DATA MANAGEMENT PLAN

Appalachian Highlands Inventory and Monitoring Network

December 2004
Revised 2007, 2014

Appalachian Highlands Inventory and Monitoring Network
National Park Service
Blue Ridge Parkway
67 Ranger Drive
Asheville, NC 28803

ACKNOWLEDGMENTS: The APHN staff would like to thank many people for their contributions to our data management program and the writing of this document, especially Debbie Angel, Dorothy Mortenson, Margaret Beer, John Boetch, Christina Wright , and the members of the Data Management Planning Working Group, who contributed to the chapters listed below:

Name	**Network**	**Chapter**
John Boetsch	**NCCN**	**II, III, and IV**
Rob Daley	**GRYN**	**III**
Patrick Flaherty	**APHN**	**IV**
Geoff Sanders	**NCRN**	**V**
Debbie Angell	**SODN**	**VI**
Sara Stevens	**NCBN**	**IX**
Margaret Beer	**NCPN**	**X**
Karen Oakley	**USGS**	**VIII**

Table of Contents

Chapters

Figures

Tables

Standard Operating Procedures (SOPs)

Appendices

1. Introduction

1.1 I&M Program Overview

The National Park Service (NPS) Inventory and Monitoring Program represents a long-term commitment by the Park Service to document the status of natural resources in parks, and the long-term trends in their condition. The National Parks Omnibus Management Act (1998) established a mandate for the NPS to fully integrate natural resources inventories, 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 mandate, the NPS initiated a permanently funded, natural resources inventory and monitoring (I&M) program in the year 2000. In order to maximize efficiency, the 270 Park Service units with significant natural resources were organized into 32 ecologically similar geographic units, called I&M networks, each of which share personnel and resources. Funding for the 32 I&M networks was phased in over several years, with the first 12 networks receiving operating funds beginning in October, 2000. The Appalachian Highlands I&M Network (APHN) was included in this first group.

1.2 APHN Organization and Management

The Appalachian Highlands I&M Network includes four National Park Service units in North Carolina, Virginia, Tennessee, and Kentucky: Big South Fork National River and Recreation Area (BISO), Blue Ridge Parkway (BLRI), Great Smoky Mountains National Park (GRSM), Obed Wild and Scenic River (OBRI) (Figure 1.1). As a "prototype" I&M park, GRSM provides input to the Network concerning protocol development and sampling design, however, the Great Smoky Mountain National Park I&M program is operationally distinct from the rest of the Network.
The APHN charter, created in 2001, describes the process used to plan, manage, and evaluate the inventory and monitoring program within the Network. Significant management and budgeting decisions are made by a Network Board of Directors, comprised of the Superintendents of the Network parks, together with the regional and Network I&M coordinators. A Science and Technical Committee, which includes Network and park resource management staff, provides technical assistance and advice to the Board of Directors. The NPS Southeast Region provides program quality assurance, oversight and other technical assistance, as requested from the Board of Directors. This management structure is designed to foster the development of an I&M program which is

responsive to the unique set of long-term resource issues and threats within each Network park.

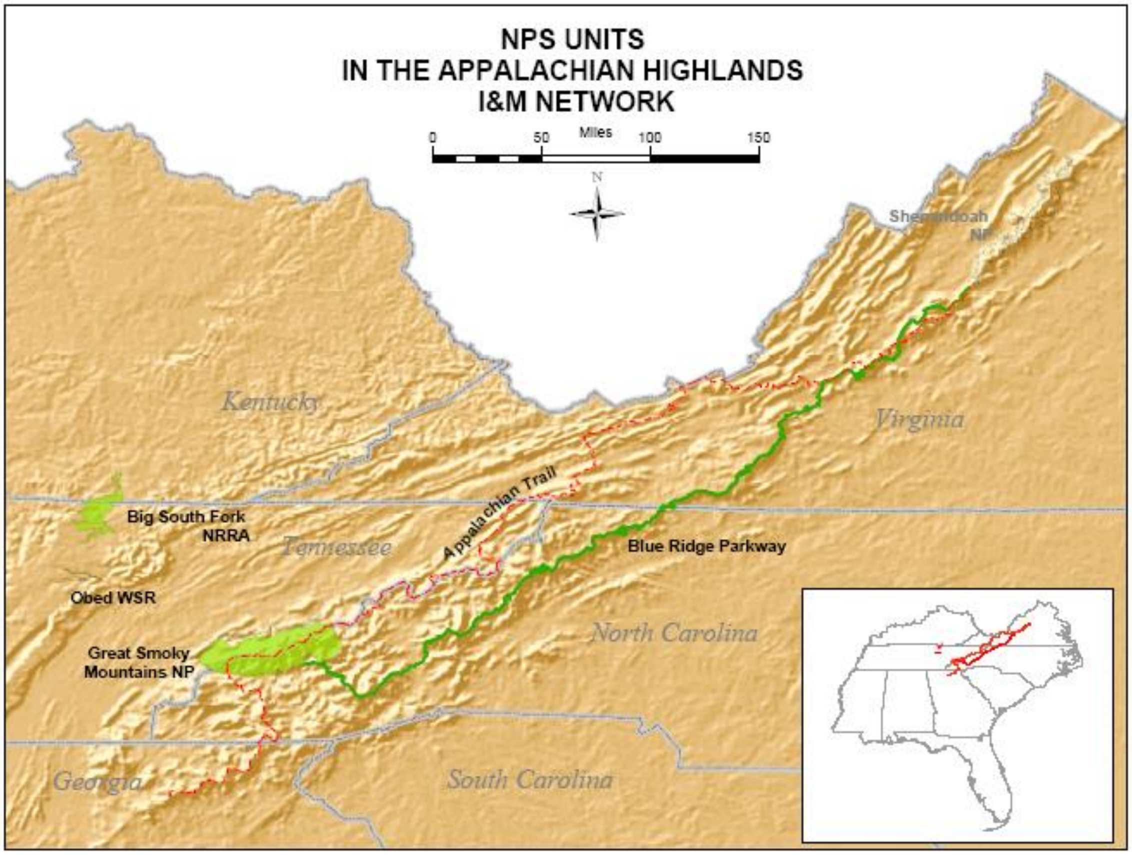


Figure 1.1. Locator map showing parks in the Appalachian Highlands I&M Network (map by Ron Cornelius - BISO).

During its first years, APHN implemented an inventory study plan, and a planning process was set in motion to accomplish long-term "vital signs" monitoring In accordance with national I&M goals, and APHN park priorities, Network activities revolve around five broad program themes:

- **CONDUCTING BASELINE INVENTORIES** of natural resources in the parks. Vascular plant and vertebrate surveys are currently underway and will document 90% of the species in each taxonomic group; detailed vegetation cover maps are also being prepared for each park from aerial infrared photos;
- **DEVELOPING A COORDINATED LONG-TERM ECOLOGICAL MONITORING PROGRAM** to efficiently and effectively monitor ecosystem status and trends over time. A draft long term monitoring plan will be completed by December, 2004; water quality monitoring will begin in 2005.
- **DEVELOPING DECISION SUPPORT SYSTEMS** (including GIS and other tools) to aid park managers in identifying, implementing, and evaluating management options;

- **INTEGRATING INVENTORY AND MONITORING** programs with park planning, maintenance, interpretation and visitor protection activities to help the parks in their efforts to make natural resource protection even more of an integral part of overall park management, and;
- **COOPERATING WITH OTHER AGENCIES AND ORGANIZATIONS** to share resources, achieve common goals, and avoid unnecessary duplication of effort and expense. A concerted effort is being made to identify and carry out cost-sharing, data sharing, and technology exchange opportunities with other agencies conducting similar inventories or monitoring.

1.3 APHN Data Management Goals

To be useful, a set of data – whether collected the previous year or 20 years ago – must also be accompanied by sufficient context about how and why it was collected, so that its value is retained beyond the lifetimes of those who collected it. Therefore, a data management program cannot simply attend to the tables, fields, and values that make up a data set. There must also be a process for developing, preserving, and integrating the context of the data that makes it interpretable and valuable.

Therefore, for our purposes, the term "data" can have a generic meaning, which refers to the entire array of information that is commonly maintained by an inventory and monitoring program, not just to the tabular and spatial data that are the primary targets of our data management efforts. These products fall into five general categories: raw data, derived data, documentation, reports, and administrative records (Table 1.1).

Table 1.1 Categories of data products and examples of each

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

These various categories of data all require some level of management to ensure their quality and availability when needed. Managing this data effectively requires a holistic view of how natural resources information is interrelated.

An integrated approach to Network data management is the cornerstone supporting the five broad program themes mentioned above (Section 1.2). This Plan describes the general procedures and practices APHN has adopted to ensure that we meet our overall data management goals, namely, that data collected by the Network are of high quality, are readily available, can be easily interpreted, and are secure for the long-term.

APHN Data Management Goals

1) To ensure that data managed by the network are of high quality, *including, designing standardized data entry, importation, and handling procedures which effectively screen for bad data, and minimize transcription and translation errors;*
2) ***To make certain network data are readily available,*** *by implementing standard procedures for distributing data, while protecting sensitive data; and designing a standardized filing system for organizing I&M information;*
3) To ensure that network data can be easily interpreted, *by considering the users' needs as the primary factor driving the design of summary reports and analyses; establishing rigorous data documentation standards; integrating common data tables and fields in the NPS database template format; and making summary information available in formats tailored to the variety of audiences interested in I&M program results;*
4) To make certain that data are secure for the long term, *including, instituting standard procedures for versioning, data storage and archiving; and maintaining the necessary hardware and software configurations to support network data management needs.*

1.4 Scope of the APHN Data Management Plan

This Plan describes the resources and processes the Network will employ to meet our data management goals. It is intended to be general in nature, establishing the principles and procedures which govern the overall data management program, while deferring detailed data handling considerations, database designs, and processing steps, to project-specific protocols. The Plan is directed primarily toward I&M staff, and park managers within the Network, however, it is also written to provide useful information to potential network partners, and others interested in how data is managed in a relatively small, long-term inventory and monitoring program, involving several parks and partners.

Chapter Two discusses I&M data management roles and responsibilities within the Network, including the roles of I&M project managers, and data managers. Chapter Three discusses the general steps involved in any generic I&M project, and how data management standards and procedures are integrated with these steps. In Chapter Four, the APHN data management infrastructure is described, including hardware, software, and LAN/WAN resources. Data acquisition, and processing considerations are covered in Chapter Five. The final chapters focus on quality assurance; project documentation procedures; data analysis and reporting; and long-term archiving of data.

This Plan is designed to lay out general data management principles and guidance for Network data management, and once it is finalized, should remain relatively unchanged. The Standard Operating Procedures (SOPs), in contrast, will be updated and revised frequently, particularly in the first years of the program.

A review of the Network monitoring program, including the data management component will be done periodically in future years after monitoring is implemented. The Data Management Plan may be revised at that time. Edits to change the language national systems in this plan were made in January 2014. Figures and diagrams 4.4 and 7.1 refer to the legacy systems: Dataset Catalog and SMMS as a metadata tools and NR Bib as a bibliography tool. These three systems were valid for when the plan was first written however over time newer systems have replaced their relevance. An effort was made to change URLs with valid web addresses and delete those links experiencing link rot, however some links may be invalid.

2. Data Management Roles and Responsibilities

2.1 Data Stewardship - A Shared Responsibility

Meeting Network data management goals requires the participation of everyone on the APHN staff, from field crews who collect data, to project managers who validate, analyze and summarize data, to the Network data manager, who ensures that "master" data sets are of high quality, and that proper data management standards and practices are adhered to. Because good data stewardship is so central to the mission of the Network inventory and monitoring program, significant staff time at all levels is devoted to that effort (Table 2.1).

Table 2.1. Appalachian Highlands I&M Network staff resources directed toward data management.

Title	# of Staff Positions	% of Time	Data Mgt. Activities	Total FTE	Total Cost (k)
Coordinator	1	30 %	data analysis, summary, and reporting, data validation and verification	.3	25.7
Data Mgr.	1	80%	data archiving and dissemination, database development, overall QA/QC	.8	54.4
Ecologists	2	35%	data analysis, summary and reporting, data validation and verification	.7	55.8
BioTechs	4	30%	data entry and verification	1.2	17.8
TOTAL				**3.1**	**153.7**

Successful data stewardship requires that Network field crews, project managers, and the data manager, work together as a team, to ensure that data are collected using appropriate methods and that resulting data sets, reports, maps, models, and other derived products, are of high quality, readily accessible, easily interpreted, and secure for the long term. All project participants receive training, briefings, and project materials about data stewardship from supervisors, project leaders, and the Network data manager. This kind of regular communication helps to emphasize the importance of good data management practices, and establishes a sense of ownership and accountability for carrying them out. The following section describes the inter-related roles of Network staff in ensuring effective data stewardship, from the earliest stages of a given project, through its useful life.

2.2 Data Management Roles

Project Managers: Because the data management aspects of every I&M project require the expertise and involvement of several people over a period of months or years, a project manager will usually be designated, and charged with keeping track of project objectives, requirements, and progress. A project manager coordinates all aspects of a project, whether it involves inventory or monitoring; whether it is oriented toward field data collection, or gathering existing data from other sources; or, whether it involves data collection and reporting on a regular basis, or is limited to periodic analysis over intervals of several years. In the APHN, the project manager is normally a Network ecologist – ideally, the person who has the best training or experience in the particular field of science which is the subject of the project.

THE PROJECT MANAGER'S DATA MANAGEMENT RESPONSIBILITIES INCLUDE:

- Developing basic project metadata documentation
- Documenting and implementing standard procedures for data collection and data handling, including deviations from those procedures
- Developing quality control measures, including certification of field operations, equipment calibration, species identification, data entry, data verification and validation
- Maintaining hard copies of data forms and archiving original forms
- Scheduling regular project milestones, including data collection periods, data processing target dates, and reporting deadlines
- Acting as the main point of contact concerning data content

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

- Develop quality assurance and quality control procedures
- Identify training needs for staff related to data handling procedures, quality control measures, and database software use
- Coordinate the design of 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
- Ensure regular archiving of project documentation, original field data, databases reports and summaries, and other products related to the project
- Create data summary procedures to automate the process of transforming raw data into meaningful information
- Identify and prioritize legacy data for conversion to desired formats
- Increase the accessibility and interpretability of existing natural resources information

The project manager is responsible for data quality during all phases of the project, including, data collection, QA/QC, analysis and reporting. Developing project documentation and metadata are crucial elements of this function. The project manager and data manager work closely together throughout the project to ensure that good data stewardship practices are followed.

Data Manager: The Network data manager has a central role in ensuring that project data conforms with program standards, designing project databases, disseminating data, and ensuring long-term data integrity, security, and availability. In order to maintain high data quality standards, and promote ready use of project data, the data manager collaborates with the project manager to develop data entry forms, QA/QC procedures, and automated reports. The APHN data manager

THE DATA MANAGER'S RESPONSIBILITIES INCLUDE:

- Developing and maintaining the infrastructure of metadata creation, project documentation, and project data management
- Creating and maintaining project databases in accordance with the best practices and current program standards
- Providing training in the theory and practice of data management tailored to the needs of project personnel
- Developing ways to improve the accessibility and transparency of digital data
- Establishing and implementing procedures to protect sensitive data according to project needs
- Establishing procedures for data dissemination
- Integrating tabular data with geospatial data in a GIS

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

- Define the scope of the project data, and create a data structure that meets project needs
- Become familiar with how project data are collected, handled and used
- Review quality control and quality assurance aspects of project protocols
- Identify elements that can be built into the database structure to facilitate quality control, such as required fields, range limits, pick-lists and validation rules
- Create a user interface that streamlines the process of data entry, review, validation, and reporting
- Ensure that project documentation is complete, complies with metadata requirements, and enhances the interpretability and longevity of project data
- Ensure proper archiving of project materials
- Identify and prioritize legacy data for conversion to desired formats

maintains spatial data themes associated with Network inventory and monitoring projects, and incorporates spatial data into the Network GIS. The data manager maintains standards for this data and the associated metadata, and develops procedures for sharing and disseminating GIS data to Network parks and partners.

2.3 Data Management Coordination

The APHN data manager works with national NPS I&M 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 with Network personnel, park staff, and cooperators, to promote and develop workable standards and procedures for the purpose of integrating datasets and making them useful for a wider variety of applications.

Key contacts for the Network data manager include park GIS and data managers, and the project leaders for each monitoring or inventory project. Regular 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 through a variety of forums, including, personal visits, phone calls, email, joint meetings, and training sessions.

The Network collaborates with other public agencies, universities and non-governmental organizations, either working together on inventory and monitoring projects, or sharing data and results from those projects. These relationships require coordination at all levels to ensure that data collected by NPS staff, cooperators, researchers, and others meet high quality standards, and that commonly accepted data management standards and procedures are adhered to.

3. Project Work Flow and the Data Management Process

3.1 Project Work Flow

This section describes the generic phases of projects, in terms of how natural resource data is generated, processed, finalized, and made available. The APHN manages short-term finite projects, such as inventories or pilot projects designed for the possibility of long-term monitoring; and long-term monitoring efforts, which have an indefinite life. From the standpoint of managing data, there is often a greater need for long-term projects to adhere to standards, in order to guarantee internal compatibility over time. Long-term projects often require a higher level of organization, documentation, and review.

However, both short-term and long-term projects share many work flow characteristics, and both generate data products needing management. Any I&M project managed by APHN is generally comprised of five primary stages: planning and approval; design and testing; implementation; product integration and dissemination; evaluation and closure (Figure 3.1). Each stage is characterized by a particular set of activities that are carried out by different people involved in the project. Following is brief description of these activities and the responsibilities they entail:

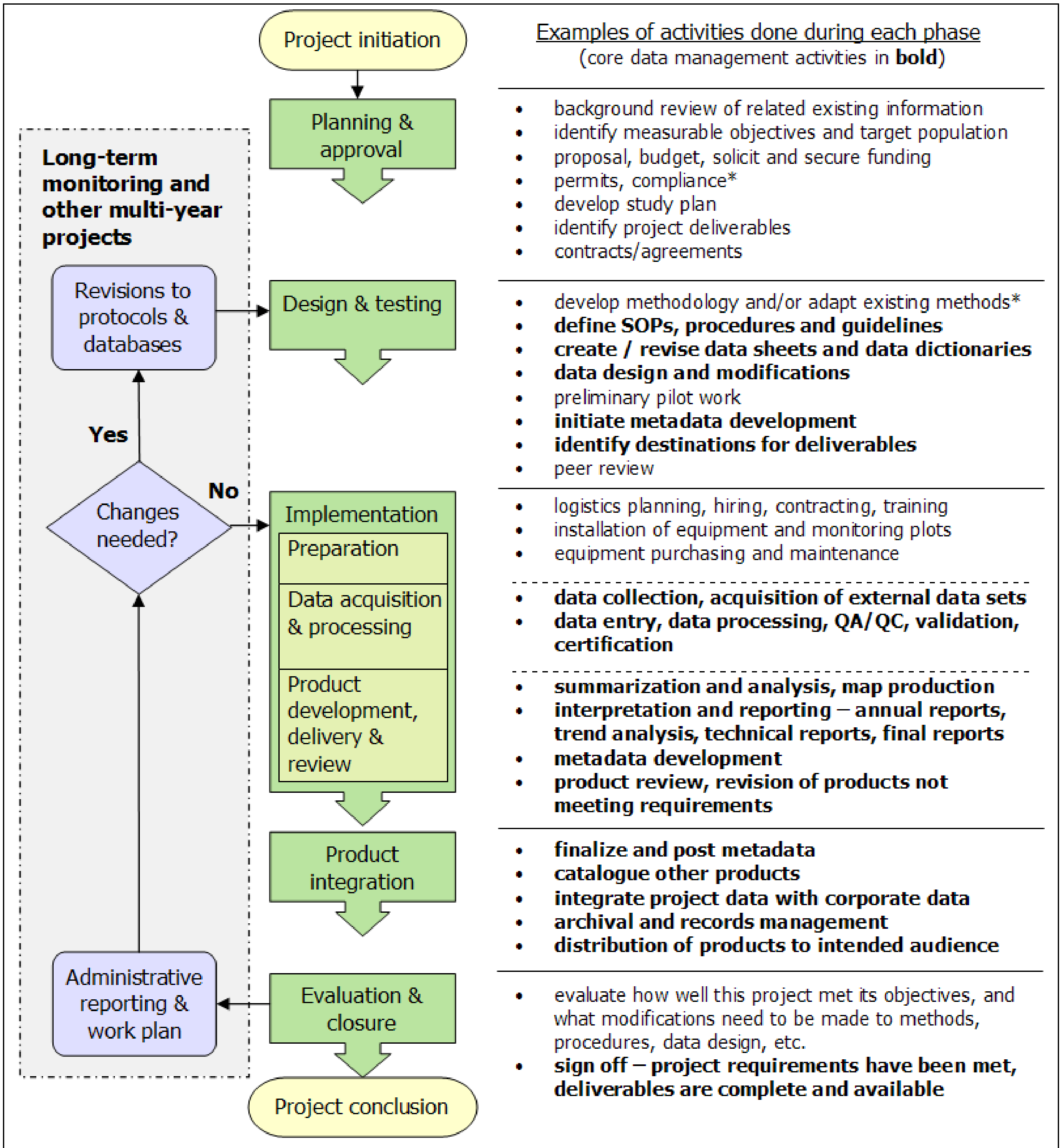
Planning and Approval

This first project phase is where many of the preliminary decisions are made regarding project scope and objectives. In addition, funding sources, permits and compliance are addressed. Establishing project objectives is the most important step in project development. It is crucial that network and park staff work together at this stage to establish why the data are needed, and how they will used. All contracts, agreements and permits should include. Primary responsibility throughout this phase rests with project managers and program administrators.

All contracts, agreements and permits should include standard language that describes the formats, specifications and timelines for project deliverables. Once a project is approved, a project tracking database is normally initiated, where contact information, project status, and due dates for deliverables are maintained. Also at this stage, the data management requirements of the project will be identified, along with the roles and responsibilities of project personnel.

Design and Testing

At this stage, specifications are established for how data will be acquired, processed, analyzed, reported, and made available to others. During design and testing, the project manager is responsible for developing project methodology, or modifying existing methods to meet project objectives. The project manager and data manager work together to develop specific procedures (SOPs) related to data acquisition, processing, analysis, and quality control. The data manager may also assist with the process of adapting existing data management SOPs to meet project objectives, and may provide a review of the clarity and completeness of the documentation.



* Determination of permit and compliance requirements often requires detailed knowledge of project methods, which are often not fully developed until after the planning and approval stage; therefore, methods development and compliance is an iterative process.

Figure 3.1. Diagram showing work flow for a generic I&M project, and how data management integrates with this process.

Also, at this stage, the project manager and data manager collaborate to develop the data design and data dictionary, where the specific data parameters that will be collected are defined in detail. Devoting adequate attention to this aspect of the project is possibly the single most important part of assuring the quality, integrity and usability of the resulting data.

Once the project methods, data design and data dictionary have been developed and documented, a database can be constructed to meet project requirements. Database development is generally the responsibility of the data manager. The resulting database applications should be thoroughly tested for problems by entering, editing, summarizing and exporting several dozen test records prior to implementing the database.

At the close of the design and testing phase, project metadata records should be initiated by the project manager. The data manager can facilitate the step by providing templates and software tools. It is also important at this stage to identify the destinations for project deliverables and to ensure that they will be produced in a manner that meets program requirements. Decisions should also be made regarding integration and permanent storage of deliverables as they are produced.

Implementation

During the implementation phase, data are acquired, processed, error-checked and documented. This is also when products such as reports, maps, GIS themes, and other products are developed and delivered. The project manager oversees all aspects of implementation – from logistics planning to data acquisition, report preparation, and final delivery. During this phase, data management staff serve as facilitators – providing training and support for database applications, GIS, GPS and other data processing applications; facilitation of data validation, summary and analysis; and assistance with the technical aspects of documentation and product development.

There are three main parts of the implementation phase:

- ❖ *Preparation* – This includes all aspects of logistics planning (housing, staging for access to remote sites), hiring and contracting, training, equipment procurement and maintenance, and installation of equipment and monitoring sites. This step may also involve preparations for handling and accessioning specimens, delivery of materials to contract labs for analysis, office or lab space acquisition, etc.

- ❖ *Data acquisition and processing* – The exact nature of data acquisition will vary widely from project to project. Many projects require a significant amount of field work and manual data collection, whereas others may solely involve the acquisition and processing of existing data. Data processing requirements also vary by project, but include all aspects of data entry and verification for accurate transcription, error-checking and data manipulation, and validation, for logical or structural problems with the data. All aspects of data acquisition should be specified in project protocols and SOPs. Similarly, quality assurance measures should be documented as part of the project metadata. Any deviation from these protocols should be documented, such as in the case of equipment failure, or adverse weather conditions, for example.

- *Product development, delivery and review* – The project staff works to develop and finalize the deliverables that were specified in the project study plan. Products that do not meet project specifications should be revised by the project manager or returned to the cooperator, or contractor, for revision. All raw and derived data products, metadata, reports and other documentation should be delivered to the data manager. Administrative records should be delivered to the appropriate network and park staff, as specified.

Product Integration and Data Dissemination

In this phase, data products and other deliverables are integrated into national and network databases, metadata records are finalized and posted in clearinghouses, and products are distributed, or otherwise made available to the project's intended audience. This is also when items that belong in collections, or archives, are accessioned and catalogued (Figure 3.2).

Product integration includes creating records for reports and other project documents in NPS Datastore, posting imaged documents to the appropriate repository, posting metadata records that have been completed and submitted by project managers, and updating NPSpecies to reflect any new species occurrence information derived from the project. This allows project information to be searchable and available to others via service-wide search engines.

Another aspect of integration is merging data from a working database to a corporate database maintained on the local network server. This occurs only after the annual working dataset has been certified for quality by the project manager. Certain projects may also have additional integration needs, such as when working jointly with other agencies, for a common database.

Evaluation and Closure

For long-term monitoring and other cyclic projects, this phase occurs at the end of each field season, and leads to an annual review of the project. For non-cyclic projects, this phase represents the completion of the project. After products are catalogued and made available, program administrators, project managers, and data managers should work together to assess how well the project met its objectives, and to determine what might be done to improve various aspects of the project methodology, and the usefulness of the resulting information.

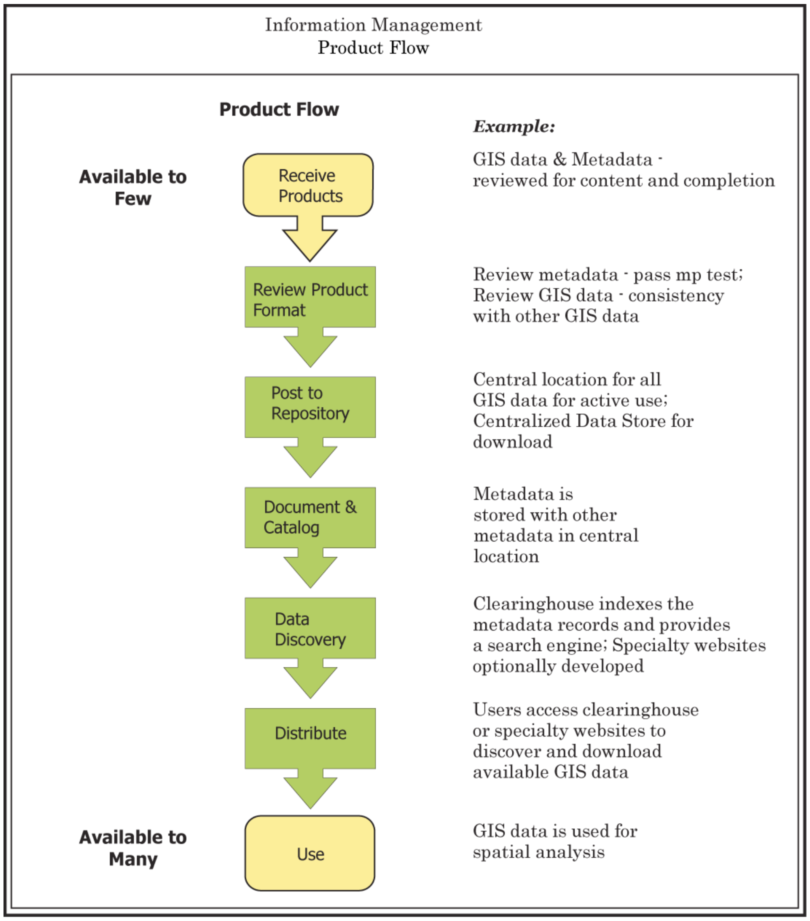


Figure 3.2. Data processing steps after receipt of project data.

4. Data Management Resources

The APHN relies on Network, park, regional, and national NPS offices to maintain the database systems, applications and software tools we use, as well as the computers and computer networks which are the foundation of our information management system. The following section describes the present APHN computer and networking environment, as well as the overall design of the NPS Inventory and Monitoring information management system and how it integrates with systems at the Network and park level.

4.1 Computer Resources Infrastructure

"Infrastructure" refers to the system of computers and computer networks that our information management system is built upon. Our Network infrastructure works with three main components: park-based local area networks (LAN); a data server maintained by network staff; and servers maintained at the national level. These components each host different parts of our natural resource information system (Figure 4.1):

National servers

- master applications – integrated client-server versions of NPS Datastore, NPSpecies, NR/GIS Metadata
- centralized repositories – Natural Resource Data Store, Protocol Clearinghouse
- public access sites – portals to NPS Datastore, NPSpecies, NPSFocus, websites for monitoring networks

Network data server

- master project databases – compiled data sets for monitoring projects and other multi-year efforts that have been certified for data quality
- common lookup tables – park name, employees, species
- project management application – used to track project status, contact information, product due dates
- network digital library – network repository for read-only **finished** versions of project deliverables for network projects (e.g., reports, methods documentation, data files, metadata, etc.)

Park LAN

- local applications – desktop versions of national applications such as NPSpecies and old Dataset Catalog
- working project materials – working databases, draft geospatial themes, draft copies of reports
- park digital library – base spatial data, imagery, and finished versions of park project deliverables

- GIS files – base spatial data, imagery, and project-specific themes that are managed from a central location

The Network depends on Information Technology specialists at BLRI, and BISO for help with hardware replacement, software installation and support, security updates, virus protection, telecommunications networking and backups of servers. APHN staff work with IT specialists as much as possible, to ensure Network infrastructure meets required standards, and that the best data security measures are in place.

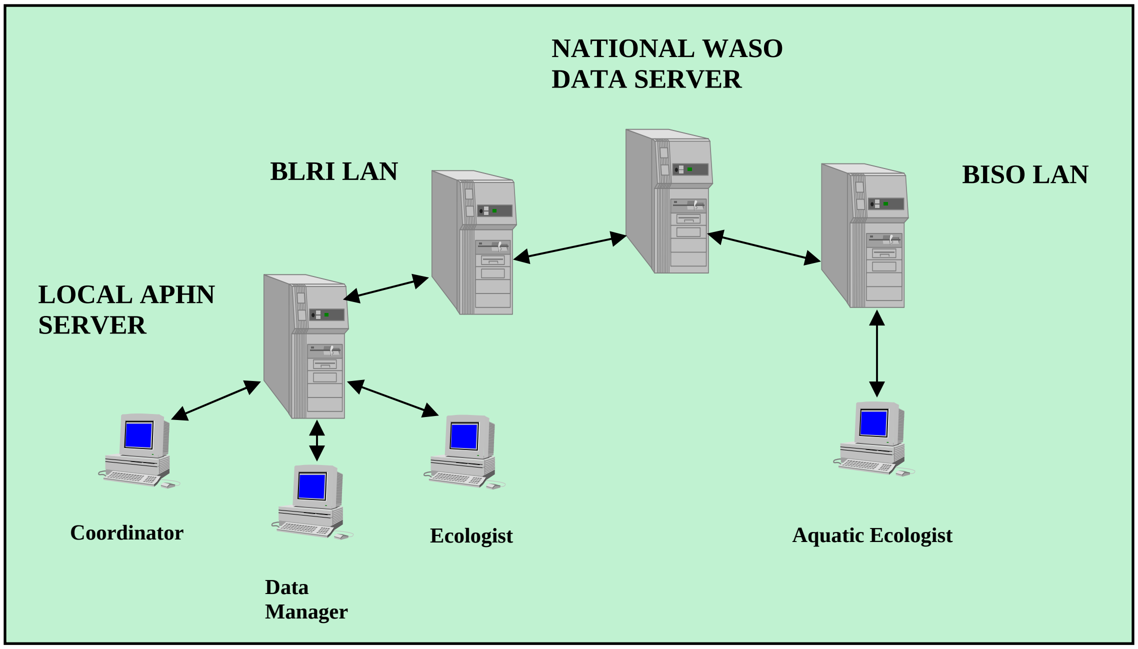


Figure 4.1. Schematic showing general connectivity of APHN computer resources.

The park and local Network servers have a hierarchical directory structure built around the same basic model. The key aspects of this model are:

1. Working files are kept separate from finished products.
2. Finished products are largely read-only
3. Standards such as naming conventions and hierarchical filing are enforced within libraries and database catalogs

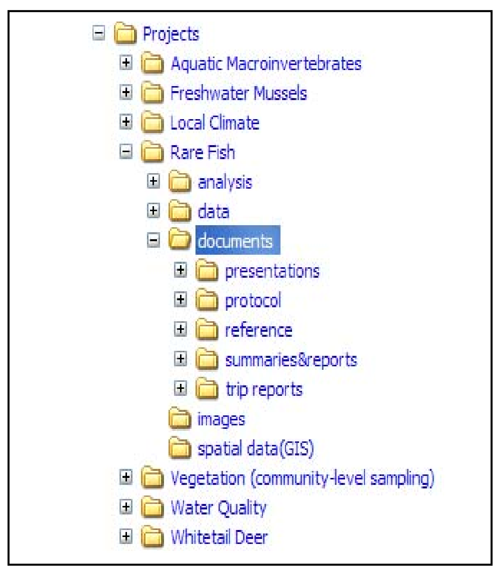


Figure 4.2 Network directory structure.

4.2 Network Systems Architecture

"Systems architecture" refers to the applications, database systems, repositories, and software tools that make up the framework of our data management effort. Rather than developing a single, integrated database system, the APHN data design relies upon standalone project databases that share design standards and links to centralized data tables. Individual project databases are developed, maintained, and archived separately.

The advantage of this design, is that it allows for greater flexibility in accommodating the needs of each project. 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.

Project database standards

Project database standards are necessary for ensuring compatibility among data sets, which is vital given the often unpredictable ways in which data sets need to be aggregated and summarized. When well thought out, standards also help to encourage sound database design and facilitate interpretability of data sets. Databases that are developed for park and network projects will contain the following main components:

> ***Compatibility with national standards***
>
> As much as possible, APHN standards for fields, tables and other database objects will mirror those conveyed through the Natural Resource Database Template. Where there are differences between local and national standards, documentation of the rationale for these differences will be developed. In addition, documentation and database tools (e.g., queries that rename or reformat data) will be developed to ensure that data exports for integration are in a format compatible with current national standards.

- *Common lookup tables* – Links to entire tables that reside in a centralized database, rather than storing redundant information in each database. These tables typically contain information that is not project-specific (e.g., lists of parks, personnel, and species).
- *Core tables and fields based on network and national templates*-These tables and fields are used to manage the information describing the "who, where and when" of project data. Core tables are distinguished from common lookup tables in that they reside in each individual project database and are populated locally. These core tables contain critical data fields that are standardized with regard to data types, field names, and domain ranges.
- *Project-specific fields and tables* – The remainder of database objects can be considered project-specific, although there will typically be a large amount of overlap among projects.

Centralized database components – common tables and data sets

Certain key information is not only common to multiple data sets, but to the organization as a whole – lists of contacts, projects, parks, species are often complex and dynamic. It is a good strategy to centralize such information so that users have access to the most updated versions in a single, known place. Centralizing also avoids redundancy and versioning issues among multiple copies. Centralized information is maintained in database tables that can be linked or referred to from multiple distinct project databases. Network applications – for project management, administrative reporting, or budget management – can also link to the same tables so that all users in the network have immediate access to edits made by other users.

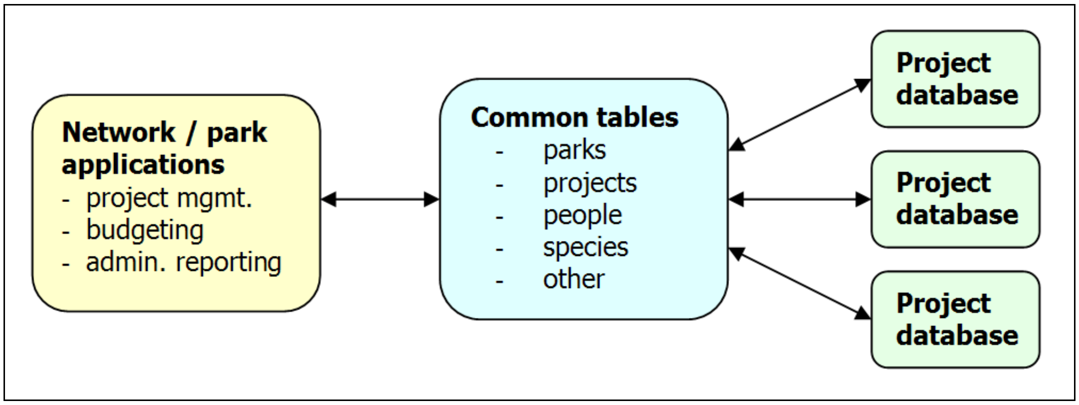


Figure 4.3. Linkage to common lookup tables.

Separating these tables by functional groupings is done primarily to reduce conflicts and performance losses associated with multiple users in MS Access (Table 4.1). Databases associated with individual projects each access the common tables via links established in each project back-end data file.

Table 4.1. Groupings for common lookup tables.

Grouping	*Description*
Parks	list of park units and networks
Projects	list of park and network projects, including inventories, monitoring, and external research projects
People	comprehensive list of contacts for parks and network, project-specific crew lists, lists of groups and users for tracking and managing access privileges
Species	comprehensive list of taxa for the park, linkage to NPSpecies, project species lists
Other Lookups	lists of watersheds, drainages, place names

4.3. National Information Management Systems

The need for effective natural resource information management cuts across NPS divisional boundaries and management strategies must be defined at the highest level possible. In this context, integrated inventory and monitoring of natural resources is multidisciplinary and requires national-level, programmatic data and information management strategies for success.

The basic strategy of natural resource and therefore inventory and monitoring information management is to provide integrated natural resource databases and information systems that enhance NPS managers' and staff's access and use of timely and valid data and information for management decisions, resource protection, and interpretation. Inventory and monitoring information needs are broadly separated into two categories:

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

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 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 (IRMA, NPSpecies, and NPS Datastore Metadata Database) utilize a distributed application architecture with both desktop and internet-accessible (master) components (Figure 4.2).

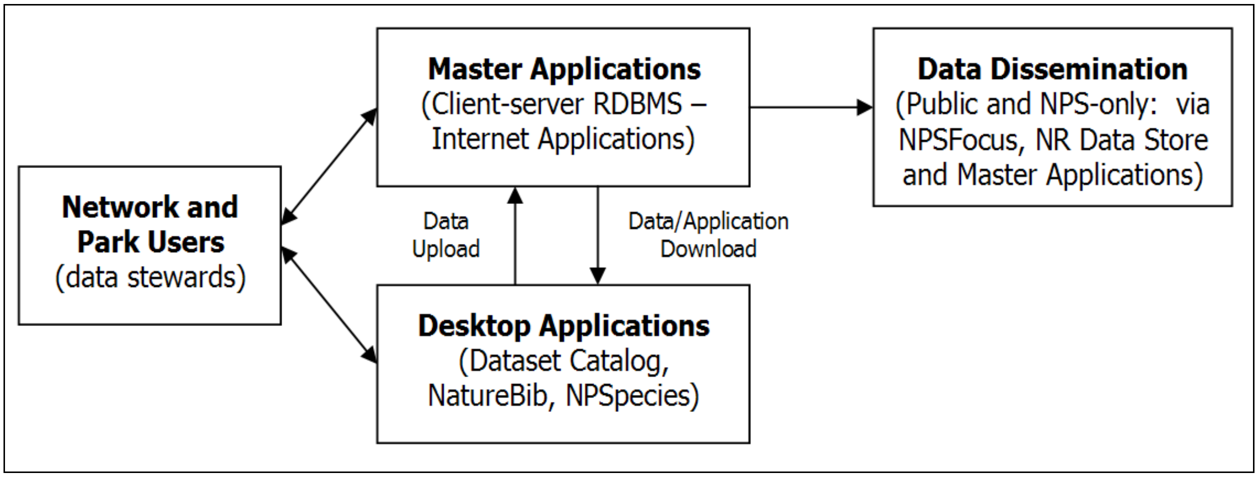


Figure 4.4. Model of the national-level application architecture.

IRMA and NR/GIS Metadata

IRMA is an on-line 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 IRMA 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 and project planning. As with other service-wide applications, the master metadata database (NR/GIS Metadata) is available through a website and will be linked to NPSpecies (the NPS species database). Hopefully it will be possible to download a version in MSAccess format from the master website. *NPS Datastore: https://irma.nps.gov/App/Reference/Welcome*

NPS Datastore

NatureBib is the master database for bibliographic references which 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 NR/GIS Metadata. It currently focuses on natural resource references, but may eventually be linked to references on cultural resources and other park operations. As with NPSpecies and NR/GIS Metadata, it is possible to download data from the master web version into the an Excel spreadsheet that can be used locally on computers without an internet connection. *https://irma.nps.gov/App/Portal*

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., vouchers and observations). Taxonomy and nomenclature are based on the 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 the MS Access version of NPSpecies. The internet-based version is the master database and is in active development as biological inventories proceed. The password-protected master version now available on the internet contains duplicate records, outdated species names, and various errors, and therefore requires clean-up and certification before any data will be available to the public. NPSpecies is linked to NPS Datastore for bibliographic references that provide written evidence of a species occurrence in a park and will be linked to NR/GIS Metadata to document biological inventory products. **See Appendix 1** for list of people with editing privileges working for the network or network parks. The MS Access application and additional details can be found at the NPSpecies website. *https://irma.nps.gov/NPSpecies/*

NPStoret

NPStoret is the master water quality and quantity database developed by the water resources division that is a subset of the EPA national database STORET. The database is still in development, but full metadata, protocol descriptions, sampling device information, parameter information, and reporting capabilities are available through a front-end form. A linked database of over 337,000 parameters for monitoring is available and can be paired-down for certain parameter measurements. A login and ID are required for the data entry and tracking of source information. The public can search, examine, and retrieve NPS-entered and other data from the STORET National Data Warehouse on the Internet. The STORET National Data Warehouse will contain a complete backup copy of the NPStoret database. The MS Access application and additional details can be found at the URL:
http://www.nature.nps.gov/water/infodata.htm STORET Web Site:
http://www.epa.gov/storet

4.4 Other National-Level Inventory and Monitoring Information Management

Natural Resource Database Template

The Natural Resource Database Template is a flexible, relational database in MSAccess 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 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 standardizing monitoring protocols. Within the monitoring protocols, modules including a written sampling protocol, database table structure, example data entry forms and quality checking routines, and

queries and reports are under development by I&M Program staff and data managers. Approved monitoring protocols, including the Database Template component, are made available through a web-based protocol clearinghouse. 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/datamgmt/applications/template/index.cfm

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 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 MSAccess that are consistent with the Natural Resource Database Template that have been developed for a particular protocol. *https://irma.nps.gov/App/ProtocolTracking*

NPS Data Store

The NPS Data Store is a key component of the data dissemination strategy employed by the I&M Program. The NPS Data Store 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 steward their data on the server. The NPS Data Store will be integrated with the master NR/GIS Metadata application to streamline programmatic data documentation and dissemination processes. See the NPS Data Store website for further information.

5. Data Acquisition and Processing

This chapter describes the steps involved in acquiring data as well as the initial stages of data processing. These steps are integral to strengthening the scientific foundation of the I&M program and to providing managers with the high quality data needed to effectively manage park resources. The Network program handles two general types of data:

- **Programmatic Data –** data produced from projects that are initiated or funded by the I&M program.

- **Non-programmatic Data –** data collected from other NPS sources or produced by external non-NPS sources.

The value of the data from these two sources is determined by the quality and usefulness of the data for addressing management, or scientific issues. The following sections outline in more detail the manner in which data are acquired and processed for these two kinds of data.

5.1. Programmatic Data

Projects initiated by the APHN typically involve I&M personnel, park staff, or cooperators/contractors. These efforts may consist of gathering existing information or conducting field data collection.

DATA DISCOVERY/DATA MINING

Data discovery, or data mining, refers to the process of acquiring information from existing sources. Data from recent studies, or historical information, called legacy data, can be very valuable, and may reveal natural resource inventory and monitoring information, or natural history collections from park areas, which NPS staff were not previously aware of. Data mining may involve remote searches of libraries or computer databases, or visiting museums or herbaria. Many electronic data sources are accessible via the internet.

SOME COMMON SOURCES OF RECENT OR LEGACY I&M DATA INCLUDE:

- National NPS databases (e.g. NPS Datastore, NPSpecies)
- Park archives and collections
- Online literature databases (e.g. First Search or Biosis)
- University and public libraries
- Regional and park GIS specialists
- Federal and state geographic data clearinghouses
- Museums and herbaria – both university and public collections

All information collected during the data discovery process is maintained by APHN in either electronic or hard copy format, depending on how the data was collected. Electronic datasets are entered in Dataset Catalog. Any geographic datasets obtained during data mining should be accompanied by FGDC compliant metadata. Information relating to the biodiversity of Network parks is entered into NPSpecies and linked to the associated reference, voucher, or observation.

Hard copies of reports, data sheets and field notes are copied and stored in file cabinets in the APHN offices. A filing system for these papers is being developed. The originals are archived in the appropriate park's collections. In the near future, the Network intends to scan important hard copy references and materials, saving them as .pdf files, in order to create a digital library, and allow easier dissemination of these documents.

FIELD STUDIES

The most common examples of projects conducted by the Network are biological inventories, and long-term vital signs monitoring. The Network Data Manager is responsible for ensuring that data collection, data entry, verification, storage and archiving for all field projects are consistent with APHN program standards. In addition to general standard operating procedures (SOPs) that define network-wide requirements, protocol-specific SOPs detail procedures and methods connected with data collection, storage and maintenance of project data. These SOPs may contain information ranging from the proper use of data entry forms or databases, to outlining calibration procedures for automated data loggers.

Field Acquisition Methods

Listed below are some of the devices available for field data collection. Protocol-specific SOPs describe how these tools may apply to individual projects.

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

- **Field Computers –** increase data collection and data entry efficiency. Data can be directly dumped from the field computers to the office desktops thereby eliminating the data entry step. Fewer chances for error as QA/QC checks can be built into the database. Could be inefficient if copious amounts of notes or comments need to be recorded in the field.

    - **Palm-top computers (PDAs) –** the small size and relative low cost of these devices makes them attractive options for collecting field. Good for small field projects but not powerful enough for large data intensive field projects. PDAs can be ruggedized fairly easily and at a relatively low cost. Most run either Windows CE or Palm operating systems which may require additional processing/programming to transfer/create the database structure in the field units.

- **Tablet PCs –** same properties as most laptops and provide the user with the convenience of a touch screen interface. They are bulkier, more expensive, and less rugged than the PDAs but are more powerful. Good for field projects that are very data intensive. Because these units run Windows XP (Tablet Edition) the project database can be directly transferred from desktop units to field units without additional programming steps.

- **Automated Data Loggers –** mainly used to collect ambient information such as weather data or water quality information. Must be properly calibrated so field crews must receive proper training and SOPs outlining the calibration/unit maintenance procedures.

  - **Permanently deployed devices –** often cost prohibitive. Data from these devices must be retrieved and batteries changed on a regular basis. These intervals should be defined in the protocol.

  - **Portable hand-held devices –** deployed for sampling only during site visits. Generally less expensive than units that are permanently deployed in the field.

- **GPS Units** – Two common types of GPS units:

  - **Handheld Garmin Units –** good for collecting general position information. Not recommended for obtaining high accuracy location information.

  - **Trimble ProXR GPS Receivers –** good for collecting highly accurate (sub-meter) location information. Because of their small size, they are often favored for field data collection.

Databases

Natural Resources Database Template (NRDT)
http://science.nature.nps.gov/im/apps/template/index.htm

All APHN field studies will have a Microsoft Access database associated with them. The Network has adopted the Natural Resources Database Template (NRDT) as the foundation for its database development program. The database template is highly flexible and can be modified and customized for each project to meet the needs and requirements of the researcher. The database incorporates mechanisms such pick lists and validation rules for quality assurance purposes.

- Field crews/project staff enter all data into the database provided to them.
- Field crews are required to periodically forward the data manager the project data files. Refer to individual protocols for the requirements regarding forwarding data.
- The data manager maintains the master copy of the database and updates it with data files received from the field crews.
- All data must undergo QA/QC procedures (see Chapter 6 of this document for more specifics relating to data verification and validation).

NPSpecies and Biodiversity Data

NPSpecies is a National Park Service database developed by WASO to store, manage and disseminate scientific information on the biodiversity of organisms in National Park Service units throughout the United States and its territories. The database is available in an on-line form (Oracle) or a desktop version (MS Access). For more specifics on NPSpecices, refer to the following web page:

https://irma.nps.gov/NPSpecies/Search/

Observations

- All observation data collected during biological inventories are entered into the NPSpecies database.

Vouchers

- Information relating to all voucher specimens collected during a biological inventory must be entered into NPSpecies.
- Once catalog and accession numbers are obtained all voucher information must also be entered into ANCS+.
- The parks where the vouchers were collected technically own the specimens and have the right to decide where the specimens will be stored.
- Refer to Appendix 5.X for more information concerning the processing of voucher data.

CHANGES TO DATA COLLECTION PROCEDURES

Changes to established data collection procedures are discouraged unless there are acceptable, valid reasons for altering the methodologies. Ideally, all problems should be identified during the design and testing stages of the project and changes implemented prior to the collection of any field data. Protocols should attempt to identify any foreseeable issues that might occur as well as contingencies to address them. Inevitably, unforeseen problems may occur which require procedures/protocols revision after data collection has begun. Significant changes to the protocols must be approved by the principal investigator, key official and the data manager. The key official must evaluate the proposed changes and determine if additional peer review is required before accepting them.

Altering data collection procedures or protocols may also occur as a result of the comprehensive review that all monitoring protocols undergo every five years. During the review, data is analyzed to determine if the current protocol accomplishing its objectives. If it is concluded that the protocol in its present form has not achieved the desired results changes could be recommended. Once again, all changes must be approved by the principal investigator as well as the key official and data manager.

5.2. Non-Programmatic Data

Two distinct kinds of non-programmatic data were defined in the introduction of this chapter: (1) *NPS* and (2) *non-NPS programmatic data*. The following section will further describe the data that fall into these classes and the manner in which the NCRN manages such data.

NON-PROGRAMMATIC NPS DATA

A large percentage of data collected in Network parks originates with projects initiated at the park level or by other NPS regional or national programs. The data collected and products produced by such efforts provide a great deal of information about park natural resources and are therefore very relevant to the mission of the I&M Program.

Park Data

Parks in the Network often use base funding or receive project funding through sources such as the Service-wide Combined Call.

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

Regional and National Programs

NPS regional and national programs support resource management efforts in the Network parks and also provide a good resource for natural resources information.

- **Air Resources –** These national-scale programs collect data, maintain databases, assure data quality, and perform the trend analysis relevant to NCRN air quality issues. The NCRN I&M program will rely on the data analyses from these national scale monitoring networks to obtain trends for many of its air vital signs.
- **EPMT -** Exotic Plant Management Teams (EPMT) collect and maintain data regarding the presence of exotic species in Network parks and the methods used to treat these species. This information is stored in the APCAM (Alien Plant Control and Monitoring Database) which is maintained by the EPMT data manager or EPMT liaison. (http://www.nature.nps.gov/biology/invasivespecies/EPMT_teams.cfm)
- **Fire Program –** Data concerning the occurrence of fires in Network parks is maintained both at the individual park level and regionally. National databases such as Fire-Pro, SACS and the soon-to-be implemented Fire Program Analysis (FPA) package (http://www.doi.gov/pmb/owf/fpa.cfm) have been and will be used to maintain information regarding resources devoted to fires as well as fire occurrences. The NPS is also involved in efforts such as the Joint Fire Science Program (http://www.nifc.gov/) that provides scientific information and support for fuel and fire management programs.
- **GIS –** Network parks are supported by the regional GIS specialist to help ensure that regional GIS data are available and accurate. Much of this data is also available through the Spatial Data Clearinghouse.
- **Water Resources –** Regional and national programs synthesize, analyze and interpret water resources data collected by parks.
- **Wildlife Management -** The Wildlife Management Program coordinates long-term monitoring and assessment of wildlife populations.

Data Processing

These data often do not require a great deal of processing because the I&M Program shares many of the file standards with Network parks and regional programs. Some basic processing steps include:

- Enter all new park biodiversity data into NPSpecies (this is especially important for park-based biological inventories) and enter all associated references into NatureBib.
- Ensure that all GIS data is in the proper projection and accompanied by compliant metadata.
- All data sets should be entered into and tracked using IRMA.

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

- Work closely with park and regional personnel to ensure that high quality data are available.
- Provide training to park staff interested in learning to use NPSpecies, NPS Datastore or the NRDT.
- Develop databases based on the NRDT that meet the needs of park resource managers.

NON-PROGRAMMATIC DATA – EXTERNAL DATA

As was the case in the previous section, external data sources often provide relevant information important to the mission of the Inventory and Monitoring program. It should be noted that such sources need not be directly connected to Network parks but may instead pertain to methodologies or protocols that could assist Network personnel with the development of a more productive program.

The APHN will rely on external sources for data to support three Vital Signs: air quality, weather, and landscape change (remote sensing data). In these cases, the agencies or organizations that collect these data have the expertise to conduct the proper quality control procedures and the capability to function as a repository and clearinghouse for the validated data.

Data Processing

Unlike the data from NPS sources, much of the data collected from external sources must undergo some degree of processing to meet program standards, however some of the basic processing steps are very similar.

- All GIS data obtained from other entities are stored in the proper format, have the correct spatial reference information and FGDC compliant metadata.
- All biodiversity data received from other entities should be entered into NPSpecies. This would include datasets like the Breeding Bird Survey. Also, if the data was taken from a report or published document, the reference must be entered into NatureBib.
- All data sets should be entered into and tracked using IRMA.

The level of data processing required for external data sets such as those used in the Vital Signs monitoring program depends on the desired output. Remote sensing datasets such as satellite imagery or aerial photography will require varying levels of processing depending on how they are received. These steps may include geospatial processing or spectral processing. Ideally, all spatial datasets will be received in a geo-referenced format and may only require geographic transformations to meet Network standards. Varying degrees of spatial and spectral processing may be necessary to adequately answer the proposed questions. The individual protocols will outline the necessary processing steps.

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

Analyses to detect trends or patterns in ecosystem processes and the condition of natural resources require data of documented quality that minimize error and bias. Data of inconsistent or poor quality can result in loss of sensitivity and incorrect interpretations and conclusions. The potential for problems with data quality increases dramatically with the size and complexity of the data set (Chapal & Edwards 1994).

The APHN data management program must ensure that our projects produce and maintain data of the highest possible quality. The Network will establish and document protocols for the identification and reduction of error at all stages in the data lifecycle. These stages include project planning, data collection, data entry, verification and validation, processing, and archiving. Although a data set containing no errors would be ideal, the cost of attaining 95%-100% accuracy may outweigh the benefit. Therefore, we consider at least two factors when setting data quality expectations:

- frequency of incorrect data fields or records
- significance of error within a data field

We are more likely to detect an error when we work with clearly documented data sets and understand what a “significant” error is within *that* data set. The significance of an error can vary with data sets and depends on where it occurs. For example, a two-digit number off by one decimal place is a significant error. A six-digit number, with the sixth digit off by one decimal place, is not a significant error. But one incorrect digit in a six-digit species number could indicate a different species. That is a significant error.

6.1. National Park Service Mandate for Quality

If we expect our current data to be useful to future users, the data must survive changes in technology. We can promote data longevity through high-quality documentation and maintenance during all phases of data management: during data collection, entry, verification, and validation. Well-documented data sets are especially important when sharing data.

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.

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 need to be accessible to be used for wise and defensible decision-making at all levels. Data need to be able to be shared and aggregated with data from other parks and from adjacent lands to support landscape-level and national planning and decision-making."

6.2 QA and QC Mechanisms

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

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

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

QA/QC procedures applied to ecological data include four procedural areas (or activities), ranging from simple to sophisticated, and inexpensive to costly:

1) defining and enforcing standards for electronic formats, locally defined codes, measurement units, and metadata
2) checking for unusual or unreasonable patterns in data
3) checking for comparability of values between data sets
4) assessing overall data quality

Much QA/QC work involves the first activity (defining and enforcing…), which begins with data design and continues through acquisition, entry, metadata development, and archiving. The progression from raw data to verified data to validated data implies increasing confidence in the quality of the data through time. The progression from raw data, to verified data, to validated data implies an increasing confidence in the quality of the data through time.

6.3 Roles and Responsibilities

Quality assurance methods should be in place at the inception of any project and continue through all project stages to final archiving of the data set. It is critical that each member of the data management group work to ensure data quality. Everyone plays a part in producing and maintaining high quality data. Anyone assigned to a project is responsible for the quality of the results generated from his or her task(s).

The data manager is responsible for:

- developing protocols and SOPs to ensure data quality
- making project managers, technicians, etc., aware of the established procedures and enforcing adherence to them
- evaluating the quality of all data and information against NPS standards before dissemination outside the network
- performing periodic data audits and quality control checks to monitor and improve the data quality program

Project managers 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
- 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.

6.4 Goals and Objectives

We must ensure that a project produces data of the right type, quality, and quantity to meet project objectives and user needs. Quality criteria should be set at a level proportionate to project-specific objectives, and these criteria should indicate the level of quality acceptable for resulting data products. The EPA (2003) defines data quality objectives as qualitative and quantitative statements that:

- clarify the intended use of the data
- define the type of data needed to support the decision
- identify the conditions under which the data are to be collected
- specify tolerable limits on the probability of making a decision error due to uncertainty in the data

The most effective mechanism for ensuring that a project produces data of the right type, quality, and quantity is to provide procedures and guidelines to assist the researcher in accurate data collection, entry, and validation. APHN 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 validation.

Although specific QA/QC procedures will depend upon the individual vital signs being monitored and must be specified in the protocols for each monitoring vital sign, some general concepts apply to all network projects. The general QA/QC procedures presented in this plan were primarily adapted from the Draft Data Management Protocol (Tessler & Gregson 1997) and the ideas contained in Michener and Brunt (2000). These general guidelines will ensure that all data collected are checked for integrity before being integrated into the monitoring program databases. Refer to SOPs and monitoring protocols for specific QA/QC procedures.

6.5 Data Collection

Careful, accurate recording of field observations in the data collection phase of a project will help 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.

Paper field notebooks or data forms have been the primary methods for ecological data collection for many years. Although paper has advantages in terms of longevity and ease of use, it does not work well under some environmental conditions, and processing options are limited until the data are transferred to digital format. As an alternative to paper, several options for electronic data collection in the field are now available, including handheld computers, automated data collection instruments, and tape recorders.

Before the data collection phase of a project begins, the data manager is responsible for providing the protocols/SOPs for data collection and storage to the project manager. All field sheets and field data recording procedures must be reviewed and approved by the data manager and documented in the protocol SOPs. The project manager, in turn, will ensure that field crews understand the procedures and closely follow them in the field. If training is needed, the data manager will work with the project manager to provide that training. Field technicians are responsible for proofing raw data forms in the field, ensuring their readability and legibility, and verifying and explaining any unusual entries. They are expected to understand the data collection forms, know how to take measurements, and follow the protocols.

6.5.1 Methods for Reducing Collection Errors

Use a formatted, project-specific data sheet as opposed to a field notebook. When electronic data collection devices are not used, data should be recorded on paper data forms. We strongly recommend acid-free paper 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.

Standardized data sheets that identify the pieces of information to be recorded and forms that reflect the design of the computer data entry interface will help ensure that all relevant information is recorded and subsequent data entry errors are minimized. Data sheets should contain as much basic preprinted project information as possible and sufficient space for recording relevant metadata such as date, collectors, weather conditions, etc. They should clearly specify all required information, using examples where needed to ensure that the proper data are recorded. Data recorders should adhere to the following guidelines:

- All information added to the data sheet must be printed and clearly legible.
- If alterations to the information are necessary, the original 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.

Consider using a handheld computer when appropriate. The use of handheld computers minimizes the need for manual data entry from field forms and associated transcription and data entry errors. Specially designed database or computer programs may be required for handheld computers, and the user interface should be customized to the project requirements. A customized data entry application has the advantage of incorporating on-the-spot QA/QC checks, so this data collection method probably provides the highest quality data when combined with point-of-entry data quality checks.

These portable units, however, are subject to environmental constraints such as heat, dust, and moisture. When handheld computers are used for data entry in the field, the data should be downloaded daily to avoid potential loss of information. Thus, if a handheld unit fails during data collection, only the current day's data are lost. Batteries should be checked prior to a data collection trip, and they should be charged at the end of every field day. The use of a memory card that will store the data in case of damage to the unit or battery failure can prevent accidental loss of data. Also, in case the unit becomes inoperable in the field, printed data sheets should always accompany field teams on data collection trips.

Use automated data loggers where appropriate. Instruments with their own data acquisition systems are useful for collecting some types of data, such as water and air quality data. These devices can be calibrated and programmed to automatically record data and store them for later download directly to a computer, thereby eliminating the possibility for manual data entry errors. Data loggers are an efficient method for recording continuous sensor data, but routine inspections are necessary, and environmental constraints, as well as power (e.g., sufficient battery charge) and maintenance requirements, are potential pitfalls when using these instruments. Regular downloads should be required since physical memory is usually limited.

Use a handheld tape recorder. Another alternative to paper field data forms is a handheld micro cassette tape recorder. Recorded observations are subsequently transcribed to paper or directly entered into computer files. As with other technological solutions, there are drawbacks including battery and tape maintenance, low environmental tolerance, and risk of failure. However, if a single data collector is in the field, tape recorders can provide an easily operated, high quality, efficient method of collecting data.

Consider calibration, maintenance, and minimum timing requirements of field equipment. Accurate field measurements are possible only if field equipment is regularly calibrated and maintained. Where appropriate, consult reference manuals for recommended calibration and maintenance procedures. Once in the field, allow sufficient time for field equipment to adjust to its environment so it will record accurate measurements (for example, when using water quality probes and GPS units). Researchers should consider maintaining records of equipment calibration and failures that accompany their field data whenever possible.

Be organized and keep a log. Organization is the key to good data collection methods. Maintaining a log of important decisions and events will help clarify information and contribute to an accurate report.

Ensure that field crews receive proper training. Although protocols and SOPs are in place, they cannot guarantee that high quality data will be collected. Prior to routine data collection for a project, conduct training sessions to ensure that field personnel have a clear understanding of data collection procedures described in the SOPs. A training program may also include a process to certify that field staff understand and can perform the specified data collection procedures. The development of a training manual may be helpful for long-term monitoring data collection efforts and those that will involve a large number of field staff. Palmer and Landis (2002) provide an outline for a training manual and suggestions for planning training sessions.

Perform quantitative assessments of data quality. Repeating measurements is the primary tool for performing quantitative assessments of data. Project managers should periodically review the work of field technicians to ensure that their work does not drift from standards during the course of the field season. Quantitative assessments may be considered if staff and funding are available, and Palmer and Landis (2002) describe several approaches that can be used.

6.6 Data Entry

'Data entry' is the initial set of operations where we transfer raw data from paper field forms into a computerized form linked to database tables. Spreadsheets should not be used for data entry (data can be exported to a spreadsheet for manipulations post entry). When data are gathered or stored digitally in the field (e.g., on a data logger), data entry consists of the transfer of data (downloading) to a file in an office computer where they can be further manipulated.

Transferring data from field projects into the computer seems a fairly simple task. But the value of the data depends upon their accuracy, and we must feel confident about the overall quality of the data. 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*. A few transcription errors are unavoidable during data entry so, all data should be checked and corrected during a data verification process.

The data manager, along with the project manager, should provide training in the use of the database to all data entry technicians and other users. The project manager makes certain that data entry technicians understand how to enter data and follow the protocols. Data entry technicians are responsible for becoming familiar with the field data forms and differences in handwriting. They must also become familiar with the database software, database structure, and any standard codes for data entry used by the Network. At minimum, data entry technicians should know how to open a data entry form, create a new record, and exit the database properly. They must learn how to commit both a 'field' entry and a "complete record" entry and to correct mistakes made while typing.

6.6.1 Methods for Reducing Data Entry Errors

Enter or download data in a timely manner. All data should be entered or downloaded into the project database as soon as possible, preferably at least once a week. Try to avoid delaying data entry until all the project data have been collected. Downloaded data should be periodically stored on CD or some other semi-permanent media.

Design efficient data entry forms and methods. A full-screen data entry form that mimics the field data forms can effectively reduce manual data entry errors due to the 1:1 correspondence of the attributes. A strategy to distinguish between validated data and newly entered data should be adopted. Data can be entered into an empty, fresh database table to avoid contaminating existing data and the new data appended to the master data only after formal verification, validation, and documentation. Alternatively, we can include validation attributes that indicate which data have been checked and validated by the project manager in the database. Regardless of strategy, we must clearly document the process for validation in the protocol data management SOP.

Build automated error checking features into the database. The most robust QA/QC measures for data entry should be built into the database design to perform automatic validation checks of data. Data entry forms reduce transcription errors through auto-filled fields, range limits, pick lists, and spelling checks. They provide controlled access to the database (i.e., forms are set for data entry only, which prevents accidental deletion or alteration of existing data). They control the sequence of data entry (i.e., certain fields require an entry before more information can be entered). They warn the operator when errors are made and provide an opportunity for correction before the data are committed to a file.

- *Auto-filled fields.* Whenever possible, the data in a field should be auto-filled by the computer. For example, if a location ID is comprised of a park code, project code, and a unique number, those elements are automatically inserted into the location ID field, ensuring that the record always contains a unique identifier.
- *Range limits.* Where the appropriate values for a particular field span a finite range, the data entry program can check the entered value against the specified minimum and maximum values for that parameter. When a value is outside the accepted range, a warning message appears and asks the user to reenter a valid value. For some fields, values outside a specified 'normal' range may be acceptable. In this case, the warning message asks the user to verify the entry before continuing.
- *Pick lists.* The data entry application may also use pop-up pick lists for standardized text items where spelling errors can occur. For example, rather than typing in a species code or name (where a misspelling generates a new species in the database), the code or name is selected from a list of valid species codes or scientific names and automatically entered into the species field. A pick list may also be used when only certain entries are acceptable. Lists are not appropriate for all written fields but should be used when appropriate.
- *Unique constraints.* Duplicate and incorrect data entry can often be caught with the application of unique constraints on data entry fields. These constraints are particularly useful when importing data from other applications.

Provide a clean, organized work environment. Desktop space near the computer should be free of clutter and distractions that could the technician to lose her place. There should be enough space for two stacks of paper documents, one from which data are being entered and one from which data have been entered. A pad or notebook and some fine colored markers should also be available for making notes. (The need for a clean workspace also applies to the verification and validation phases.)

If possible, use two data entry technicians for data entry. When one technician reads the data from the field data forms and another enters them into the computer, the work is often faster and results in a lower error rate. If only one person is available, he should work at a slower pace to avoid errors. Like many monotonous tasks, data entry can be done in a personal rhythm that reduces the tediousness of the work.

6.7 Verification and Validation Procedures

We appraise data quality by applying verification and validation procedures as part of the quality control process. These procedures are more successful when preceded by effective quality assurance practices. *Data verification* checks that the digitized data match the source data, while *data validation* checks that the data make sense. It is essential that we validate all data as truthful and do not misrepresent the circumstances and limitations of their collection. Failure to follow SOPs for data entry, validation, and verification will render a data set suspect. Although data entry and data verification can be handled by personnel who are less familiar with the data, validation requires in-depth knowledge about the data.

The data manager establishes SOPs for verification and validation and provides them to the project manager. The project manager will ensure that the SOPs are followed. The project manager or designee will validate the data after verification is complete. The project manager is also responsible for reviewing all data products and reports before they are released outside the Network. The data and project managers will evaluate the results of verification and validation and determine any procedural or data form revisions that may be indicated by the results. Technicians will follow the SOPs for verification of data, make required changes, and document those changes.

6.7.1 Methods for Data Verification

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 we have verified the computerized data as accurately reflecting the original field data, we can archive the paper forms and manipulate and analyze most data on the computer.

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

1) *Visual review at data entry.* The data entry technician verifies each record after input. She compares the values recorded in the database with the original values from the hard copy and immediately corrects any errors. This method is the least complicated since it requires no additional personnel or software. Its reliability depends entirely upon the person keying data and thus, is probably the least reliable data verification method.
2) *Visual review after data entry.* All records are printed upon the completion of data entry. The values on the printout are compared with the original values from the hard copy. Errors are marked and corrected in a timely manner. 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).
3) *Duplicate data entry.* The data entry technician completes all data entry, as normal. Random records are selected (every **n**th record) and entered into an empty replica of the permanent database, preferably by someone other than the person keying the permanent data. Then we use a query to automatically compare the duplicate records from the two data sets and report any mismatches of data. Then we manually review any disparities and correct if necessary. This method adds the overhead of retyping the selected records, as well as the creation of a comparison query, but it becomes increasingly successful as the value of **n** decreases. Professional data entry services frequently use this method.

We can also use the entered data to calculate simple summary statistics with statistical software. These summary statistics can help us catch a duplicate or omitted entry. For example, we can view the number of known constant elements, such as the number of sampling sites, plots per site, or dates per sample. We can pose the same question in different ways; differences in the answer provide clues to errors. The more checks we devise to test the completeness of the data, the greater our confidence that we have completely verified the data.

To minimize transcription errors, our policy is to verify 100% of records to their original source by permanent staff. In addition, 10% of records are reviewed a second time by the project manager, and we report the results of that comparison with the data. If the project manager finds errors in her review, then we verify the entire data set again.

Although we may have correctly transcribed the data from original field notes or forms, they still might be inaccurate or illogical. For example, entries of stream pH of 25.0 or a temperature of 95°C in data files raise doubt about their accuracy; and such entries almost certainly are incorrect, whether or not they were properly transcribed from field forms. This process of reviewing computerized data for range and logic errors we call *validation*, and it can accompany data verification *only* if the operator has comprehensive knowledge about the data. More often, validation is a separate operation carried out *after* verification by a project specialist who can identify generic and specific errors in particular data types. Corrections or deletions of logical or range errors in a data set require notations in the original paper field records 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 data set since field sheets or other sources of data could easily be 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.

6.7.2 Methods for Data Validation

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

Data entry application programming. Certain components of data validation are built into data entry forms. The simplest validation during data entry is range checking, such as ensuring that a user attempting to enter a pH of 20.0 gets a warning and the opportunity to enter a correct value between 1.0 and 14.0 (or better yet, within a narrow range appropriate to the study area). Not all fields, however, have appropriate ranges that are known in advance, so knowledge of what are reasonable data and a separate, interactive validation stage are important.

Edwards (2000) suggests the use of ‘illegal data’ filters, which check a specified list of variable value constraints on the master data set (or on an update to be added to the master) and create an output data set. This output data set includes an entry for each violation, along with identifying information and an explanation of the violation. They illustrate the structure of such a program, written in the SAS® programming language.

A caveat should be interjected regarding the operative word ‘illegal’. Even though a value above or below a given threshold has never before been observed and the possibility that it could occur seems impossible, such an observation is not always an illegal data point. Edwards (2000) points out that one of the most famous data QA/QC blunders to date occurred when NASA’s computer programs deleted satellite observations of ozone concentrations that were below a specified level, seriously delaying the discovery of the ozone hole over the South Pole.

Outlier Detection. According to Edwards (2000), "the term outlier is not (and should not be) formally defined. An outlier is simply an unusually extreme value for a variable, given the statistical model in use." Any data set will undoubtedly contain some extreme values, so the meaning of 'unusually extreme' is subjective. The challenge in detecting outliers is in deciding how unusual a value must be before it can (with confidence) be considered 'unusually' unusual.

Data quality assurance procedures should not try to eliminate outliers. Extreme values naturally occur in many ecological phenomena; eliminating these values simply because they are extreme is equivalent to pretending the phenomenon is 'well-behaved' when it is not. Eliminating data contamination is a better way to explain this quality assurance goal. If contamination is not detected during data collection, it is usually only be detected later if an outlying data value results. When we detect an outlier, we should try to determine if some contamination is responsible.

We can use database, graphic, and statistical tools for ad-hoc queries and displays of the data to detect outliers. Some of these outlying values may appear unusual but prove to be quite valid after confirmation. Noting correct but unusual values in documentation of the data set saves other users from checking the same unusual values.

Other exploratory data analyses. Palmer and Landis (2002) suggest that in some cases, calculations for assessments of precision, bias, repetition, completeness, and comparability may be applicable and that for certain types of measurements, evaluation of a detection limit may also be warranted (the authors provide examples of procedures that may be applicable). Normal probability plots, Grubb's test, and simple and multiple linear regression techniques may also be used (Edwards, 2000; the author provides SAS and Splus code for constructing normal probability plots and examples of output showing normal and non-normal distributions).

6.8 Version Control

The APHN manages files from a multitude of sources, comprised of many formats with many iterations of a particular product. Some of the files are complete, some are works-in-progress, and for others the status cannot be determined. In addition to files it collects, the Network also generates many files, some of which fall into the complete, works-in-progress, and undetermined status categories. Determining the status of a single file can be difficult, but determining the current file within a series of similarly named files can be almost impossible.

Version control is the process of documenting the temporal integrity of files as they are being changed or updated. Change includes any alteration in the structure or content of the files, and such changes should not be made without the ability to undo mistakes caused by incorrect manipulation of the data. Whenever we complete a set of data changes, we should save the file with a unique name, a simple act that should become routine for all data handlers.

Prior to any major changes to a file, we should store a copy of the file with the appropriate version number. This allows the tracking of changes over time. With proper controls and communication, versioning ensures that only the most current version is used in any analysis.

The data manager determines the version control method that will be used, and other network personnel are responsible for accurately designating versions for any files upon which they have worked. Software tools that assist in file management can be helpful–for example, to create databases that include fields to record the file's revision history. Backup routines can be built into the databases that allow for automatic file renaming and archiving. Important program files can be catalogued in a simple index or more formally tracked and archived using professionally developed version control software.

6.8.1 Version Control Options

Dates. Using a date provides logical version control. The date is usually formatted as YYYYMMDD or YYMMDD, where DD is optional (depending on the frequency of changes). One drawback to this method: dates may be hard to read, which can confuse users who open the wrong version of a file.

Sequential numbers. We can designate versioning of archived data sets by adding a number to the file name, e.g., 001 or V1.0 for the first version. We assign each additional version a sequentially higher number. We should also document the date that a file becomes a new version, perhaps through assignment of database folder (or directory) names.

Version control software. We can eliminate the work of differentiating multiple versions of documents by using version control software to append modifying characters to the file name. Such software applications track changes made to a document, add comments related to the different document iterations, and retrieve the document at any recorded stage of development. These applications are available in either desktop or online formats.

Version control software should:

- Preserve previous versions of documents for possible recovery
- Track documents as they change during the course of the developmental and editorial phases of document/report creation
- Prevent conflicts between multiple collaborators by prohibiting multiple edits to the same file at the same time
- Evaluate the document creation process by tracking who changes a file, when they make the change, and what changes they make
- Reduce storage requirements by eliminating multiple copies of complete documents

6.9 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 can 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. Therefore, the APHN must evaluate and identify the types of information it will disseminate that will be subject to the guidelines. 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. Also, mechanisms must be in place for receiving and addressing comments/complaints pertaining to the quality of data.

Data are distributed to the public through the APHN I&M web page, national web sites such as the Biodiversity Data Store and the NPS Data Store, and public access databases such as NPSpecies and NPS Datastore. Any information distributed through any of these mechanisms must undergo internal QA/QC procedures and be approved for release.

Data Quality Review Methods

The Network will establish guidelines and protocols to ensure compliance with DO #11B. These protocols will document both internal and external review procedures for data and information disseminated outside the network, as well as a process for processing complaints about data quality.

Edwards (2000) suggests the initiation of quality circles, regular meetings of project managers, the data manager, and data management personnel for discussing data quality problems and issues. These meetings promote teamwork attitudes while focusing brainpower on data quality issues. Participants become more aware of quality issues and learn to anticipate problems. Moreover, all participants develop a greater appreciation of the importance of their role in data quality and the entire monitoring effort.

Value of Feedback from QA/QC Procedures

Quality assurance procedures may need revision to improve quality levels if random checks reveal an unacceptable level of data quality. Quality checks should not be performed with the sole objective of eliminating errors, as the results may also prove useful in improving the overall process. 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, 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.

We can modify field data forms 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 we find the same type of validation error occurring repeatedly in different data sets, 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.

Monitoring Conformance to Plans and Standards

Data managers can use periodic data audits and quality control inspections to maintain and improve their data quality program. They must verify that staff is operating in conformance with the data quality procedures specified in this plan and the protocol specific data management plans. They should track and facilitate the correction of any deficiencies. These quality checks promote a cyclic process of continuous feedback and improvement of the both the data and quality planning process.

Periodic checks by the data manager to see if network staff are adhering to the data quality procedures established in the Data Management Plan and protocols SOPS may include verification of the following:

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

The results of quality assessments should be documented and reported to the research staff and the network coordinator. The project manager and coordinator are responsible for ensuring that non-conformities in data management practices are corrected.

Communicating Data Quality

The Network will use data documentation and metadata to notify end users, project managers, and network management of data quality. A descriptive document for each data set/database will provide information on the specific QA/QC procedures applied and the results of the review. Descriptive documents or formal FGDC-compliant metadata will document quality for spatial and non-spatial data files posted on the Internet.

7. Data Documentation

7.1. Purpose of Metadata

Data documentation is a critical step toward ensuring that data sets are useable for their intended purposes well into the future. This involves the development of metadata, which can be defined as information about the content, quality, condition and other characteristics of data. Additionally, metadata provide the means to catalog datasets, within intranet and internet systems, thus making their respective datasets available to a broad range of potential data users.

While the importance of metadata is universally accepted within the data management community, there are many approaches to data documentation, involving varying levels of detail. Following are some considerations which an I&M network manager should consider in the development of data documentation strategies.

- 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).

7.2. NPS Integrated Metadata System Plan and Tools

The NPS Metadata System Plan is limited to three recommended desktop applications for collecting metadata. These include Dataset Catalog (developed by the I&M Program), and two commercial off the shelf metadata tools, ArcCatalog and SMMS. The following is a brief description of each of these tools:

Dataset Catalog: Dataset Catalog is a tool for cataloging abbreviated metadata on geospatial and biological data sets providing parks and networks a means to inventory, organize, and maintain information about data set holdings locally. Dataset Catalog is not intended to be an exhaustive metadata listing, but assistance for 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 to be imported into other metadata tools. Version 2.1 (in development) will include the ability to export records in Extensible Markup Language (XML). In 2004 the I&M Program recommends that all relevant datasets at I&M parks and networks be cataloged in at least simple Dataset Catalog format. IRMA has replaced this Dataset Catalog system.

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. IRMA and ArcCataloghave replaced SMSS for FGDC Biological Profile and other geospatial metadata creation.

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. There are editors to enter metadata, a storage schema, and property sheets to view the data. With ArcCatalog users can view GIS data holdings, preview geographic information, view, edit, work with tables. Metadata within ArcCatalog is stored exclusively as Extensible Markup Language (XML) files.

IRMA, ArcCatalog and the NPS Datastore have replaced SMMS and Dataset Catalog as of 2014. 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 NPS Data Store comprises a web based system to integrate both data dissemination and metadata maintenance.

7.3. Metadata Process/Workflow

Step 1. Identify Relevant Data Sets and Compile Pertinent Metadata
Data utilized by the Network can be grouped, at least initially, in three broad categories based on origin. These categories include legacy datasets (primarily data collected prior to the inception of the Network), non-programmatic datasets (ongoing data collection efforts conducted by entities outside the direct responsibility of the Network), and programmatic datasets (data collected by the Network).

<u>Legacy Data</u>
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. A detailed amount of documentation may not be possible, however the data and all supporting documentation should be assembled and

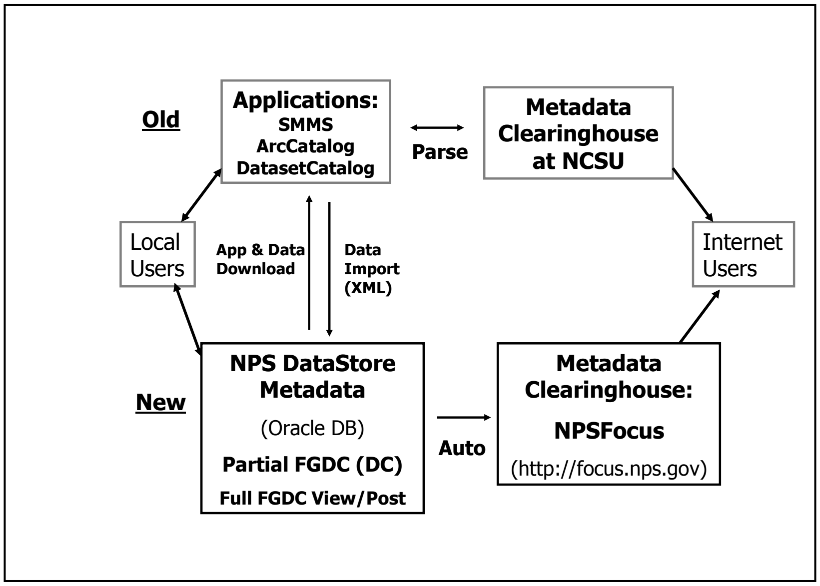


Figure 7.1. NPS integrated metadata system.

reviewed. Many legacy data sets will need to be converted to a standard database format for incorporation and future analyses. Data entry, validation and verification procedures will follow those contained within this DMP. A processing and revision log will be maintained with the dataset for capture of pertinent metadata.

<u>Non-programmatic Data</u>

Networks and/or prototypes are not the only entities gathering relevant inventory and/or monitoring data pertinent to park management. The Network and Prototype will make every effort to capture and assimilate all relevant data. When documenting non-programmatic datasets, the outside entity will be contacted and a request will be made for available metadata and/or a metadata interview will be conducted. As with legacy data, data files may need to be converted to a standard database format for analysis. Data entry, validation and verification procedures will follow those contained within this DMP. A processing and revision log will be maintained with the dataset, for capture of additional metadata.

<u>Programmatic Data</u>

For new projects, metadata development will begin by interviewing the principal investigator to explain what will be needed to properly document the data. In most instances, this will include completion of a basic metadata survey for inclusion in the data manager's project file, as well as submission of supporting documentation (proposal, SOPs, etc.). In addition, a database structure will be developed by, or in close consultation with, the data manager, to ensure compliance with the principles and procedures contained within this DMP. Updates and revisions to the metadata will be conducted in tandem with data submissions.

Step 2. Create IRMA Record

The Network will develop a simple IRMA record for relevant spatial and non-spatial data. This approach provides brief metadata for all Network data holdings in a searchable, centralized location. Managers can identify and prioritize datasets for which formal metadata will be developed. Prioritization of datasets for further documentation will be based upon current or anticipated future use. Datasets, which will be used repeatedly in analysis or with high probability for data sharing, will be addressed first. All GIS layers will be documented with applicable FGDC metadata standards.

Step 3. Select Metadata Tool and Complete Record

The Network will use IRMA and stand-alone Metadata Tools and Editor as primary metadata tools for documenting non-spatial datasets; ArcCatalog will be used for documenting spatial datasets.

Step 4. Make Information Available

At a minimum, metadata and associated data will be submitted to NR-GIS Metadata and Data Store. This will be accomplished using the recommended desktop applications. Additionally, information on data holdings should be conveyed in a meaningful manner for park resource managers, researchers, and others with a potential interest/stake in park management and/or research endeavors.

Similar to metadata creation, the mechanisms and formats for accomplishing this are varied. In addition to FGDC text files, Dataset Catalog can output a list of all records, single record reports, and/or a data dictionary report. Dependant upon the target audience these standardized outputs can be useful in conveying information on program data holdings and summaries of database structures. Obviously, customized queries and reports can also be generated. Other standardized outputs include ArcCatalog stylesheets. The NPS Metadata ArcCatalog Extension contains custom stylesheets, which can be invoked from the metadata toolbar. These can be used to depict pertinent details in a more coherent format than standard metadata outputs.

8. Data Analysis and Reporting

8.1. Data Analysis

Data Analysis. Appropriate analysis of monitoring data is directly linked to the monitoring objectives, the sampling design, and management uses of the data. Analysis methods need to be considered when the objectives are first identified and the sampling design is selected, rather than after the data are collected. Failure to adequately consider analysis methods during monitoring program development could result in use of sampling designs that are either inadequate or too complex to meet the monitoring objectives.

It is important that the data analysis Standard Operating Procedures (SOPs) for each Vital Sign, ensure that the sampling designs and analysis methods we use meet Network monitoring objectives. In addition, making the connection between the analyses that are produced and the decisions that are faced by park managers is critical. Interpretation of these analyses will emphasize the use of simple, graphical displays and visual summaries, so that the implications of monitoring results to management decision making are readily apparent.

With the exception of water quality, the protocols and accompanying monitoring designs for APHN Vital Signs are in the early stages of development. Therefore, decisions concerning the monitoring designs and analyses appropriate for addressing those monitoring objectives have yet to be made. The specific objectives of the Network water quality monitoring program are to:

- Determine seasonal and annual variability in concentrations of bacteria, nutrients, sediment, selected trace metals, and physical parameters at major streams and rivers within BISO, BLRI and OBRI; and, to
- Determine if concentrations of bacteria, nutrients, sediment, selected trace elements, and physical parameters are changing over time and estimate the rates of change.

For water quality monitoring, summary statistics (minimum, maximum, median, mean) and annual time series graphs (concentration vs. time) will be generated by site and compared with state use classifications. Boxplots of each water-quality parameter by site will be produced, to compare and contrast data visually. Annual comparisons will also be compared to historic data.

Trends will be calculated using non-parametric analyses after about 5 years on sites where monthly data is collected and after about 8 years for sites where bimonthly data is collected. Loads and yields can be calculated for sites with continuous discharge stations when there is an adequate period of flow and water quality data (about 3 to 5 years of data).

8.2 Reporting of Monitoring Information

As part of the NPS effort to "improve park management through greater reliance on scientific knowledge", a primary purpose of the Inventory and Monitoring Program is to develop, organize, and make available natural resource data and to contribute to the NPS institutional knowledge by transforming data into information through analysis, synthesis and modeling. The NPS 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, where the key role of the I&M Program is to provide park managers and interpreters 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, and as stated by the National Park Advisory Board, the findings "must be communicated to the public, for it is the broader public that will decide the fate of these resources".

Toward this end, the APHN is developing strategies for effectively sharing information with Network parks, scientists, cooperators, adjacent land managers and other potential collaborators. Information from Network data mining and inventory projects is being entered into the NPS master web-based databases, and monitoring data will be added as that part of the program gets underway (NPSpecies, NPS Datstore, and the Natural Resource Database Template). The Network has developed a website (http://science.nature.nps.gov/im/units/aphn/index.cfm), as well as network newsletters to disseminate information and updates to parks and the public. Network staff and cooperators present posters and give presentations at professional meetings, write papers for publication in technical journals and other publications, and write popular articles for park brochures and newspapers. We are currently working with the Southern Appalachian CESU, the Appalachian Highlands Science Learning Center and park interpreters to more effectively interpret inventory and monitoring results to the parks and the public. We are also exploring the possibility of sharing information from network projects with the Southern Appalachian Information Node of the national Biological Infrastructure (NBII). NBII is a collaborative effort that links information, high-quality biological databases, and analytical tools with information consumers such as government agencies, academic institutions, non-government organizations, and private industry.

Table 7.1 summarizes the content of reports and other information products of the I&M monitoring effort, intended audience, reporting schedule, and responsible entities for each. Table 7.2 is a schedule of formal reports and publications related to the I&M program.

Table 7.1. Reporting of Monitoring Data
(NOTE: ALL PUBLIC NEWS RELEASES WILL BE COORDINATED THROUGH THE PARKS' RESOURCE MANAGERS AND SUPERINTENDENTS PRIOR TO RELEASE)

Monitoring Protocol	Information Content	Target Audience & format	Responsible Person	Schedule
Water Quality	Summary of baseline, trends in pH, temperature, DO, specific conductivity, major ions, aquatic macro-invertebrate species diversity & numbers	-Park Managers, cooperators (executive summaries; briefings)	APHN aquatic ecologist, data manager	At least annually; eventfully
		- Public (news releases, brochures, network newsletters, website)	" "	Eventfully
		- Presentations at professional meetings; journal articles; informal presentations	" "	" "
Air Quality	Summary of baseline, trends in ozone levels, deciviews (visibility), nitrate and sulfate deposition, particulates	-Park Managers, cooperators (executive summaries; briefings)	APHN ecologist, data manager	Annually
		- Public (news releases, brochures, network newsletters, website	" "	Eventfully, or at least annually
Rare fish	Distribution and numbers for the year; trends	-Park Managers, cooperators (executive summaries; briefings)	APHN ecologist, hydrologist, data manager	At least annually, or more often if noteworthy events are observed.
		- Public, without specific locations (website, newsletters, news releases) SENSITIVE INFORMATION – no site-specific details		
Rare mussels	Distribution and numbers for the year; observed age structure, evidence of reproduction, trends	-Park Managers, cooperators (executive summaries; briefings)	APHN ecologist, hydrologist, data manager	Annually, or more often if noteworthy events are observed (e.g., first reproduction of reintroduced endangered mussels)
		- Public, without specific locations (website, newsletters, news releases); SENSITIVE INFORMATION – no site-specific details		
Rare plants & cobblebars	Species cover, relative abundance, trends	-Park Managers, cooperators (executive summaries; briefings)	APHN ecologist, data manager	Annually
		-Public (network newsletters, website) – SENSITIVE INFORMATION - no site-specific details		- Periodically, or eventfully
Weather	Annual rainfall, snowfall, temperatures	-Park Managers, cooperators (executive summaries), other scientists working in the parks,	APHN data manager	Annually

Monitoring Protocol	Information Content	Target Audience & format	Responsible Person	Schedule
	(average, extreme highs, lows), storm frequency, frost dates	interpreters		
Forest vegetation structure, composition, landscape pattern	Percent cover by dominant forest or other vegetation types, changes and trends	-Park Managers, cooperators (executive summaries; briefings)	APHN ecologist, data manager	Annually
		- Public (news releases, brochures, network newsletters, website)	" "	Periodically, when noteworthy trends are documented
Plant poaching	Changes in species composition in vulnerable communities, disappearance of target species (trilliums, orchids, galax, bloodroot, etc.)	-Park Managers, Law Enforcement, cooperators (executive summaries; briefings with key details on location, seasonality of poaching events)	APHN ecologist, data manager	Eventfully (fresh evidence of poaching will be reported immediately); Annual summaries of data
		- Public (news releases, network newsletters) with prior review and clearance from parks Law Enforcement, Resource Managers; no details on specific site locations	" "	Periodically, when situations warrant
Land use patterns	Changes and rates of change in development and land use patterns adjacent to and upstream of the parks; effects on park resources	-Park Managers, cooperators (executive summaries; briefings)	APHN ecologist, data manager	Initial baseline report, repeated every 5 years
		-public release – prior coordination is required with parks' superintendents through Resource Management before any public release of this type of information SENSITIVE INFORMATION	APHN Coordinator	To be determined.
Landscape pattern	Changes in composition and structure of dominant vegetation types (loss of dominant forest trees to disease, insect pests, etc.)	-Park Managers, cooperators (executive summaries; briefings)	APHN ecologist, data manager	Initial baseline report, repeated every 5-10 years
		-Public (newsletters, website) -Professional presentations, journal articles		Periodically, or eventfully

9. Data Dissemination

9.1 Data Ownership

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

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

- 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.

Provisions from OMB Circular A-110 are included in all standard Network agreements and contracts to ensure that U.S. Government regulations regarding ownership and use of information produced as a result of the agreement are adhered to. In addition, every cooperative agreement or contract must include a list of deliverables and products. Details on formatting and media types that will be required for final submission must also be included. Examples of deliverables include, but are not limited to, field notebooks, photographs (hardcopy and digital), specimens, raw data, and reports. 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 APHN 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

The main mechanism for distribution of APHN inventory and monitoring data to the broader public will be the internet. 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. APHN will use the following applications and repositories to distribute data, formal and informal reports and publications:

- **NPS Datastore**–a master web-based database housing natural resource bibliographic data for I&M Program parks

- **NPSpecies**–a master web-based database to store, manage and disseminate scientific information on the biodiversity of all organisms in all National Park units

- **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

- **Appalachian Highlands I&M Network Website**–provides detailed information about the network and the 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. http://science.nature.nps.gov/im/units/aphn/index.cfm)

Table 9.1 Data types that will be uploaded to web-based applications.

Web Application Name	Data types available at site
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 inventories, monitoring, and projects
NR-GIS Metadata and Data Store	Metadata (spatial and non-spatial and products)
APHN Website	Reports, publications, newsletters and metadata

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 Network will disseminate 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. NPS Focus has been released as an Intranet version only (http://focus.nps.gov/), a public version is projected in the near future.

9.3 Handling Sensitive Data

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.

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)).

The following guidance for determining whether information should be protected is suggested in the draft Director's Order #66 (the final guidance will be contained in 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 archiving of those records to avoid their unintentional release

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.

Classification of sensitive I&M data will be the responsibility of park superintendents through park and Network staff. Network staff will work with park staff to classify sensitive data on a case by case basis. APHN 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 responsible for making principal investigators aware of sensitive resources. 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.

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 both the NPSpecies and NPS Datastore databases are discussed at the following websites:

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

10. Data Maintenance, Storage and Archiving

This chapter describes procedures for the long-term management and maintenance of digital data, documents, and objects that result from APHN projects and activities. The overall goals of these procedures are:

- to avoid the loss of information over time
- to ensure that information can be easily obtained, shared, and properly interpreted by a broad range of users.

Effective long-term data maintenance is inseparable from proper data documentation, and an essential part of any archive is accompanying explanatory materials (Olson and McCord 1998). This chapter will refer to, and elaborate on metadata standards and dataset documentation procedures that are more fully explained earlier in this document and Chapter VII (Data Documentation) of this plan.

10.1 Digital Data Maintenance

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

Following the lead of the National Park Service and the national I&M program, APHN has adopted MS-Access as its database standard and ArcGIS as its spatial data management standard. APHN will remain current and compatible with NPS or national I&M version standards for these software programs.

The changing and elimination of specific programs/software is a significant cause of information loss. Data can quickly become inaccessible to users if stored in out-of-date software programs or on outmoded media. Maintaining digital files involves managing the ever-changing associated infrastructure of hardware, software, file formats, and storage media. As software and hardware evolve, data sets 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).

Any data set for which data entry or updates is still occurring will be stored under the "Active Data Projects" directory on the APHN server. The "Project Archive" directory is reserved for data sets that will no longer change.

Short-term data sets
For short-term data sets created or managed by APHN, upon project finalization a set of ascii tab-delimited text files will be created for each data table comprising the data set. These files will be accompanied by a readme.txt file that explains the contents of each file, file relationships, and field definitions. The ascii files are in addition to the native software structure of the dataset. Creating the text files will help ensure the data are usable in a wide range of applications or platforms.

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

Monitoring projects will also have variable long-term data archiving requirements. Raw data sets that are later manipulated or synthesized will need archiving in the original form. Modifications to protocols will typically require complete data sets to be archived before modifications are implemented. With frequent changes to the monitoring project, it is necessary to preserve interim data sets (data "milestones") over the long term. Data archiving requirements for ongoing projects will be detailed in the data management SOPs for each monitoring project.

Version control
Previous versions of databases will be saved in their native format and archived in addition to the current version. Documentation of version updates and associated details will be part of the archive metadata document, and revision information and history will be included in tables within the database files themselves as well. File names of the archived revisions should clearly indicate the revision number or date.

Spatial data
Spatial data sets that are essential to APHN (ie, base layers) will be maintained in a format that remains fully-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 ArcGIS format if they are no longer supported. At this time there is no practical way to save GIS data in a software or platform-independent format.

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

10.2 Storage and Archiving Procedures – digital data

Directory structure for electronic archives
Digital data need to be stored in a repository that ensures both security and ready access to the data in perpetuity. APHN relies on the Blue Ridge Parkway for IT support. As of 2007, APHN has an on-site 450 GB server with a level-5 RAID (redundant array of independent disks) for shared data storage, combined with a nightly schedule of full backup to replicate server. A periodic offsite tape backup will run and be saved to an offsite location to recover from fire or other catastrophes. Offsite storage through secure IM and individual user folders is available on Blue Ridge Parkway server.

Directory structure for individual projects
The organization and naming of folders and files should be intuitive to users unfamiliar with a specific project. A standardized structure may not be practical however, all project archives will include several to most of the following elements:

- administrative documents such as agreements, correspondence, research permits
- programmatic documents including protocols, procedures, supporting documents
- interim data sets or "milestones"
- data sets reformatted or manipulated by APHN
- data sets original form – ascii
- conceptual or statistical models used for data interpretation
- final report
- readme files -- includes an explanation of directory contents, project metadata (including a dataset catalog report), and version documentation

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

Backup Procedures for digital data

The BLRI computer network that contains APHN data has a RAID 5 system with tape backup performed every night. The backup period for data lasts roughly 30 days for retrieving data. Shadow copy software is used for the Windows 2003 server for immediately recovering files lost or corrupted in the middle of editing.

Backups of data that reside on the personal computers of staff are the responsibility of each staff member. The preferred method is for staff to regularly copy important files onto a personal directory on the APHN shared network, where daily backups are performed.

10.3. Storage and Archiving Procedures – documents and objects

This section applies to documents such as final reports prepared by staff or contractors, program administrative documents, contracts and agreements, memoranda of agreement, and other documents related to APHN administration, activities and projects. This section also applies to physical items such as natural history specimens, photographs, or audio tapes. In most instances these documents and objects are essential companions to the digital data archives described above.

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

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

Section N of Appendix B, which provides guidelines on natural resource-related records (including, specifically, the results of Inventory and Monitoring Programs), indicates that all natural resource records are considered “permanent,” that is, 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.

Documents

All paper documents managed or produced by the APHN will be housed in three locations:

1. APHN central files, Asheville, NC (BLRI computer network)

These files contain project files, administrative documents and non-record copies of documents that are archived at an off-site facility (see item 2, below). Examples include: meeting minutes, correspondence, memoranda of understanding, contracts and agreements, research permits, interim and selected final reports produced by the program.

2. Network park museums. Network park archives will be used for original documents and associated materials produced by the network (e.g. photographs, field notes, permits) that are a high priority to maintain under archival conditions. Examples include: original inventory reports and accompanying slides and maps; original vegetation mapping reports; APHN monitoring reports. Copies of these reports will be maintained in the APHN central files, and all will have an electronic equivalent (e.g. pdf) for distribution or reproduction.

For all materials submitted for archiving, APHN will assist with cataloging, and will provide essential cataloging information such as the scope of content, project purpose, and range of years, to facilitate ANCS+ record creation and accession. APHN will also ensure that materials are presented using archival-quality materials (e.g. acid-free paper and folders, polypropylene or polyethylene slide pages).

3. Network park central files. High-quality copies of park-related documents resulting from APHN projects, along with electronic versions, will be provided 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. APHN will not manage documents at the park level.

Specimens

Specimens collected under the direction of the APHN will be provided to the network repository/museum in which they were collected for curation, or to a repository approved by a park (where the specimens are considered on loan). APHN will assist with cataloging, and will provide park curators with associated data required for cataloging each specimen. This data will be in comma-delimited format (.csv) format for automated uploading into ANCS+. Data provided to non-NPS curators will be in Excel format.

Photographs

Archivists have been reluctant to fully embrace digital photography and some have expressed concern that, with the accelerating rate of technological change, documentary heritage is in danger of being lost in the information age (Cox 2000).

Slides are labeled using indelible pigment ink, or using laser-printed archival-quality slide labels. Slide labels will 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 are stored in polypropylene slide sleeves at the APHN office. In addition, all slides are 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 are stored in individual polypropylene sleeves and within archival boxes. Each photo is labeled on the back, using archival-quality labels that are either laser-printed or hand-labeled, with the same information elements required for slides. If a contractor is submitting photographs, corresponding TIFF files must also be submitted

Every image, regardless of format, has an entry into the APHN Photo Database (see SOP TBA), where attributes such as electronic file name, keywords, project, photo description, photographer, date, and location are catalogued. All photo files and the associated photo database are housed on the archive portion of the APHN server (see Figure 1).

Role of curators in storage and archiving procedures

Curators for parks within APHN are an ongoing source of expertise, advice, and guidance on archiving and curatorial issues. Project managers should involve park curators when projects are in the planning stage, to ensure specimen curation and document archiving is considered, and that any associated expenses are included in project budgets.

11. Literature Cited

Cox ,I. J., M.L. Miller, J.A. Bloom, NEC Research Institute. The International Conference on Information Technology: Coding and Computing (ITCC'00) March 27 - 29, 2000. Las Vegas, Nevada p. 6 Watermarking applications and their properties.

Chapal, Scott E., and Don Edwards. "Automated smoothing techniques for visualization and quality control of long-term environmental data." *Environmental Information Management and Analysis: Ecosystem to Global Scales*. Eds., William K. Michener, James W. Brunt, and Susan G. Stafford. Bristol, PA: Taylor & Francis Inc., 1994. 141-158.

Landis, E.B. and C.J. Palmer. 2002 Lake Mead NRA Resource Management Division, Quality System Management Plan for Enviromental Data Collection Projects http://science.nature.nps.gov/im/datamgmt/dmw032004/docs/11RMQSMP_2.pdf

Michener, W.K. and Brunt, J.W. Ecological Data: Design Management and Processing, Blackwell Science, 2000

Tessler, Steven & Joe Gregson. 1997. Draft Data Management Protocol. National Park Service. Document available at: http://www.nature.nps.gov/im/dmproto/joe40001.htm

Standard Operating Procedure 1

Data Dictionary

1. Purpose

To standardize the contents of data dictionaries developed to accompany and document all data sets or databases. The data dictionary serves as metadata for each non-spatial data set or database, describing the creator, contents, data elements and units, and any derived or calculated fields. In addition, the data dictionary may also serve as guidance to data entry personnel to understand the entire contents of the database.

2. Scope and Applicability

The standards described in this procedure pertain to all non-spatial data sets or data bases that are newly developed or legacy data.

3. Procedures and General Requirements

- The data dictionary should be developed along with the database or data set, by its creator, and should stay with the database at all times. By specifying the contents and data elements of the database, there is less chance of error by the data collectors and data entry staff, because everything is described in advance.

- The introductory section should include: (1) Name, project identification number and date, (2) List of parks included in the database, (3) Contact information for the project lead, (4) Database software used and data filename, (5) Short abstract of the database contents that includes the purpose and general description, (6) References or citations that accompany the database, and (7) List of tables used in the database (8) Definitions of Acronyms used in describing data, fields, agencies.

- Following the introduction, each data element in the database will be described. Data elements will be organized under table headings, as they exist in the database itself. This description will include the field name as used in the database, a brief description of the data included in the field, standard format for data entry, units (if necessary), and any standard designations – such as those used for no data (e.g. ND, -9999).

- For any calculated fields or manipulated/transformed data, all formulas or transformation methods will be described and included in the description of the data element.

- The data dictionary will be an MS Word or ASCII text file that is always stored along with the database itself. The dictionary file name will match that of the database (e.g. database = filename.mdb, dictionary = filename_dic.txt) for easy reference.

- If the database is migrated to newer software versions, edited, or otherwise changed; these changes should also be documented in the data dictionary. Remember, the purpose of the data dictionary is to provide a user who had no involvement in the study, information to make use of the data in the database - appropriately and without making assumptions regarding the content.

4. **Responsibilities**
    - Project Scientist or Data Manager: It is the responsibility of the database creator to also develop the data dictionary for the database. Keep in mind that this is metadata for the database and that the data dictionary is necessary to ensure the usability of the data beyond the life of the project (or the tenure of project personnel). Do not assume that anything is self-explanatory, document the contents.

    - Data Entry Personnel: Become familiar with the contents of the data dictionary, prior to beginning data entry. The data dictionary should answer the majority of questions you might have regarding the contents of the database. If any of the data elements are still unclear after reading the data dictionary, the dictionary should be revised to provide better documentation for current and future users. Work with the project scientist or data manager to ensure the data dictionary reflects the necessary changes.

    - Database Users: Read the data dictionary before using the database to familiarize yourself with the contents. In addition, read any additional reference material that accompanies the database. All datasets have limits to their scope and applicability and you should know the purpose and limits of the data, before you use the data.

Standard Operating Procedure 2

Database Specifications

1. **Purpose**
   To standardize the overall design and functioning of databases used for I&M Program projects. This document should be used in association with the data dictionary for the specific database to understand how the database works.

2. **Scope and Applicability**
   The standards described in this document are relevant to all I&M Program databases. While the specific fields included in the database may differ on a project by project basis, the general structure, organization and functioning of the database should be the same as the information presented here. Therefore, a separate data dictionary will be developed for each database to define all data elements; the data dictionary should be used in association with this document as a starting point and users guide to I&M Program databases.

3. **Procedures and General Requirements**
   - All APHN I&M databases are based upon the concepts implemented in the Natural Resource Database Template (NRDT), available online at: http://science.nature.nps.gov/im/apps/template/index.htm. The NRDT is a relational database in MS Access that may be used as a stand-alone database or linked to ArcView GIS software using other tools from the I&M Program such as ArcView to Access (http://www.nps.gov/akso/gis/).

   - The NRDT is built around two core tables (tblEvents and tblLocations) which contain all of the data pertinent to sampling events (date, time, conditions) and the location where the sampling event occurred. These two tables are standard to every database developed by the APHN I&M Program.

   - The following sections go through an implementation of the NRDT concept, using a plant inventory database. This should provide users a general example of how databases in the APHN are developed, their content, and potential reporting tools.

   Data Entry
   Data entry is initiated through the database switchboard, beginning by selecting the park where the sampling occurred. After setting the park, data entry personnel may return to the switchboard and progress to the main data entry screen, enter or edit field notes. Location, event, observer, and all sampling information is entered here, through the use of pick lists or command buttons to supplementary data entry screens. The species pick list was generated from the current species lists from APHN parks, which are based upon ITIS (Integrated Taxonomic Information System, from the USDA).

Data Verification
Once data entry is complete for a particular set of field notes, the next step in the process is to verify that the data was entered correctly. Several elements of data verification are built directly into the database. By selecting the "Verify taxonomy and Species Counts" button on the main switchboard, data entry personnel are moved to the data verification screen.

Voucher Specimens
All species recorded as having photo or specimen vouchers can be processed using the "Process Specimens (vouchers)" or the "Process Photo Vouchers" buttons on the main switchboard. Once again, the event/location of interest is selected and all supporting voucher information (Species ID, specimen condition, catalog number, accession number etc.) may be entered. Please check the Print Label check box if a voucher specimen label will be required.

Data Maintenance and Reports
If species codes or other taxonomic information changes, it is possible to update the species codes through the data maintenance screen. The Reports screen allows the user to: (1) Generate Current Species Lists, (2) Preview Herbarium Labels, (3) Print Herbarium Labels, and (4) Generate ANCS+ and Upload Spreadsheet – which combines the fields necessary for ANCS+ records into a spreadsheet which may be uploaded into ANCS+.

4. Responsibilities

- <u>Project Scientist:</u> The project scientist will work with the data manager at the outset of the project to develop a database that supports both project (protocol) and programmatic needs. Please review the database specifications and the data dictionary that were developed specifically for your project, prior to starting field work, to ensure that the database will meet project needs and will capture all raw data. In addition, please use the database as provided or discuss necessary changes with the data manager. Changes to the database may result in problems with the database itself, or incompatibilities with other I&M Program data management needs. Work with data entry personnel (especially those not involved in data collection) to ensure they understand the database and your field data sheets before entering any data.

- <u>Data or Project Manager:</u> The data manager will work with the project scientist to develop a database that meets both project and I&M Program needs. Database development is typically a back-and-forth process that results in an implementation that works for everyone involved. Upon completion of the database, the data manager will develop a data dictionary that described all data fields in the database, to ensure the greatest level of understanding by all database users and data entry personnel. Finally, the data manager will also work with data entry personnel to ensure that data are being entered properly and that quality assurance methods are being followed.

- Data Entry Personnel: Please review the data dictionary and the database specifications document before attempting to use the database you have been provided. Become familiar with the project field data sheet before beginning data entry.

Standard Operating Procedure 3
GIS Specifications

1. **Purpose**
To standardize data formats, coordinate systems, and spatial scale and resolution for all spatial data collected through the NPS Inventory and Monitoring Program or NPS Natural Resources programs.

2. **Scope and Applicability**
This document was created from the NPS GIS Specifications document that is referred to below in section 3, Reference Documents. This document applies to all spatial data and GIS products or deliverables. The use of GPS in the collection of spatial data is covered in a separate, GPS Specifications SOP document.

This document assumes the user is already familiar with GIS and does not provide any information or training in GIS itself.

3. **Reference Documents**
   - National Park Service. 2003. GIS (Geographic Information System) data specifications for resource mapping, inventories and studies. National GIS and Inventory and Monitoring Program Website. 10 pages. This document can be found at: http://science.nature.nps.gov/im/gis/standards.htm

4. **Procedures and General Requirements**
Data Formats:
   - All data and related file names should adhere to 8.3 file naming conventions: names do not contain spaces or special characters. In addition, field names should not exceed 10 characters due to limitations in ArcINFO and dbase.

   - All vector data will be supplied as an ArcINFO coverage, ArcINFO interchange file (*.e00) or ArcView Shapefile that is compatible with the current version of ArcGIS. All coverages should be created as double precision data sets. If the data was originally created in single precision, it should be converted to double precision before submitting the final product. Both coverages and shapefiles will also have defined projections (and in the case of shapefiles, an accompanying *.prj file) and FGDC compliant metadata.

   - All raster data will be supplied as an ArcINFO GRID and ArcINFO interchange file (*.e00) that is compatible with the current version of ArcGIS. Other possible raster file formats include: .BIL and MrSID files. However, the project or data manager must approve any deviations from the preferred standard (GRID), described above. All raster files will also be submitted with FGDC compliant metadata.

a. Most all digital imagery (e.g. aerial photos) will be supplied as tagged image file format (.TIFF) files with the proper header file for geo-referencing purposes. In special circumstances (e.g. large image files) other spatial data formats (ERDAS Imagine, MrSID etc.) may be acceptable. All digital imagery will be properly geo-referenced, accompanied by appropriate header or world files, and submitted with FGDC compliant metadata.

Coordinate Systems:

b. All spatial data collected or submitted for NPS programs will be geo-referenced and provided in the standard projections, noted below. The steps used to get the data in the proper projection must be documented in the accompanying FGDC compliant metadata. Prior approval from the project or data manager must be received to deviate from these specifications.

Park	Projection	Datum	Spheroid	False Easting	False Northing	Units
All	Geographic	NAD83	GRS 1980			Decimal Degrees
BLRI	UTM, Zone 17	NAD83	GRS 1980	500,000	0	Meters
BISO	UTM, Zone 16	NAD83	GRS 1980	500,000	0	Meters
OBRI	UTM, Zone 16	NAD83	GRS 1980	500,000	0	Meters

Spatial Scale and Resolution:

c. Specific scale and spatial resolution requirements for image data will be specified in the contract or cooperative agreement for each project. For example, vegetation classification projects under the NPS/USGS vegetation classification and mapping program will use 1:12,000 color infrared aerial photographs (or better) with 60% overlap (endlap) and 30% sidelap.

d. In general, new data should be compiled with an accuracy level better than US National Map Accuracy Standards (NMAS) for a 1:24,000 scale map; unless other requirements exist (which would then be included on a protocol-specific basis). However, all spatial data collected will be analyzed for their spatial accuracy and will meet or exceed NMAS for the appropriate scale (for more information, please see http://mapping.usgs.gov/standards).

e. Any calculations done with location data should be done at double precision with the results rounded or truncated to the appropriate propagated error limits. All calculations and processing completed on the spatial data shall be reported in the FGDC compliant metadata that accompanies the GIS layer.

f. For maps on publication scales larger than 1:20,000, not more than 10 percent of the points tested shall be in error by more than 1/30 inch, measured on the publication scale; for maps on publication scales of 1:20,000 or smaller, 1/50 inch. These limits of accuracy shall apply to positions of well-defined points only. Well-defined points are those that are easily visible or recoverable on the ground: monuments or markers, such as benchmarks and property boundary monuments; intersections of roads and railroads; and corners of large buildings or structures (or center points of small buildings).

g. The following table provides the allowable horizontal error for some common scales:

Scale	Allowable Error
1:40,000	33.8 meters (111 feet)
1:31,680	16.1 meters (53 feet)
1:24,000	12.2 meters (40 feet)
1:20,000	10.1 meters (33 feet)
1:12,000	6.1 meters (20 feet)
1:9,600	4.9 meters (16 feet)
1:4,800	2.4 meters (8 feet)
1:2,400	1.2 meters (4 feet)
1:1,200	0.6 meters (2 feet)

Standard Operating Procedure 4

GPS Specifications

1. **Purpose**
   To describe the procedures used to collect GPS data of acceptable quality in the field. This document also provides information on instrument settings, field operations, and standards in the recording of positional data.

2. **Scope and Applicability**
   GPS (Global Positioning System) is currently a constellation of 25 Department of Defense satellites that orbit the earth approximately every 12 hours, emitting signals to Earth at precisely the same time. The position and time information transmitted by these satellites is used by a GPS receiver to trilaterate a location coordinate on the earth using three or more satellites.

   The standards described in this procedure pertain to all data collected with hand-held or backpack GPS receivers. This document focuses upon field operations and instrument settings to be used when collecting GPS data, as well as including some recommended standards for the collection of positional data. This document is not intended to serve as a training manual for those new to GPS data collection. For more detailed information regarding the collection of GPS data, please consult the documents below in Section 3, Reference Documents.

3. **Reference Documents**
   - A reference manual for creating quality GIS data with GPS technology is available at
     http://www.nps.gov/gis/gps/gps4gis/

4. **Procedures and General Requirements**
   - Garmin Unit Settings:
     (1) Estimated Horizontal Error (EHE) should be less than or equal to 12 meters. This will keep you just within National Map Accuracy Standards for a 1:24,000 scale map – which is the maximum acceptable.
     (2) Use a minimum of 4 satellites (3D) for every position.
     (3) For point data, collect 90-120 positions at 1-2 second intervals and averaged.
     (4) For line or polygon data, use a 2-5 second interval for walking and road driving, depending upon the road type and speed of the vehicle. Force (e.g. wait for) a position at each corner and use a minimum of 3 positions to define any curve or change in direction.

   - Trimble Pathfinder Unit Settings:
     (1) Positional dilution of precision (PDOP) should be less than or equal to 6, however, start with a maximum PDOP setting of 4 and shift to 5 if data

collection is not successful. This will allow you to meet National Map Accuracy Standards for a 1:5,000 scale map.

(2) Use a minimum of 4 satellites (3D) for every position.
(3) Signal to Noise Ratio (SNR) should be less than or equal to 5.
(4) Set Elevation Mask to 15.
(5) Be sure to check the Antenna Height setting – which should be the typical height at which the antenna will be carried. If the antenna is attached to a pole, it must be located above the user's head and the antenna height setting should be the height at the top of the pole. Whenever possible, the antenna should be clear of any obstructions.
(6) For point data, collect a minimum of 30 positions at 1 second intervals and averaged.
(7) For line or polygon data, use a 2-5 second interval for walking and road driving, depending upon the road type and speed of the vehicle. Force (e.g. wait for) a position at each corner and use a minimum of 3 positions to define any curve or change in direction.

- <u>All GPS Units:</u>
  i. Always do differential corrections, either real-time or post-processed (real time differential correction is preferred).
  ii. Check the graphics data collection screen regularly to see if you are getting multi-path or other apparent distortions to the data. Garmin requires the user to monitor the screen and stop data collection during poor PDOP or SNR windows. Trimble receivers set to the appropriate mask (see above) will stop collecting data automatically.
  iii. Be aware of the possibility of multi-path interference and use offsets or other methods to keep the antenna away from building overhangs, tall fences or walls, and heavy canopy, whenever possible.
  iv. If maximum accuracy is required, it is important to sync the collection rate with the base station logging rate. Stations log anywhere from 1 to 30 second data. It is recommended that logging rates be in multiples of 1 or 5 for best differential corrections. Setting logging rates other than 1 and 5 may reduce the number of positions that are in sync with base data and reduce accuracy.
  v. Map all features in a single area on a single day or on consecutive days.

- <u>Park Unit Data Standards:</u> All digital geospatial data should reference the coordinate system appropriate for its use and it should be documented in the metadata. When collecting GPS data in the field, the standard projection for the APHN and the Southeast Region is the Universal Transverse Mercator (UTM) with the following parameters:
  - *Projection:* Universal Transverse Mercator
  - *Datum:* North American Datum 1983
  - *Spheroid:* GRS 1980
  - *False Easting:* 500,000
  - *False Northing*: 0

  - *Units:* Meters

- Horizontal/Vertical Accuracy and Precision: All spatial data collected will be analyzed for spatial accuracy and will meet or exceed the National Map Accuracy Standards for the intended scale and use (for more information, please see http://mapping.usgs.gov/standards). Any calculations done with location data should be done at double precision with the results rounded or truncated to the appropriate propagated error limits. All calculations and processing completed on the spatial data will be reported in the metadata.

- Positional coordinate data should not be recorded in NAD27 in the field. Datum conversions should be done as an office, post-processing activity using software that utilized a full NADCON datum conversion in order to assure accuracy and precision

- To state again, real-time differential correction techniques should be employed whenever possible for efficiency and time savings. The distance between the base station and the remote GPS receiver should be kept to a minimum.

5. **Responsibilities**
   - Data Collectors: Be certain that you understand how to use the GPS unit, data logger, and any study-specific data collection forms prior to going into the field and collecting data. Ask questions or request training as needed. Plan your trips to the field in advance, and target times of the day when the most satellites will be overhead. Check the latest almanac at http://www.trimble.com/gpsdataresources.html) to get the best satellite availability for maximum field productivity. Spend time at the beginning of the project going over the data you are collecting and differentially correcting.

   - Data or Project Manager: Provide data collection personnel with training and instruction in the use of GPS units, data loggers, and any study-specific data collection forms prior to field collection. If possible, accompany all new data collectors in the field to ensure quality data collection procedures. Be available to answer questions from data collection personnel. If possible, examine data from initial data collection efforts, to catch systemic errors.

Standard Operating Procedure 5

Archiving Information(Inventories or Projects)

1. **Purpose**
   To describe procedures used to archive finished inventories or projects with completed products not likely subject to change. By having exact digital copies in three places and the documentation of the hard copies of the inventory after signing off on deliverables, the original data would be available in case of IT or natural disasters.

2. **Scope and Applicability**
   Mandatory for all inventories or projects administered or paid for by APHN budgets or directed by APHN staff. Other park projects and research as they become available should be archived in this similar manner. Not exclusive to research on NPS lands, adjacent Forest Service or State Land inventories could be archived in this similar manner if helpful in gaining more knowledge of encroaching or similar natural resources.

3. **Procedures and General Requirements**
   (1) Put copy of data in Archived Inventories, the stand-alone folder on APHN server under the proper park code. BISO/Herps for example would have specified folders for: Reports, Database, Photos, Photocopied field notes, etc. Create read-only files before copying into correct folder. The Archived Inventories folder has security where files cannot be edited from this folder.
   (2) Send archives to NPS Datastore(create metadata), NatureBIB(create metadata) and NPspecies (fill out correct Data2web form). The development of IRMA and Service Oriented Architecture(SOA) will change this process of archiving at the WASO level.
   (3) Place exact copy of .zip(zipped) data on BLRI server inpblris3, under IM directory folder Inventory, as read-only. This folder will have the same security where files cannot be edited from this folder.
   (4) Document locations of hard copy reports or CDs, ie(BLRI Botanist's Office, OBRI Biologist's Office, APHN Nora's Office or Bookshelf in Data Manager's Office.) Attach Readme file to the copy on our APHN server and BLRI IM folder detailing what was done for project or inventory archiving. If changes occur after inventory has wrapped up that significantly alters the data, a versioning date will be attached to new documents or data.

4. **Responsibilities**
   All network staff permanent and seasonal should be aware of this process and when tasked to archive finished inventories should reference this document.

Appendix 1

NPSpecies list of Editing Permissions as of 1/14/2014

For NPSpecies database there will be established fields that will be filled out similarly each time through the database. The APHN data manager will be the Point of Contact (POC) who will be in charge of who has NPspecies editing and entry capabilities for the network parks. At the Ft. Collins, WASO level any changes to the databases are documented to the user. Link to Users: Check List for Updated Users: https://irma.nps.gov/NPSpecies/User

PARK /NETWORK AFFILIATION	EMPLOYEE
APHN Data Manager(POC)	Patrick Flaherty
APHN Monitoring Coordinator	Robert Emmott
APHN Ecologist	Nora Murdock
BLRI Biologist	Bob Cherry
BLRI Botanist	Chris Ulrey
OBRI Natural Resource Chief	Rebecca Schapansky
OBRI Botanist	Justin Coffey
BLRI Biologist	Lillian McElrath
BLRI Natural Resource Chief	Bambi Teague
BISO GIS Specialist	Chad Harrold
APHN Hydrologist	Jim Hughes
AHSLC Science Coodinator	Paul Super

Appendix 2

APHN software and computer resources: 1/21/2014

Application/Function	Software Package/Version
Network Operating System	Microsoft 2010
Web Development	Macromedia Dreamweaver
Environment Operating System	Microsoft Windows NT 4.0 and Windows 2000
Database	Microsoft Access 2010
Word Processing	Microsoft Word 2010
Spreadsheet	Microsoft Excel 2010
Presentation	Microsoft PowerPoint 2010
Graphics	Adobe Photoshop
Digital Images	Adobe Photoshop, Thumbs Plus
Bibliographic	SMMS, Thumbs Plus
GIS Desktop	ArcGIS 10.2
Scheduling	BISON Connect
Statistics	Microsoft Excel 2010
Windows Utilities	Windows Imaging, Decompression tool
File Transfer	NERO CD burning software, Google Docs
Virus Protection	Symantec Anti-Virus Corporate Version

APHN electronic tools for data collection
Cannon EX Photoshot digital camera, Garmin 5 GPS unit, Office 2010