is to verify 75% of records to their original source by permanent staff. In addition, 10% of records will be reviewed a second time by the project manager, and the results of that comparison reported with the data. If errors are found during the project manager review, data will be re-verified.

### **6.5.4 NCBN Data Validation Methods**

The following methods will be used to validate all NCBN project data. Specific procedures for data validation depend upon the vital sign being monitored and will be included within each monitoring protocol.

- The use of **numeric range limits** as described in Section 6.4.2 will be used within both inventory and long-term monitoring databases as often as possible. Not only do numeric range limits act as a quality control procedure, but their use help to validate data fields with known value limits. 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 will be important.

- All ***GPS data are projected*** and viewed with the appropriate spatial data set for the park in a GIS in order to validate that all points fall within the planned sampling area, whether it be within the boundary of the park, or specific to a transect or plot within the specified sampling area.
- Other validation methods include the ***detection of outliers***. 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 (with confidence) be considered ‘unusually’ unusual. Each NCBN monitoring protocol will address the use of statistical methods for detecting outliers in reference to the specific protocol. One of the most common tests now used for detecting contamination in samples is Grubbs’ test (Grubbs & Beck 1972; ASTM E 1994). This test will only be used for those NCBN data sets that follow a normal distribution. For others that are clearly non-normally distributed, the Grubbs test will not be used, and other methods for detecting outliers, such as t-tests and F-tests in ANOVA and regression may be considered. (Note: Please see NCBN monitoring protocols for details on detecting outliers specific to each protocol.)

### ***6.6 Monitoring Conformance to Plans and Standards***

Periodic data audits and quality control inspections to maintain and improve the Network’s data quality program will be conducted by project leaders and data management staff. Verification that staff is operating in conformance with the data quality procedures specified in this plan and the protocol specific data management plans will be made. Deficiencies will be tracked and corrected as needed. These quality checks will help to promote a cyclic process of continuous feedback and improvement of both the data and quality planning process.

Periodic checks with Network staff will assure that:

- data collection and reporting requirements are being met
- data collection and reporting procedures are being followed
- verification and validation procedures are being followed
- data file structures and maintenance are clear, accurate and according to plan
- revision control of program documents and field sheets are adequate
- calibration and maintenance procedures are being followed
- seasonal and temporary staff have been trained in data management practice
- metadata collection and construction for the program proceeds in a timely manner
- data are being archived and catalogued appropriately for long term storage will result in a complete quality assessment that will identify existing non-conformities in data handling procedures that can be corrected before years of data collection

### ***6.7 Data Quality Review and Communication***

The National Park Service requires QA/QC review prior to communicating/disseminating data and information. Only data and information that adhere to NPS quality standards will be released.

Director's Order #11B states that all information (e.g., brochures, research and statistical reports, policy and regulatory information, and general reference information) distributed by the NPS (including information obtained from sources outside of the NPS) must be accurate, reliable and timely in nature. 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 documented and accompany all data. Also, mechanisms must be in place for receiving and addressing comments/complaints pertaining to the quality of data.

Data will be distributed to the public through the NCBN I&M web page, national web sites such as the Biodiversity Data Store and the NR-GIS Data Store, and public access databases such as NPSpecies and NatureBib (See Chapter 9 on Data Distribution). Information distributed through any of these mechanisms will undergo Network QA/QC procedures and be approved for release.

#### ***6.7.1 Feedback from QA/QC Measures to Improve Data Quality***

QA/QC procedures should not be performed with the sole objective of eliminating errors, as the results may also prove useful in improving the overall process of producing quality data. For example, if the month and day are repeatedly reversed in a date field, the data entry technicians may require retraining about the month/day entry order. If retraining is unsuccessful in reducing the error's occurrence, the computer program may need to be rewritten so that month and day are entered separately, field length limits are enforced, or a pick list is created. In this manner, the validation process will serve as a means of improving quality as well as controlling the lack of quality.

Field data forms can be modified to avoid common mistakes when necessary. With knowledge of validation errors and exploratory data results in hand, the field data forms as the source of the logic errors can be more easily re-evaluated. Often minor changes, small annotations, or adding check boxes to a field form can remove ambiguity about what to enter on the form. Perhaps surprisingly, when the same type of validation error is found occurring repeatedly in different datasets, the field form—not the field crew—is usually at fault. Repeated errors found during validation can also mean that protocols or field training are at fault, which can then be recognized and corrected.

#### ***6.7.2 NCBN Quality Review of Completed I&M Products***

Another important component of the Network's quality assurance plan is to conduct final review of all project products produced. All data products, including those from inventory, vegetation mapping efforts and monitoring, developed directly from I&M Program funding are to undergo a complete review by trained, technical staff at the North Carolina State University, NPS Field

Technical Support Center (NCSU FTSC). The Network has been working closely with the NCSU FTSC to develop review guidelines and procedures for all project deliverables. See Appendix 23–NCBN Quality Assurance Plan) for a full description of procedures used by FTSC staff to review, QA/QC and document both the Network’s inventory and monitoring data and associated products. These final products are then made available, as well as archived on site at the NCSU FTSC (See Chapter 10 for archiving details).

The following provides a short overview of how NCSU handles the QA/QC and Review process:

*Reports*–All reports are reviewed for proper content and organization. The Table of Contents in all reports is compared to the opened report to ensure that all tables, appendices, and maps are included.

*Spatial Data*–All spatial data are projected in the local UTM zone and a shapefile created. If shape files were submitted as part of the data products these are compared to the newly created file. A number of features are checked including whether or not all of the points fall within the boundary of the park. Survey locations match the survey locations in the raw data. Attribute tables have a key field such as LocationID in order to link to the raw data. See Appendix 26–Northeast Region I&M Program QA/QC Procedures and Metadata Development for Final Biological Inventory Products.

*Raw data*–All raw data are reviewed in reference to the final project report. Comparisons are made such as whether or not all surveys mentioned in the final report are reflected within the data, and the number of survey locations matches those identified in the report.

*Metadata*–If metadata accompanied either the spatial or biological data, these data are checked for compliance with FGDC Compliant Metadata Standards. If no metadata accompanied the data, both FGDC compliant metadata and the Biological Data Profile are created for each project. Data quality are documented in all formal FGDC-compliant metadata for spatial and non-spatial data files posted on the Internet. The Network uses data documentation and metadata to notify end users, project leaders, and Network management of data quality.

Once the review process is complete for all deliverables, all biological databases, spatial data files, reports, and metadata are compiled, zipped into one file along with a descriptive Readme file that provides information on the specific QA/QC procedures applied and the results of the review. These products are then made available on the appropriate NCBN and National NPS websites and repositories, as well as permanently archived at NCSU FTSC.

## ***6.8 Roles and Responsibilities***

If possible, the Network will have a quality assurance plan in place at the time of implementation of Network monitoring projects. This QA plan will be maintained through all project stages to final archiving of the data set. It is critical that each member of the NCBN team (cooperator, Network staff, park staff) work to ensure data quality. Everyone plays a part in producing and maintaining high quality data. Anyone assigned to a project will be responsible for the quality of the results generated from his or her task(s).

The data manager is responsible for overseeing the:

- development of protocols and SOPs to ensure data quality
- adherence by project managers, technicians, etc., of the established QA/QC procedures
- evaluation data quality and information against NPS standards before dissemination outside the Network
- periodic data audits and quality control checks to monitor and improve the data quality program

Project leaders must:

- be aware of quality protocols and convey their importance to technicians and field crews
- ensure compliance with the protocols
- validate data after the verification process is complete
- perform periodic data audits and quality control checks to monitor and improve the data quality program
- review all final reports and information products

Technicians must follow established protocols for data collection, data entry, and verification established in the inventory and monitoring protocol data management SOPs.

### ***6.9 Creating a QA/QC Audit Trail***

The process of subjecting data to QA/QC and then modifying data files leads to the creation of what is called an audit trail (Gotelli 2004). Along with all alterations or reviews of datasets, methods used for QA/QC, verification and validation will be noted in the metadata accompanying each NCBN dataset. Changes made to a dataset (when errors are found) will be noted in the metadata as well. The audit trail created for each project will include the raw data and metadata, the new data file that includes any necessary corrections, the analyzed data and so on. This audit trail which includes these series of files and documentation, will enable future users of NCBN data to recreate the process by which the final, analyzed dataset was created.

### ***6.10 Acknowledgements***

This chapter was adapted from concepts and material developed by Debbie Angell (SODN), Susan Huse (NCBN), and Kristina Callahan (NCSU).

## Chapter 7

### Data Documentation

#### 7.1 Metadata

“*Metadata*” means “data about data.” Metadata describe *how*, *when*, and *by whom* a particular set of data was collected. Metadata also tell how the data are *formatted*. Metadata can include information about the quality, condition, and characteristics of a dataset. Thus, metadata enable us to create and maintain a framework for cataloging datasets, which helps us make datasets more readily available to a broad range of users.

Metadata usually come in the form of documents, which can be printed, provided on line, and stored in various ways. But because metadata can be created and stored in many ways, they can be inconsistent, incomplete, and inaccurate. The Network can prevent many problems by maintaining a consistent framework for data documentation.

Unfortunately, although the Network and the data management community agree about the importance of a consistent framework, individuals, departments, and groups within the NPS and the scientific community take diverse approaches to processing and documenting their data. Thus, this chapter suggests guidelines and sources to consider in developing a complete data documentation strategy. Consistency within the Network and NPS system should be a primary goal of all staff.

#### 7.2 Sources for Data Documentation Guidance

The following federal sources suggest methods for developing a data documentation strategy:

- Executive Order 12906, signed by President William Jefferson Clinton in 1994, mandates federal agencies to “...document all new geospatial data it collects or produces, either directly or indirectly...” using the Federal Geographic Data Committee (FGDC) [Content Standard for Digital Geospatial Metadata](#) (CSDGM). In addition, EO 12906 directs agencies to plan for legacy data documentation and provide metadata and data to the public.
- The FGDC [Biological Data Profile](#) contains all the elements of the CSDGM and includes additional elements for describing biological data sets. Metadata created in compliance with the Biological Data Profile can be added to the [National Biological Information Infrastructure](#) (NBII) Clearinghouse. Although not a requirement, completion of the Biological Data Profile for appropriate data sets is recommended.
- All GIS data layers must be documented with applicable FGDC and NPS metadata standards. The NPS GIS Committee requires all GIS data layers be described with FGDC standards and the [NPS Metadata Profile](#).
- While there are numerous tools available for developing metadata, the [NPS Integrated Metadata System Plan](#) is limited to three recommended desktop

applications: Dataset Catalog, ArcCatalog, and Spatial Metadata Management System (SMMS).

### ***7.3 NPS Integrated Metadata System Plan and Tools***

The NPS Integrated Metadata System Plan is limited to three recommended desktop applications for collecting metadata:

- Dataset Catalog (developed by the I&M Program)
- ArcCatalog (a commercial off the shelf metadata tool)
- SMMS (a commercial off the shelf metadata tool)

An additional tool, not yet recommended by NPS, can be useful for validating metadata records:

- Metadata Parser (and its utility for metadata creation)

This section briefly describes these tools and their potential usefulness for creating metadata.

**Dataset Catalog:** [Dataset Catalog](#) is a tool for cataloging abbreviated metadata on geospatial and biological data sets pertaining to park(s) and/or a network. It provides parks and/or networks a means whereby they can inventory, organize, and maintain information about data set holdings locally. Although Dataset Catalog is not intended to be an exhaustive metadata listing, it does assist parks and networks in beginning to meet the mandates of EO 12906. With the current version of Dataset Catalog (version 2), records can be exported as an FGDC text file, which can then be imported into other metadata tools. Version 2.1 (in development) will include the ability to export records in Extensible Markup Language (XML). The I&M Program recommends that all relevant datasets at I&M parks and networks be cataloged in at least simple Dataset Catalog format. The CUPN-MACA Prototype plan to follow this recommendation.

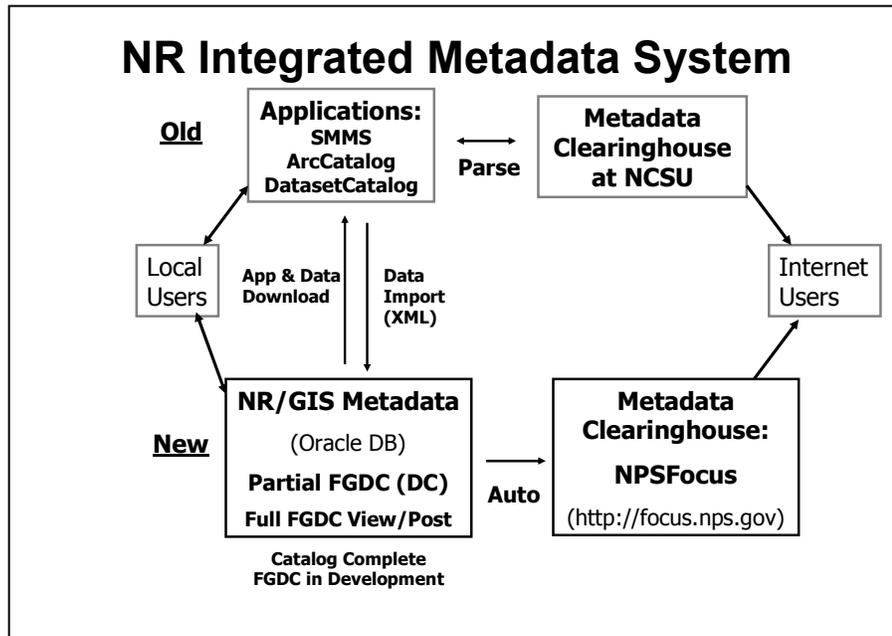
**Spatial Metadata Management System:** [SMMS](#) is a tool with the capability to create, edit, view, and publish metadata that is compliant with FGDC requirements. SMMS uses an MS Access database structure combined with an advanced FGDC-compliant metadata editor. The software allows selection of views depending on whether the user wants the full standard, biological, or the minimal compliant view of Sections 1 and 7. There is online Help to describe the purpose, usage or mandatory status of metadata elements. The context-sensitive help file provides the FGDC definition for each field on the screen. In addition to Help files, there are sample metadata records for most sections that provide "real world" examples. The NPS Integrated Metadata System Plan recommends SMMS for FGDC Biological Profile and other geospatial metadata creation. The CUPN-MACA Prototype GIS staff have been using SMMS for several years. However, this platform may be dropped in the near future as customization of ArcCatalog metadata tools make it the optimal tool.

**ArcCatalog:** [ArcCatalog](#) is a management tool for GIS files contained within the ArcGIS Desktop suite of applications. With ArcCatalog, users can browse, manage, create, and organize tabular and GIS data. ArcCatalog comes with support for several popular metadata standards that allow one to create, edit, and view information about the data. There are editors to enter metadata, a storage schema, and property sheets to view the data. ArcCatalog lets users view GIS data holdings, preview geographic information, view and edit metadata, work with tables, and define the schema structure for GIS data layers. Metadata within ArcCatalog is stored exclusively as Extensible Markup Language (XML) files. The NPS Integrated Metadata System Plan recommends ArcCatalog for gathering GIS-integrated geospatial metadata. An optional but highly recommended extension for ArcCatalog is the [NPS Metadata ArcCatalog Extension](#) developed by NPS Midwest Region GIS Technical Support Center. The extension fixes several ArcGIS 8 metadata errors and provides added functionality for NPS users. Development is also underway to provide Biological Profile editing capability and NPS Profile support. This tool is under review by CUPN-MACA Prototype GIS staff.

**Metadata Parser:** The [MetaParser](#) (mp) program is used to validate metadata records by checking the syntax against the CSDGM and to generate compliant output files for posting to clearinghouses. It generates a textual report indicating errors in the metadata, primarily in the structure, but also in the values of some of the scalar elements where values are restricted by the standard.

Until recently many NPS data stewards collected, parsed and stored metadata (and GIS data sets) in the NPS GIS Clearinghouse managed by North Carolina State University (NCSU). See Figure 7.1. However, efforts are currently underway to unify and streamline metadata development. This new approach utilizes existing desktop metadata creation applications, as well as an online integrated metadata database (NR-GIS Metadata) and a web based data server (NR-GIS Data Server). [NR-GIS Metadata and Data Store](#) will comprise a web based system to integrate both data dissemination and metadata maintenance. It will be possible to update Dataset Catalog records in the NR-GIS Metadata database or in the source desktop application (i.e., ArcCatalog, Dataset Catalog, SMMS). Non-sensitive NR-GIS Metadata records are automatically posted to [NPS Focus](#).

Figure 7.1 NR Integrated Metadata System



#### 7.4 Network Data Documentation Strategy

Data utilized by the Network can be grouped, at least initially, into two broad categories *Network-based data* and *Network-integrated data* (see Chapter 5 for details):

- Network-based data are those data collected by Network staff and/or cooperators.
- Network-integrated data are those data collected by other entities (parks, universities, other agencies, other NPS programs), but identified as important natural resource data, necessary for the Network to manage.

Under network-integrated data there are two data sub-categories, *current or ongoing datasets* and *legacy datasets*. The Network defines current or ongoing datasets as those datasets that are pre-determined for acquisition. These can either be park-based or from external NPS sources. They are protocol dependent datasets necessary for the network to acquire at the time of their collection or soon after. Legacy datasets on the other hand are those data found and compiled through the data mining process. These include vertebrate and vascular plant species data, other important natural resource inventory data, specimen or voucher data, bibliographic data, and existing monitoring datasets.

## Network-based data

1. Metadata Questionnaire completed (see Appendix 37–*NCBN Metadata Questionnaire*.) GPS Metadata Questionnaire has been completed (see Appendix 30–*Northeast Region GPS Metadata Form*.)
2. FGDC compliant metadata developed for all spatial data (see Appendix–*Northeast Region I&M Program Guidelines for developing FGDC compliant metadata for both spatial and biological data*.) Metadata is produced by the network using the Metadata Editor in ESRI ArcGIS 8.x - ArcCatalog. All data is entered into the GUI interface and the metadata is exported as a text file. The text file is parsed using mp.exe (metadata parser) created by the U.S. Geological Survey. When all errors have been resolved, the metadata parser is used to export the final metadata into three formats (txt, xml, and html).
3. Biological Data Profile is developed for all biological databases (see Appendix 5–*Northeast Region Metadata tools used in the creation of the FGDC biological data profile*). The general and spatial information is collected using the Metadata Editor in ESRI ArcGIS 8.x – ArcCatalog. Taxonomic information is retrieved from the online Taxonomic Metadata Tool created by the Integrated Taxonomic Information System. Data dictionary information is retrieved from the Microsoft Access database using the Data Dictionary Harvester Tool in the NPS Dataset Catalog. Notepad is used to combine these three sets of information. Spatial Metadata Management System (SMMS 5.1) is used to complete the metadata. The U.S. Geological Survey Metadata Parser (mp.exe) with the Biological Data Profile (BDP) configuration file is used to parse the metadata and create the final metadata record in three formats (txt, xml, and html).
4. FGDC or Biological Data Profile records are imported into the Dataset Catalog.
5. Metadata distributed through national and network based repositories (See Chapter 9)

## Network-Integrated Data

### *Current or ongoing datasets*

1. These datasets must be acquired from outside sources with some form of metadata or documentation for it to be useable by the Network. If this data is not acquired with fully compliant FGDC metadata, the Network completes the steps above under Network-based data to develop appropriate documentation for these datasets.
2. If these datasets are acquired with FGDC compliant metadata, the metadata record is catalog in the dataset catalog.

3. Metadata will not be distributed on network-integrated datasets.

#### *Legacy datasets*

1. In many cases, legacy data are initially identified as part of data mining efforts. Unfortunately, many of the legacy datasets will be missing pertinent information, and the originator may no longer be in contact. Thus, an “adequate” level of documentation may not be possible. However, the data and all supporting documentation related to it will be assembled and reviewed (see chapter 5 for further information on legacy data acquisition and processing).
2. At a minimum, a dataset catalog record will be created for all legacy datasets based either upon interview with the data collector (if possible), or on any accompanied documentation.
3. Depending upon the extent of documentation, legacy data metadata will or will not be distributed outside of the network. If park-based data it will be left up to the discretion of park staff.

## Chapter 8 Summaries, Analyses, and Reports

Within the Northeast Coastal and Barrier Network, data are distributed and used in many ways. This chapter describes how the Network reports and distributes raw and summarized data for all Network-based Inventory and Monitoring projects. Chapter 4 of this DM Plan describes project work flow. Chapter 5 describes Network data acquisition and processing. Chapter 9 describes how and where the Network will distribute Network-managed data.

### 8.1 Data Users

To better present data to users, NCBN staff must identify potential Network users and the types of information they will need. Table 8.1 shows the two main categories of Network data users: *general* and *primary*.

**Table 8.1 NCBN data product end users and the types of data required.**

NCBN Data Users	Data Used
General users (park superintendents, resource managers, park planners and other park staff, researchers, cooperators, other agencies, the public, educators, the media)	Summary and historical data
Primary users (the scientific community, researchers, cooperators, partners, park staff)	Preliminary, raw, summary and historical data

*(Note: This diagram was adapted from the EMAP Information Management Plan, 2000)*

*General users* represent the broadest spectrum of NCBN data users. General users range from park planners to the interested public. These users have a correspondingly broad range of data and information needs, which range from summary information that might be used for elementary educational purposes to park management planning.

*Primary users* are, typically, users within the scientific community who apply NCBN data to a wide-range of research and analysis activities.

These two types of users require the same high standards of data reporting:

- accessibility to data (web)
- documentation available with data (including both metadata and associated protocols and reports)
- data and documentation files in standard formats
- data made available in a timely fashion

NCBN will support all data users by providing quality assured raw data, summarized data, metadata, and reports. All products that result from the use of Network data will become

available as soon as results have been published and all data management tasks have been completed.

## 8.2 Data Delivery

NCBN will use two stages of data delivery, *data exchange* and *final distribution*. Data exchange involves the sharing of raw data on an ad hoc basis (EPA EMAP IM, 2000). These data will have limited-access and be available by password only on the Network’s webpage. Data exchange will be limited to Network partners, researchers, and park staff. These data will be made available soon after data collection (raw digital data) and prior to final product development.

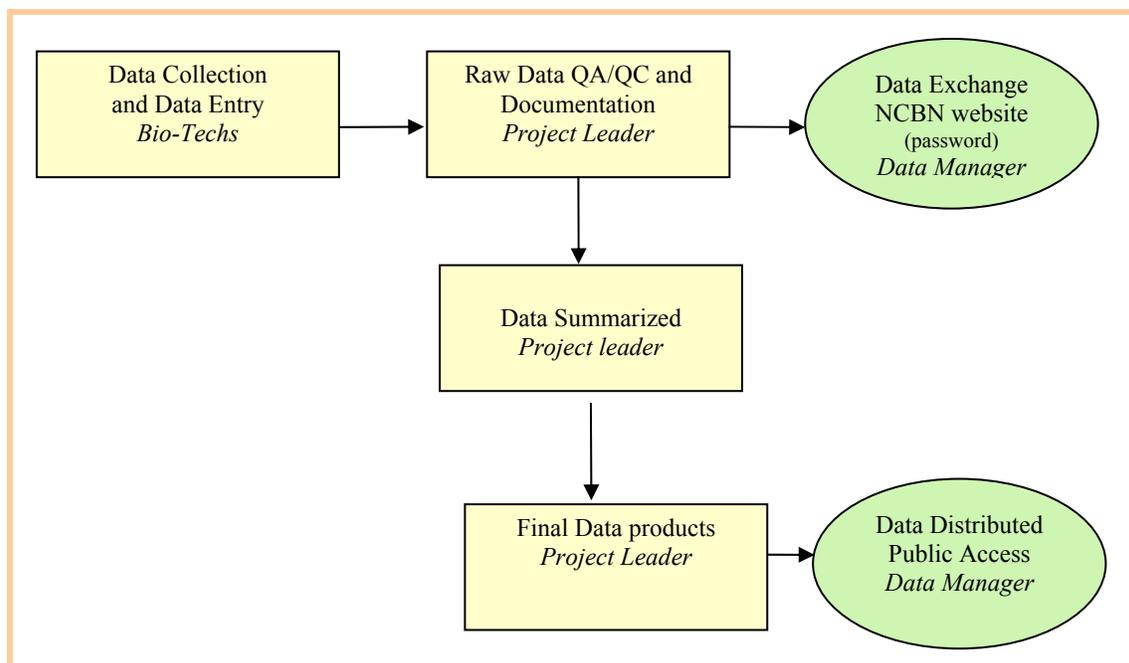
Final distributed data will include all quality-assured, documented data and information products (both documented raw and summarized data). Distribution of these data will be made available and searchable on publicly accessible repositories and websites following publication (See Chapter 9 for details).

## 8.3 Preparation of Data for Exchange and Distribution

All data, including digital raw data, undergo numerous preparations and review before exchange or full distribution. Roles and responsibilities (see Chapter 2) in data preparation follow the basic path shown in Figure 8.1. Data collection and entry will be conducted by Network Biological Technicians (Biotech). Each Biotech will be responsible for entering data, preferably on a weekly basis during the field season, or directly following the field season (within one month). These data will be entered into a “working copy” of that particular project’s monitoring database.

All data will undergo quality control measures as described in chapter 6. Metadata will be created for this “working copy” of the database and will be archived along with the project protocol and database. At that point, these data (raw digital data) will be posted on the NCBN website and will be password protected. Network researchers, cooperators, and park staff will be able to request access to these datasets via the Network Data Manager.

**Figure 8.1 NCBN data preparation for exchange and final distribution**



Data for public access will go through the same process but will require additional processing steps. After data are entered into the “working copy” of the project database, they will be documented and archived. Finally, they will be summarized and incorporated into the project’s annual report (see Chapter 7 of the NCBN Phase 3 Plan for reporting schedules).

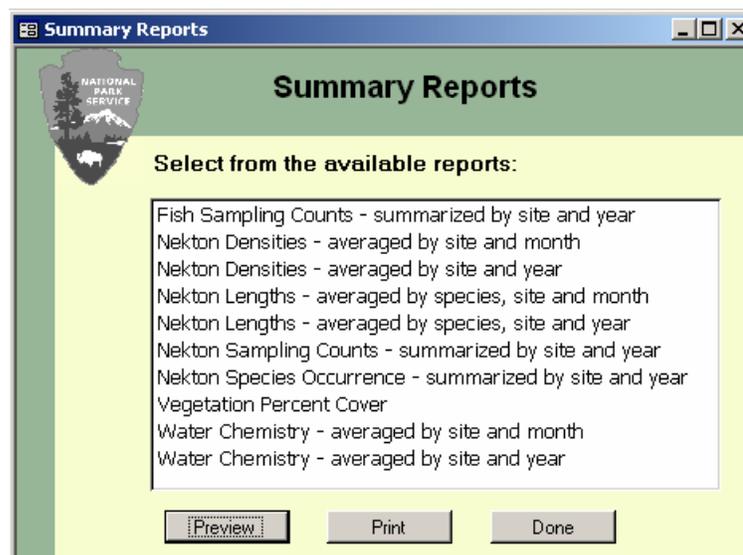
If a trend report is scheduled for that project, data are exported for more robust data analysis. These analyses will also be incorporated into the project’s trend report that replaces the annual report that year. Each NCBN protocol is on a six-year trend report cycle. Final data products are produced, including spatial data products and publications. Once reviewed, documented, and archived, all data products are widely disseminated on publicly accessible repositories and websites (see Chapter 9 of the DM Plan).

#### ***8.4 Monitoring Data Summaries and Analyses***

Specific details on the content of each Network Vital Sign project’s data summaries and analyses should be described in their related protocol (See NCBN Phase 3 Plan and Protocols). All Network Project Leaders (either NPS staff or cooperator) will be responsible for annual summaries and trend analyses for their project. Additional statistical guidance and assistance will be contracted out by the Network as needed.

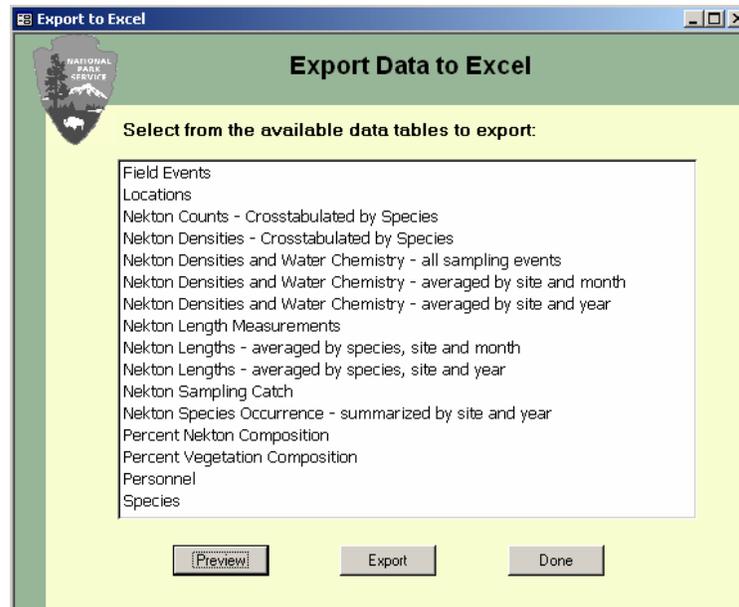
As part of the NCBN Monitoring Database Template, used for all Network monitoring projects, automated data summaries and analyses have been specifically designed to meet the reporting needs of each protocol. Automated reports and data summaries within each vital signs project database help to standardize analyses and reporting procedures across projects. Figure 8.2 shows the Network’s Nekton monitoring database summary reports form. A similar form exists for each project database.

**Figure 8.2 Example of NCBN Monitoring Database Form for selection of automated summary reports for the NCBN Nekton monitoring protocol**



To conduct more complex analyses of Network monitoring data, an “Export to Excel” function has been added to the NCBN monitoring database to facilitate transfer of data to other statistical software packages (Primer, SAS, etc.). To export a digital version of these summary data for direct inclusion in a text document or for use in a spreadsheet or other program, the “Export Data to Excel” from the Analysis and Export menu is available. A list of summary tables is displayed in Figure 8.3. Once selected, a preview of the table or export is available.

**Figure 8.3 Example of NCBN Monitoring Database Form for selection of automated summary report exports for the NCBN Nekton monitoring protocol**



### **8.5 NCBN Monitoring Project Reports**

Two report types will be generated from Network vital signs monitoring projects: *annual* and *trend reports*. These reports will be made available to the public on the NCBN website and will follow the **Northeast Region Natural and Social Science Study Proposal and Deliverable Guidelines** (available at: [www.nps.gov/nero/science/](http://www.nps.gov/nero/science/)). Project reports will be completed by either Network cooperators contracted to implement a Network monitoring protocol or an assigned Project Leader.

#### *Annual Project Reports*

Reports for each NCBN vital signs monitoring project will be completed annually. These reports will include the following:

- summary of monitoring activities for the previous year and significant findings
- annual data summaries
- discussion of problems that may impede monitoring results as well as proposed corrective actions

- discussion of changes made to protocol SOPs

*Trend Reports*

Multi-year, monitoring project reports will be produced on a rotating basis every six years (Table 8.2). These reports will be based upon the project’s long-term trend analyses. These reports are intended for publication in the Regional Technical Report Series. Trend reports will include the following:

- summary of monitoring activities and significant findings
- discussion of patterns and/or trends in the condition of resources being monitored
- interpretation of results in a multi-park, regional or national context
- detailed information on issues or problems of potential management concern
- detailed information on project data management
- discussion of problems that may impede monitoring results as well as proposed corrective actions
- discussion of changes made to protocol SOPs

**Table 8.2 Network Vital Signs Project Trend Report and Protocol Review Schedule** (Note: This table is taken directly from the NCBN Phase 3 Plan (Table 7.1))

Project	Protocol	Year and Report T = Trend Report R = Program and Protocol Review												
		2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	
<b>Geomorphologic Change</b>	Shoreline Position													
	Coastal Topography		T						R	T				
<b>Salt Marsh</b>	Salt Marsh Nekton													
	Salt Marsh Vegetation			T					R		T			
	Salt Marsh Elevation													
<b>Estuarine Eutrophication</b>	Ecosystem Indicators of Eutrophication				T				R			T		
	Estuarine Nutrient Inputs													
<b>Visitor Use and Impacts</b>	Park Use as an Agent of Change					T			R				T	
	Visitor Impacts to Park Resources													
<b>Landscape Change</b>	Landscape Change							T	R					T

**8.6 Species Inventory Summaries, Analyses and Reports**

*Summaries and Analyses*

Inventory PIs determine the types of summaries and analyses to use for Network inventory data. NCBN requires that all analytical methods be discussed in the project proposal and again in the final report. Some standard data summaries are required of all cooperators working on species inventories such as:

- species richness for each sampling site
- species distribution
- number of rare species and their distribution

All inventory data are summarized in GIS for each park as well. Maps are produced displaying species richness by site and distribution.

*Species Inventory Reports*

All inventory project PI’s are required to submit a progress report on an annual basis and a final report at the close of the project. These reports must describe the background and historical information compiled for the taxonomic group of study. They must describe the current inventory results, including species status and/or distribution. All Network inventory reports must also provide a discussion on management concerns pertaining to that resource.

Network cooperative agreements for inventories state that at least one set of draft products, including documentation, data and reports will be submitted prior to product finalization. Draft Network inventory reports go through rigorous review by Network staff and at least one subject-matter expert within the local scientific community. Draft data products are reviewed by Network data management staff. All inventory reports are made available for download on the NCBN website and distributed in hardcopy format to the parks.

**8.7 Other Network-generated Reports**

Table 8.3 provides a list of other reports developed by the Network.

**Table 8.3 Additional Network-generated reports**

<b>Type of Report</b>	<b>Purpose of Report</b>	<b>Primary Audience</b>
Annual Administrative Report and Work Plan (AARWP)	Account for funds and FTEs expended; Describe objectives, tasks, accomplishments, products of the monitoring effort; Improves communication within park, Network, region, Program;	Superintendents, Network staff, regional coordinators, and Servicewide program managers; admin. Report used for annual Report to Congress.
Program and Protocol Review reports	Periodic formal reviews of operations and results (5 year intervals); Review protocol design and products to determine if changes needed; Part of quality assurance – peer review process;	Superintendents, park resource managers, Network staff, Servicewide Program managers, external scientists
Scientific journal articles and book chapters	Document and communicate advances in knowledge; Part of quality assurance – peer	External scientists, park resource managers, Network staff

Symposia, workshops and conferences	review process; Review and summarize information on a specific topic or subject area; Communication of latest findings with peers; Helps identify emerging issues and generate new ideas;	Park resource managers, Network staff, external scientists
Regional Highlights	Interesting and new highlights from Network activities	Park and Regional Staff

*Annual Administrative Report and Work Plan*

On an annual basis the Network develops what is called the *Annual Administrative Report and Work Plan* (AARWP). This report is submitted to the Regional office for review and to the National I&M Program office for review and distribution. NCBN also makes this report available on the Network’s website for download.

An AARWP is required for all NPS I&M Networks. It must contain an accounting of funds and FTE’s expended. It summarizes the objectives, tasks, accomplishments and products of the Network for all projects. Particular sections of this report include Network water quality initiatives and products. Network vital signs are also discussed in relation to GPRA goals in order to assist parks in developing 5 year strategic plans. This report is reviewed annually by the Network’s Board of Directors and provides communication between park and Network staff.

*Articles, Highlights, Workshops*

Other reporting mechanisms used by the Network will include conferences, workshops, symposia, journal articles, and regional highlights. See Table 8.3.

## Chapter 9 Data Ownership

### ***9.1.1 National Park Service Policy on 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.
- 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.*

### **9.1.2 Establishing Data Ownership Guidelines**

The Northeast Coastal and Barrier Network has established guidelines for the ownership of data and other research information. 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 Northeast Coastal and Barrier Network must include OMB Circular A-110 cited under the *Reports and Deliverables* Section of all agreements and contracts. The following shows appropriate language to use when citing Circular A-110:

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

Every cooperative or interagency agreement or contract must include a list of deliverables and products clearly defined within each agreement or contract. Details on formatting and media types that will be required for final submission must be included. Agreements and contracts must list all products expected to result from the project. These include, but are not limited to, field notebooks, photographs (hardcopy and digital), specimens, raw data, and reports.

The following statement must be included in the Reports and Deliverables section of all NCBN agreements and contracts:

“All reports and deliverables must follow the most recent version of the Northeast Region Product Specifications.”

See Appendix 14–*Northeast Region I&M Program Product Specifications*.

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

## **9.2 Data Distribution**

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 NCBN Program will make certain that:

- Data are easily discoverable and obtainable
- Data that have not yet been subjected to full quality control will not be released by the Network, unless necessary in response to a FOIA request
- Distributed data are accompanied by complete metadata that 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

To accomplish this, the NCBN will use a number of distribution methods that will allow information collected and developed as part of the program become widely available to park staff and the public.

### **9.2.1 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:

NCBN Information Management Plan

- **NatureBib**—a master web-based database housing natural resource bibliographic data for I&M Program parks ([NatureBib Home Page](#))
- **NPSpecies**—a master web-based database to store, manage and disseminate scientific information on the biodiversity of all organisms in all National Park units ([NPSpecies Home Page](#))
- **Biodiversity Data Store**—a digital archive of document, GIS dataset and non-GIS dataset files that document the presence/absence, distribution and/or abundance of any taxa in National Park Service units ( [Biodiversity Service Center Home Page](#))
- **Natural Resource and GIS Metadata and Data Store**-online repository for metadata and associated data products. ([NPS NR-GIS Metadata and Data Store Home Page](#)) (**Note:** Currently under development).
- **Northeast Coastal and Barrier Network Website**—provides detailed information about the Network and it’s I&M Program. Metadata on all inventory and monitoring products developed as part of the Network’s I&M plan will be posted to this site. Data and products will either be available through the site, or users will be directed to where the data are stored. ([Northeast Coastal and Barrier Network Home Page](#))

Table 9.1 provides a list of data types that will be uploaded to these sites.

**Table 9.1 Web data types**

<b>Web Application Name</b>	<b>Data types available at site</b>
NPSpecies	Data on Park Biodiversity (species information)
NatureBib	Park Related Scientific Citations
Biodiversity Data Store	The raw or manipulated data and products associated with Inventory and Monitoring data that have been entered into NPSpecies.
NR-GIS Metadata and Data Store	Metadata (spatial and non-spatial and products) and associated I&M products.
NCBN Website	Reports and metadata for all Inventory and Monitoring Data produced by the Network.

Currently, the NR-GIS Metadata and Data Store and the Biodiversity Data Store are under development. Until procedures and further guidance become available for the use of these two repositories, the Northeast Coastal and Barrier Network will disseminate all data developed as part of its I&M Program via the Network website. When both repositories are completely operational, the Network will upload all applicable data and information to each of those sites as needed.

Because Network data will reside in the repositories listed above, this data will automatically be searchable via the integrated metadata and image management system and search gateway called

NPS Focus. This system is being built with Blue Angel Enterprise software for metadata management and the LizardTech Express Server for image management. Currently ten NPS and two non-NPS databases have been integrated into the NPS Focus prototype in either full or testbed form for one stop searching. NPS Focus has been released as an Intranet version only (<http://focus.nps.gov/>), a public version is projected in the near future.

Of the ten databases uploaded to date to NPS Focus, NatureBib and the NR-GIS Metadata and Data Store are most applicable to existing Network bibliographic and spatial metadata. The Network will continue to upload data and information to these two sites, which will coincide with the ability to search for these records through the NPS Focus portal. As NPS Focus reaches further development, other databases and repositories utilized by the Network are expected to be searchable through this portal as well.

### ***9.2.2 Data Classification: Protected vs. Public***

All data and associated 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, etc. Network staff must carefully identify and manage any information that is considered sensitive. The Network must clearly identify and define those data needing access restrictions and those to make public.

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 <http://www.doi.gov/foia/> for further information.

(For more general information, see Appendix 24–*Northeast Coastal and Barrier Network FOIA and Sensitive Data 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 Parks 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)).

## NCBN Information Management Plan

The following guidance for determining whether information should be protected is suggested 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 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
- Management and archival of those records to avoid their unintentional release

Classification of sensitive I&M data will be the responsibility of the Northeast Coastal and Barrier Network staff, the park superintendents, and investigators working on individual projects. Network staff will classify sensitive data on a case by case, project by project, basis. They will work closely with investigators 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 principal investigator(s) working on each project. The investigators, 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.

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
- 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.

### ***9.2.3 Access Restrictions on Sensitive Data***

The Northeast Coastal and Barrier Network Staff are 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 project liaison, or Data Manager identifies all potentially sensitive park resources to the principal investigator for each project. Reciprocally, the principal investigators for each project must identify any known references to potentially sensitive park resources.

For each project, the Network staff provides a complete list of all references to potentially sensitive park resources in each park to the park superintendent for review. Each superintendent then determines which information should be protected.

When preparing or uploading information into any Network database, the Network staff ensures that all protected information is properly identified and marked. The Network staff work together to ensure that all references to protected information are removed or obscured in any reports, publications, maps, or other public forum.

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, (see section 9.2.1) are maintained on both secure and public

servers, and all records that are marked 'sensitive' during uploading will only become available on the secure servers.

Procedures for assigning a sensitivity level to specific records when uploading to the NPSpecies or NatureBib databases are discussed in the following management plans:

- *Appendix 10–Northeast Region Inventory and Monitoring Program NPSpecies Data Management Plan*
- *Appendix 12– Northeast Region I&M Program NatureBib Data Management Plan*

Also, see the following websites:

- <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. However, limits to how these data are subsequently released must also be clearly defined. It is crucial that the Network staff institute quality control and quality assurance measures to ensure that the person doing the uploading of records into the online applications is familiar with the procedures for identifying and entering protected information.

#### ***9.2.4 Public Access to Network Inventory and Monitoring Data***

According to FOIA (specifically the 1996 amendments), all information routinely requested must be made available to the public via reading rooms and/or the internet. Network project data will be available to the public at one or more internet locations:

- The Northeast Coastal and Barrier Network web site
- Public servers for the NPSpecies and NatureBib databases
- Public server for the Biodiversity Data Store
- Public server for the NR/GIS Data Store

The Network will regularly provide updated information about inventories and monitoring projects, including annual reports and detailed project reports through the Network web site. Information on species in the National Parks, including all records generated through the Inventory and Monitoring Program, will be maintained and assessable through the NPSpecies database. Bibliographic references that refer to National Park System natural resources will be accessible through the NatureBib database. Documents, maps, and data sets containing resource information from all sources, and their associated metadata, will be assessable through the Biodiversity Data Store and/or NR/GIS Data Store. Each of these databases/repositories will be available via both a secure server and a public server, and the public can access all information in these databases except those records marked as 'sensitive.'

### ***9.2.5 Data Availability***

Both raw and manipulated data resulting from the Network's inventory and monitoring projects will be fully documented with FGDC compliant metadata and made available to the public via the Network's website. The metadata for all datasets will be made accessible to the public as soon as they are provided and verified by the Investigator(s) or project managers.

Datasets for short-term studies (inventories) will be provided to the public on the NCBN website, two years following the year the data were collected or following publication of the Investigator's results (whichever comes first). Long-term (monitoring) studies will be provided to the public in four or five year intervals, or when trends analyses have been completed and reported on by the Network. (This will be specific to each Network monitoring protocol, refer to the Network's Phase 3 plan for further information). Before data are posted, the Investigator or project leader will be asked to verify the final dataset and metadata if necessary. After the Network staff and Investigator verify the dataset, the data will be made accessible to the public, provided no sensitive information has been identified.

NCBN staff will notify investigators prior to making datasets available to the public. This will allow each investigator the opportunity to request in writing to further restrict access to the dataset by the public. Network staff will review the Investigator's request and determine whether the request will be granted and for how long the dataset will remain restricted.

### ***9.2.6 Data Acquisition Policy***

The NCBN will develop a dataset acquisition policy that will be made available to all NCBN website users who wish to acquire program data and information. This policy will include such things as:

- Mandatory questionnaire will be available on the website (This questionnaire must be completed and mailed to the Network Data manager before data can be acquired. This questionnaire will allow Network staff the ability to maintain a distribution log specifying recipient name and contact information, intended use of data, export file format, delivery date and method, and data content description noting range by date and geography of data delivered.)
- Statement about use and appropriate citation of data in resulting publications
- Request that acknowledgement be given to the National Park Service Inventory and Monitoring Program within all resulting reports and publications

All data sets with public access available on the NCBN website will be accompanied by the Network's acquisition policy.

### ***9.3 Data Feedback Mechanisms***

The NCBN website will provide an opportunity for NPS staff, cooperators and the public to provide feedback on data and information gathered as part of the Network's I&M Program. A "comments and questions" link will be provided on the main page of the site for general questions and comments about the Network's program and projects. A more specific "data error feedback" link will direct comments to the NCBN staff pertaining to errors found in website accessible data. Annual reporting of progress will be presented to the Board of Directors and to the Technical Committee on a yearly basis, and feedback will be expected during and following these presentations.

#### ***9.3.1 Data Error Feedback Procedures***

The following feedback procedures describe the process which NCBN will use to receive and verify data errors identified by public and private data users:

- Web Users send in a notification about an alleged error through the NCBN website. NCBN staff then sends an acknowledgment to the notifier.
- NCBN staff then inputs the information into a data error log table incorporated in either each of the Network monitoring databases or a specific error tracking database developed for the Network.
- NCBN then determines if the data questioned by the notifier are correct or incorrect. If the data are correct, then the Network staff informs the notifier that no corrections are to be made and the information stands. If the data are incorrect, the Network staff makes the appropriate corrections and notifies the original data collectors (cooperator, other agency, park staff, etc...).
- Once data are corrected, the Network website will be refreshed with the corrected information.
- Throughout this process, the Network staff will continually inform the notifier via e-mail of the status.

### ***9.4 Acknowledgements***

Thanks to Marc Albert (NCBN) for developing the FOIA and sensitive data appendix as well as Wendy Schumacher (WASO) and Debbie Angell (SODN) for their review and assistance.

## **Chapter 10**

### **Data Maintenance, Storage, and Archiving**

Long-term maintenance and management of digital information are vital to the Network's 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 (Olson and McCord 1998). 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.

NCBN has entered into a cooperative agreement with the North Carolina State University NPS Field Technical Support Center (NCSU FTSC) for the maintenance, storage, and archiving of datasets produced as part of the Network's inventory, vegetation mapping, and vital signs monitoring program. The Network is also negotiating a new agreement with the University of Rhode Island FTSC for the maintenance and storage of digital data developed by Network and Regional I&M staff stationed at the University of Rhode Island. These files include the Network's administrative and working data files.

#### ***10.1 Digital Data Maintenance***

Digital data maintained in the NCBN data archive at NCSU FTSC 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.

Following the lead of the National Park Service and the National Inventory & Monitoring program, NCBN has adopted MS-Access as its database standard and ArcGIS as its spatial data management standard. NCBN 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 (Vogt-O'Connor 2000). 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, NCBN archiving procedures include saving datasets in both their native format (typically MS-Access 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 Datasets***

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 an MS-Access database or Excel spreadsheets; spatial data such as shapefiles, coverages, geo-referenced images, etc.; and a final 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
- 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 NCBN. 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 Monitoring of Datasets***

Data produced as part of the Network's long-term monitoring projects will have variable archiving requirements that depend upon the data types collected (specifically the Network's Geomorphologic Change Protocol—see associated protocols for specific details). 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 (data "milestones") 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 in MS-Access or Excel spreadsheet format as well as spatial data depending on 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.

Table 10.1 describes the fields required by Network staff when submitting digital biological datasets or products to NCSU FTSC for archiving. This simple tracking system will be made available on the web for all Network staff.

**Table 10.1 Digital Data Tracking Database**

<b>Field Name</b>	<b>Description</b>	<b>Format</b>
Project	Project title	Text
Subject	Standardized subject name (e.g. birds, small mammals, vegetation map)	Text
Park	4-letter acronym(s) for park name(s)	Text
Transmitter	Name, address, phone number of person who sent the data	Text
Date Received	Date digital data were received	Date
Transmittal Doc	Name of transmittal documentation data file and storage location	Text
Media Type/Count	Type and number of media received (e.g. 3 CDs, 1 DVD, 1 Firewire drive)	Text
Data File Count	Number of data files received	Numeric
Data File Names	The name of each data file and its associated metadata record or data dictionary	Text
Server	The name of the server on which the data are stored	Text
Distributed To	Names of people to whom copies of the data have been distributed; for each, date copy sent	Text

### ***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 ArcGIS shapefiles, coverages and geo-databases; 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 ArcGIS formats (shapefiles, coverages, and geo-databases) will be maintained in formats that are accessible by the current ArcGIS version. ArcGIS has maintained compatibility with previous data formats, and while shapefiles have retained all functionality in ArcGIS, coverages may require conversion to a newer ArcGIS format if they are no longer supported. If necessary, NCSU FTSC is responsible for converting ArcGIS data to an upgraded format.

In addition to archiving ArcGIS 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 GPS data (e.g., .COR files) will be archived, and the ability to correct GPS data over time will inevitably be lost. Thus, these data will be archived in their native format and in ASCII comma delimited text format.

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 Storage and Archiving Procedures—Digital Data***

The archiving procedures described below are designed to ensure both the security of and ready access to NCBN 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 NCBN.

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:

1. name and address of the sender
2. brief project description, including topic and park name(s)
3. type and number of digital media (e.g., 3 CDs, 1 DVD)
4. 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)

## NCBN Information Management Plan

For digital aerial photography:

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

Upon receipt, NCSU FTSC staff will:

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

NCBN 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 an Excel spreadsheet, referred to as the Digital Data Tracking Database (see Table 10.1). Information about sets of digital aerial photography should be entered into a similar spreadsheet, the Aerial Imagery Tracking Database (see Table 10.2).

Copies of the Digital Data Tracking Database (Table 10.1) and the Aerial Imagery Tracking Database (Table 10.2) will be sent to the Network once a year or as requested. 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.

**Table 10.2 Aerial Imagery Tracking Database**

Field Name	Description	Format
Park	4-letter acronym for park name	Text
Name	Name, address, phone number of person who sent the photography	Text
Date Received	Date digital photos were received	Date
Media Type/Count	Type and number of media received (e.g. 3 CDs, 1 DVD, 1 firewire drive)	Text
Image Files	Number of image files received	Numeric
Date Flown	Date photography was flown	Date
Season	Season of photography (“leaf-on” or “leaf off”)	Text
Flightlines	Number of flight lines	Numeric
Photocenters/Airborne GPS	Name(s) of data file(s) containing photocenters or airborne GPS data	Text
Contractor	Name and address of the contractor who produced the aerial photography	Text
Analog Photos Available	Whether or not analog air photos are available (“Yes” or “No”)	Text
Digital Mosaic Available	Whether or not a digital mosaic of the photography is available (“Yes” or “No”); If yes, name of digital mosaic data file and storage location	Text
Distributed To	Names of people to whom copies of the data have been distributed; for each, date copy sent	Text
Notes	Notes regarding missing image files, problems with particular images, etc.	Text

### ***10.2.1 Storage Facilities for Electronic Archives***

Currently, NCBN data are archived at NCSU on servers maintained by the Center for Earth Observation. 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.

Beginning in January 2005, the NCBN data archive will be moved to the NC State enterprise storage system, which will meet State of North Carolina certified critical data standards. Servers will be located in a secure facility and data backups will be performed weekly.

### ***10.2.2 Directory Structure for Electronic Archives***

The NCBN 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:

- administrative documents such as agreements, correspondence, and research permits
- programmatic documents including protocols, procedures, and supporting documents
- interim datasets or milestones
- datasets submitted by contractors
- datasets reformatted or manipulated by NCBN (e.g., data converted to NRDT format, datasets migrated to current software formats)
- datasets—in ASCII format
- conceptual or statistical models used for data interpretation
- final reports
- README files—including an explanation of directory contents, project metadata (including a dataset catalog report), and version documentation

After final data and reports have been submitted, draft products will not need to be maintained.

### ***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.

Data stored on NCSU ITD servers are backed up weekly and backups are retained for 28 days off site. A working Disaster Recovery Plan/Business Continuity Plan has been established in the event of a disaster.

Regional and NCBN data and information produced by staff stationed at URI are backed up by the URI FTSC on a Quantum Super DLT Tape Drive: SDLT 320 (TR523AA-EY). Data are backed up at URI on a daily and weekly basis. Data are stored both on-site and two off-site secure locations.

#### ***10.2.4 Data and Network Security***

Beginning January 2005, NCBN data archives maintained by NCSU will be 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 NCBN data on the NCSU ITD servers is restricted to a limited number of designated, authenticated employees.

#### ***10.3 NCBN Storage and Archiving of Official and Non-official Records***

The guidelines in this section apply to both official and non-official records. *Official records* are the original records created or received in performing the daily work of the Network. By law, the National Archive and Records Administration (NARA) has responsibility for the official records of the federal government. NPS official records are managed according to the *Records Management Guideline*, NPS-19. For NCBN, the Northeast Region Boston Support Office (NER BOSO) records officer provides guidance. Official records most applicable to the Network's daily work include budgets, personnel records, central correspondence, and contracts. The official copies of these are maintained by the NER BOSO office.

*Non-official records* are those records produced by the Network including all natural resource records such as field notes, reports, maps, etc.. Directions for managing these materials and other digital materials are provided in NPS Director's Order 19: Records Management (2001) and its appendix—NPS Records Disposition Schedule (NPS-19 Appendix B, revised 5-2003). NPS-19 states that all records of natural and cultural resources and their management are considered mission-critical records. NPS-19 further states:

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

Section N of NPS-19 Appendix B provides guidelines on natural resource-related records (including, specifically, the results of Inventory and Monitoring Programs). Section N indicates that all natural resource records are considered "permanent;" that is, they are to be transferred to the National Archives when 30 years old. It also indicates that non-archival copies of natural resource-related materials are "potentially important for the ongoing management of NPS resources" and should not, in any instance, be destroyed.

### ***10.3.1 Hard Copy and Digital Reports Produced by the Network***

All paper and digital reports managed or produced by the NCBN will be housed in one of three locations:

1. NCBN central files, University of Rhode Island, Kingston, RI/Northeast Region, Boston Support Office, Boston, MA
2. Technical Information Center (TIC)
3. Network Parks' Central Files and/or Museums

1. NCBN central files, University of Rhode Island, Kingston, RI/Northeast Region, Boston Support Office, Boston, MA

These files contain project files, copies of administrative documents, and non-record copies of documents. Examples include: meeting minutes, correspondence, memoranda of understanding, contracts and agreements, research permits, and interim and selected final reports produced by the program or under its auspices. NCBN will use acid-free paper and folders for all permanent records in the central files. In addition to maintaining these paper records, NCBN will maintain electronic versions when possible. These will be housed and backed-up on a regular schedule on the URI Field Technical Support Center (URI FTSC) server. The central files will be maintained by a URI FTSC staff member under the guidance of the Network data manager and coordinator.

2. Technical Information Center (TIC)

TIC provides temperature and humidity-controlled facilities, a professional archival staff, and meets all museum standards set by NPS. This repository will be used for technical reports by the Network a high priority to maintain under archival conditions.

3. Network Parks' Central Files and/or Museums

NCBN will provide high-quality copies of park-related documents resulting from NCBN projects, along with electronic versions, to park resource management staff. Parks may choose to accession these materials into their museums, incorporate them into their central files, or house them in their resource management library, as they deem appropriate. NCBN will not manage documents at the park level.

### ***10.4 Specimens***

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

### ***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 (Cox 2000).

NCBN recommends that staff and contractors provide photos to the Network as 35mm slides (preferably Kodachrome or Ektachrome), which have a proven long-term stability (Wilhelm and Brower 1993). If contractors cannot provide slides, NCBN requests digital photos in their original size and format accompanied with documentation. See Appendix 14—*Northeast Region I&M Program Product Specifications*.

Slides should be labeled using indelible pigment ink or laser-printed archival-quality slide labels. Slide labels should include: a unique ID, project name, photographer, photo date, a brief identification of contents (e.g., species name, plot ID), and geographic location (UTMs or description). All slides will be stored in polypropylene slide sleeves at the NCBN central office until an appropriate archival center has been identified by the Network. In addition, all slides will be scanned and saved as TIFF files, and these electronic copies are used as the primary means of distributing or reproducing the images.

If photographs are provided, they will be stored in individual polypropylene sleeves and within archival boxes. Each photo will be labeled on the back, using archival-quality labels that are either laser-printed or hand-labeled, with the same elements of information required for slides. If a contractor submits photographs, corresponding TIFF files must also be submitted

Every image, regardless of format, must have an entry into the NCBN Photo Database (Appendix to be developed), where attributes such as electronic file name, keywords, project, photo description, photographer, date, and location will be catalogued. All photo files and the associated photo database are housed on the archive portion of the URI FTSC server.

#### ***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:

1. The name and address of the sender
2. The park name(s)
3. The number of photos transmitted (for each park)

Then NCSU staff will proceed as follows:

1. 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.

2. NCSU enters information about the set of aerial photos (as appropriate and as it becomes known) into an Excel spreadsheet, referred to as the Air Photo Tracking Database (see Table 10.3).
3. NCSU sends a copy of the Air Photo Tracking Database (Table 10.3) to the Regional I&M Coordinator and to the Network Coordinator every January. In future, this database will be posted on the World Wide Web.
4. Aerial photographs are then stored in plastic sleeves, lying flat, in a fireproof cabinet, in a dark, cool, temperature-controlled environment.

### ***10.6 Role of Curators in Storage and Archiving Procedures***

NCBN Park curators are an ongoing source of expertise, advice, and guidance on archiving and curatorial issues, and they have a role in almost every project undertaken by the Network. Project managers should involve park curators when projects are in the planning stage, to ensure that all aspects of specimen preservation or document archiving will be considered, and that any associated expenses are included in project budgets.

**Table 10.3 Air Photo Tracking Database**

Field Name	Description	Format
Park	4-letter acronym for park name	Text
Name	Name, address, phone number of person who sent photos	Text
Date Received	Date photos were received	Date
Date Flown	Date photography was flown	Date
Season	Season of photography (“leaf-on” or “leaf off”)	Text
Flightlines	Number of flight lines	Numeric
Photos	Number of photos received	Numeric
Photocenters/Airborne GPS	Name(s) of data file(s) containing photocenters or airborne GPS data	Text
Contractor	Name and address of the contractor who produced the aerial photography	Text
Digital Imagery Available	Whether or not the photos are available in digital form (“Yes” or “No”); If Yes, name(s) of digital file(s) and storage location	Text
Digital Mosaic Available	Whether or not a digital mosaic of the photos is available (“Yes” or “No”); If yes, name of digital mosaic data file and storage location	Text
Checked Out	Name of person to whom photos are checked out and expected return date	Text
Notes	Notes regarding missing photos, problems with particular photos, etc.	Text

**10.7 Acknowledgements**

This chapter was originally adapted from material written by Margaret Beer (NCPN). Additional material was provided by staff at the NCSU FTSC.

## References

- Archaeological Resources Protection Act of 1979 (ARPA). 16 USC 470aa-470mm; PL 96-95
- Brunt, James W. 2000. Data management principles, implementation and administration. Pp. 25-47 in Michener, William K., and James W. Brunt (eds), 2000, *Ecological Data: Design, Management and Processing*, Methods in Ecology Series, Malden, MA: Blackwell Science Ltd.
- Chapal, Scott E., and Don Edwards. 1994. Automated smoothing techniques for visualization and quality control of long-term environmental data. Pp. 141-158 in Michener, William K., James W. Brunt, and Susan G. Stafford (eds), 1994, *Environmental Information Management and Analysis: Ecosystem to Global Scales*. London: Taylor & Francis Ltd.
- Cox, R. J. 2000. Searching for authority: archivists and electronic records in the New World at the fin-de-siècle. *First Monday* 5:1. [http://firstmonday.org/issues/issue5\\_1/cox/index.html](http://firstmonday.org/issues/issue5_1/cox/index.html)
- 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 Science Ltd.
- Federal Cave Resources Protection Act of 1998 (FCRPA). 16 USC 4301-4310; PL 100-691
- Freedom of Information Act (FOIA). 5 USC 552; PL 89-554, 90-23
- Gotelli, Nicholas J. and Aaron M. Ellison. 2004. *A Primer of Ecological Statistics*. Sinauer Associates, Inc., Sunderland, MA.
- Gregson, Joe. 2000. *Natural Resources Data and Information Systems Handbook Outline*. NPS I&M Program. Unpublished.
- Gregson, Joe. 2002. *Draft Natural Resource Data and Information Systems Management Handbook Chapter 2*. NPS I&M Program.  
<http://www1.nrintra.nps.gov/im/datamgmt/nrdismhb.htm>
- Grubbs, F.E., and G. Beck. 1972. Extension of sample sizes and percentage points for significance tests of outlying observations. *Technometrics* 14:847-854.
- 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.
- Michener, William K. and James W. Brunt. 2000. *Ecological Data: Design, Management, and Processing*. Blackwell Scientific, Ltd., London

## NCBN Information Management Plan

- Michener, W. K., J. W. Brunt, J. J. Helly, T. B. Kirchner, and S. G. Stafford. 1997. Nongeospatial metadata for the ecological sciences. *Ecological Applications* 7:1, 330-342.
- National Historic Preservation Act (NHPA). 16 USC 470-470x-6; PL 89-665, 96-515
- National Parks Omnibus Management Act of 1998. 16 USC 5901-6011; PL 105-391
- National Park Service. 2004. Natural Resource Inventory and Monitoring Data Management Vision and Framework Draft.  
[http://science.nature.nps.gov/im/datamgmt/docs/DMVision\\_Framework.doc](http://science.nature.nps.gov/im/datamgmt/docs/DMVision_Framework.doc)
- National Park Service Management Policies 2001.
- National Park Service. 2002. Director's Order #11B: Ensuring quality of information disseminated by the National Park Service.
- NPS Director's Order #66: Freedom of Information Act and the Protection of Exempted Information (Drafts 12-4-03 and 4-12-04).
- Olson, R.J. and R.A. McCord. 1998. Data Archival. In: Michener, W.K., J.H. Porter, and S.G. Stafford. Data and information management in the ecological sciences: a resource guide. LTER Network Office, University of New Mexico, Albuquerque, NM. pp 53-57.
- 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.
- Palmer, Craig J., and Eric. B. Landis. 2002 Draft. Lake Mead National Recreational Area Resource Management Division: Quality System Management Plan for Environmental Data Collection Projects.
- Tessler, Steve, and Joe Gregson. 1997. Draft Data Management Protocol. NPS I&M Program.  
<http://www1.nrintra.nps.gov/im/dmproto/joe40001.htm>
- U.S. Environmental Protection Agency. 2001. National Coastal Assessment Quality Assurance Project Plan 2001-2004. United States Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Gulf Ecology Division, Gulf Breeze, FL. EPA/620/r-01/002.
- U.S. Environmental Protection Agency. 2003. Guidance for geospatial data quality assurance project plans. EPA QA/G-5G. EPA/240/R-03/003. Washington, DC: Office of Environmental Information.

## NCBN Information Management Plan

- U.S. Environmental Protection Agency. 1999. EMAP Information Management Plan 1998-2000. United States Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Gulf Ecology Division, Gulf Breeze, FL. EPA/620/R-99/001a.
- USDI National Park Service. 2003. NPS records disposition schedule. NPS-19, Appendix B – Revised, 5-03. 61 p.
- USDI National Park Service. 1999. Data management protocol. Denali National Park and Preserve, Denali Park, AK. 34 p.
- Vogt-O'Connor, D. 1996. Care of archival compact disks. National Park Service Conserve O Gram 19:19. NPS Museum Management Program, Washington, DC. 4 p.
- Vogt-O'Connor, D. 1997. Caring for photographs: general guidelines. National Park Service Conserve O Gram 14:4. NPS Museum Management Program, Washington, DC. 4 p.
- Vogt-O'Connor, D. 2000. Planning digital projects for preservation and access. National Park Service Conserve O Gram 14:4. NPS Museum Management Program, Washington, DC. 4 p.
- Wilhelm, H. and C. Brower. 1993. The permanence and care of color photographs: traditional and digital color prints, color negatives, slides, and motion pictures. Preservation Publishing Company, Grinnel, IA. 744 p.