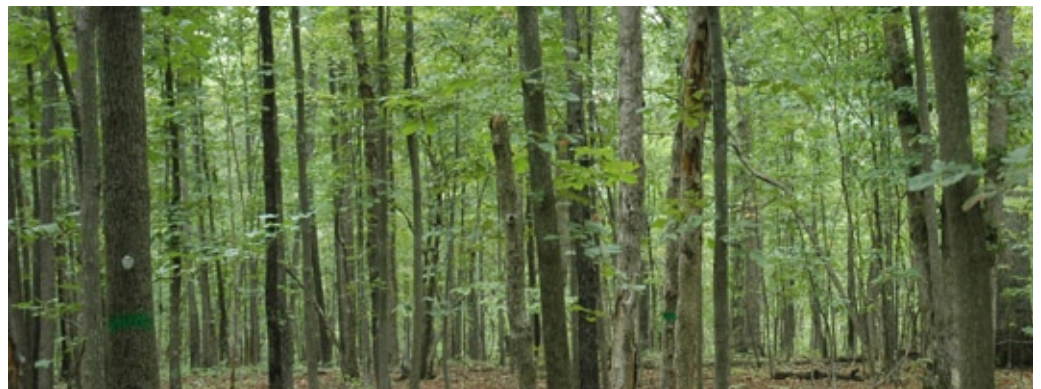




Mid-Atlantic Network

Data Management Plan

Natural Resource Report NPS/MIDN/NRR—2009/077



ON THE COVER

Fall foliage at Valley Forge National Historical Park; silver spotted skipper (*Epargyreus clarus*) on milkweed (*Asclepias spp.*) at Shenandoah National Park; forest at Fredericksburg and Spotsylvania National Military Park

Mid-Atlantic Network

Data Management Plan

Natural Resource Report NPS/MIDN/NRR—2009/077

Kristina K. Callahan and Sarah M. Wakamiya
Mid-Atlantic Network
National Park Service
120 Chatham Lane
Fredericksburg, VA 22405

January 2009

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

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

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

Views, statements, findings, conclusions, recommendations and data in this report are solely those of the author(s) and do not necessarily reflect views and policies of the U.S. Department of the Interior, NPS. Mention of trade names or commercial products does not constitute endorsement or recommendation for use by the National Park Service.

Printed copies of reports in these series may be produced in a limited quantity and they are only available as long as the supply lasts. This report is also available from the I&M Mid-Atlantic Network Web site (<http://science.nature.nps.gov/im/units/midn/>) and the Natural Resource Publications Management Web site (<http://www.nature.nps.gov/publications/NRPM>) on the Internet, or by sending a request to the address on the back cover.

Please cite this publication as:

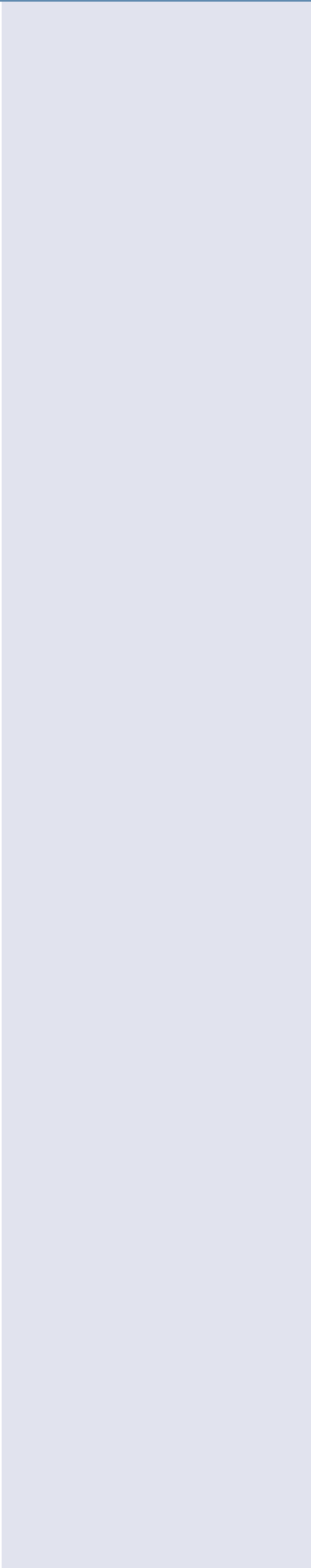
Callahan, K.K., and S. M. Wakamiya. 2009. Mid-Atlantic Network data management plan. Natural Resource Report NPS/MIDN/NRR—2009/077. National Park Service, Fort Collins, Colorado.

Change History

Version numbers will be incremented by a whole number (e.g., Version 1.3 to Version 2.0) when a change is made that significantly affects requirements or procedures. Version numbers will be incremented by decimals (e.g., Version 1.6 to Version 1.7) when there are minor modifications that do not affect requirements or procedures included in the plan.

The following revisions have occurred to this plan since January 31, 2009

Original Version #	Date of Revision	Revised By	Changes	Justification

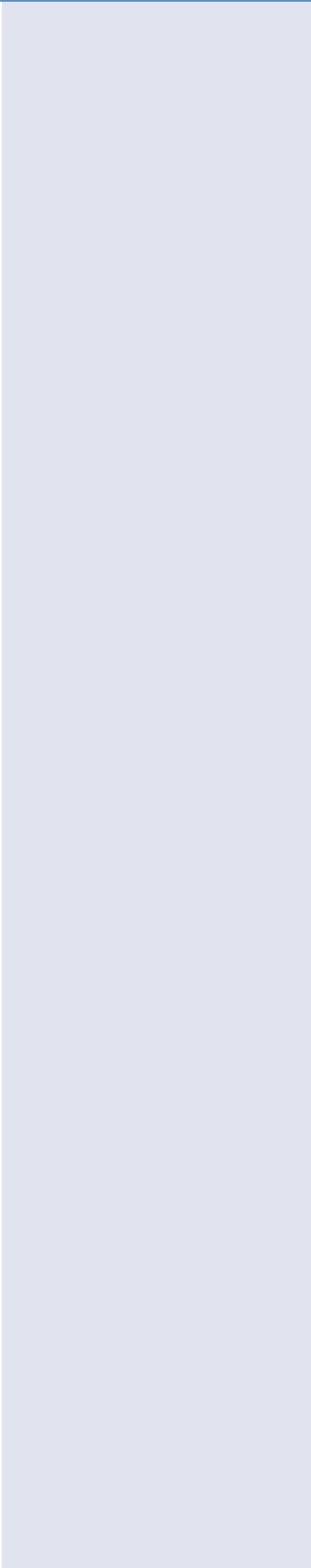


Contents

Executive Summary	xv
Chapter 1: Introduction and Background.....	1
1.1 Inventory and Monitoring Program.....	1
1.2 The Mid-Atlantic Network.....	1
1.3 Data Management Overview.....	1
1.4 MIDN Data Management Goals	2
1.5 National Data Management Plan.....	2
1.6 Intended Audience and Layout of the Data Management Plan	3
1.7 Revisions	3
Chapter 2: Infrastructure and Systems Architecture	5
2.1 Introduction	5
2.2 Objectives	5
2.3 Computer Infrastructure	5
2.4 Introducing Bob.....	6
2.5 Local Workstations.....	6
2.6 Data File Folder Structure	6
2.7 MIDN Websites	7
2.8 Standard Operating Procedures.....	7
Chapter 3: Project Management and the Data Life Cycle	9
3.1 Introduction	9
3.2 Objectives	9
3.3 Project Work Flow	9
3.3.1 Planning and approval	9
3.3.2 Design and testing.....	9
3.3.3 Implementation	10
3.3.4 Product integration	11
3.3.5 Evaluation and closure	11
3.4 Project Data Life Cycle	11
3.5 Standard Operating Procedures.....	11
Chapter 4: Data Management Roles and Responsibilities	13
4.1 Introduction	13
4.2 Objectives	13
4.3 National Level Staff and Coordination	13
4.4 Northeast Region Data Management Staff.....	13
4.5 Mid-Atlantic Network Staff	14
4.5.1 Coordinator/Ecologist	14
4.5.2 Data Manager/Biologist	14
4.5.3 Seasonal Biotechnicians	15
4.5.4 North Carolina State University Field Technical Support Center (NCSU FTSC)	15
4.6 Park Level Data Management Staff	15
4.7 Data Management Coordination	16
Chapter 5: Databases.....	17
5.1 Introduction	17

5.2	Objectives	17
5.3	Project Database Standards.....	17
5.4	Natural Resource Database Template	17
5.5	Natural Resource Monitoring Protocols Clearinghouse	17
5.6	Standard Operating Procedures.....	17
Chapter 6: Data Acquisition, Processing and Reporting		19
6.1	Introduction	19
6.2	Objectives	19
6.3	Types of Data	19
6.4	Mid-Atlantic Network Data	19
6.4.1	Inventory Projects	19
6.4.2	Monitoring Data.....	20
6.5	National Park Service Data	20
6.5.1	Data Mining	20
6.6	External Data	21
6.7	GIS Data	21
6.8	Standard Operating Procedures.....	21
Chapter 7: Quality Assurance and Quality Control		23
7.1	Introduction	23
7.2	Objectives	23
7.3	File Naming.....	23
7.4	Standards and Review Procedures.....	23
7.5	Monitoring Protocol Development	24
7.6	Standard Operating Procedures.....	24
Chapter 8: Documentation		25
8.1	Introduction	25
8.2	Objectives	25
8.3	Network Standards	25
8.4	Spatial Data Documentation	25
8.5	Tabular Database Documentation.....	26
8.6	Vital Signs Protocol Documentation	26
8.7	Standard Operating Procedures.....	26
Chapter 9: Data Ownership and Sharing		27
9.1	Introduction	27
9.2	Objectives	27
9.3	Data Ownership	27
9.4	Data Sensitivity	27
9.5	Access Restrictions to Sensitive Data.....	28
9.6	Standard Operating Procedures.....	28
Chapter 10: Data Dissemination		29
10.1	Introduction	29
10.2	Objectives	29
10.3	Data Distribution.....	29
10.3.1	NatureBib	29

10.3.2 NPSpecies	29
10.3.3 NPS Data Store	30
10.3.4 NPS Focus	30
10.3.5 Electronic Technical Information Center (ETIC)	30
10.3.6 Northeast Region (NER) Philadelphia Science Office (PHSO)	31
10.3.7 U.S. Geological Survey (USGS)-NPS Vegetation Mapping Website	31
10.3.8 Mid-Atlantic Network Website	31
10.4 Feedback Mechanisms	31
Chapter 11: Records Management and Object Curation	33
11.1 Introduction	33
11.2 Objectives	33
11.3 Digital Data Maintenance	33
11.3.1 Short-Term Data Sets	34
11.3.2 Long-term Monitoring of Data Sets	34
11.3.3 Spatial Data	35
11.4 Storage and Archiving Procedures for Digital Data	35
11.4.1 Storage Facilities for Electronic Archives	37
11.4.2 Directory Structure for Electronic Archives	37
11.4.3 Backup Procedures for Digital Data	37
11.4.4 Data and Network Security	37
11.5 MIDN Storage of Hard-Copy and Digital Reports	38
11.6 Specimens	38
11.7 Photographs	38
11.7.1 Aerial Photographs	38
11.8 Role of Curators in Storage and Archiving Procedures	39
11.9 Standard Operating Procedures	39
Chapter 12: Project Tracking and Documentation	41
12.1 Introduction	41
12.2 Objectives	41
12.3 Project Tracking	41
Chapter 13: Implementation	43
13.1 Introduction	43
13.2 Objectives	43
13.3 Network Implementation	43
Literature Cited	45

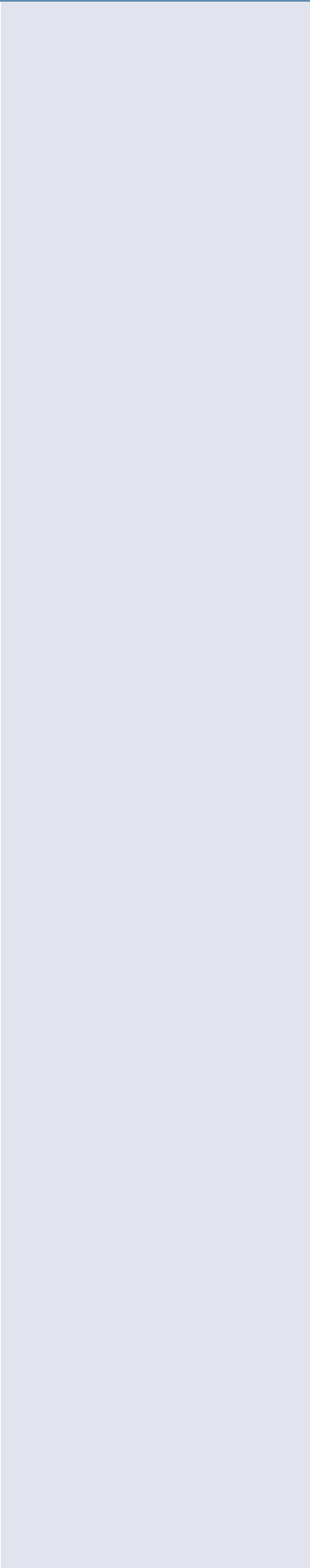


Tables

Table 1.1. Categories of data products and project deliverables	2
Table 6.1. Examples of park-based natural resource information that is processed or documented by the Network	20
Table 11.1. Digital Data Tracking Database.	35
Table 11.2. Air Photo Tracking Database.....	39

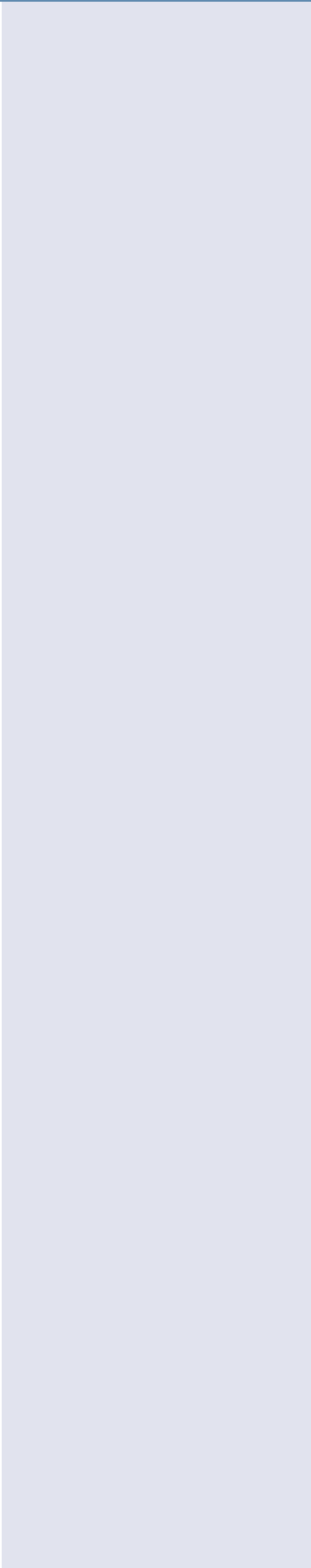
Figures

Figure 1.1. National park units in southeastern Pennsylvania and Virginia that comprise the Mid-Atlantic Inventory and Monitoring Program Network	1
Figure 1.2. Conceptual Layout of the MIDN Data Management Plan.....	3
Figure 2.1. Screen capture of the MIDN data file folder structure.....	6
Figure 2.2. Screen capture of the public Mid-Atlantic Network website.....	7
Figure 2.3. Screen capture of the internal Mid-Atlantic Network website.....	8
Figure 3.1. Project and data workflow with core data management activities in bold	10
Figure 10.1 Screen capture of the NPS Data Store website.	30



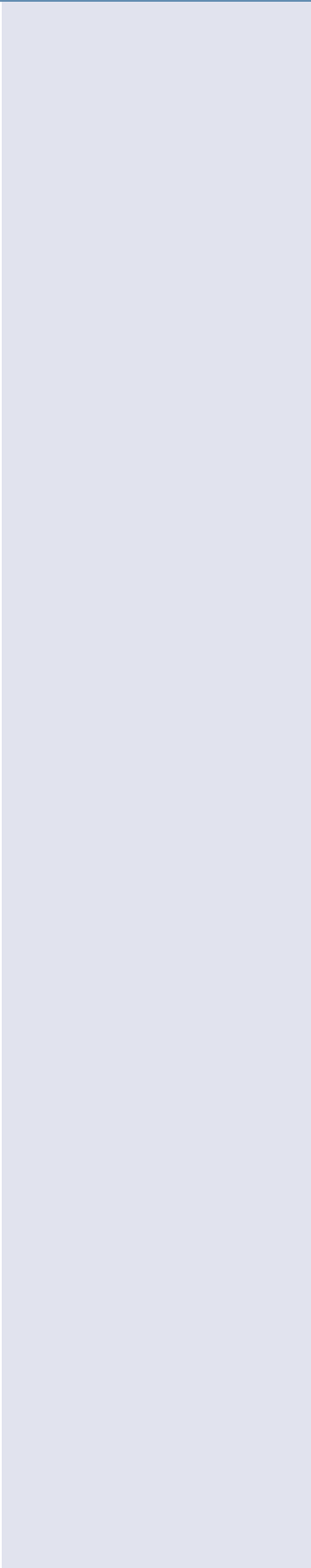
Supplemental Documents

MIDN Directory Structure Guidance Document
NER Inventory Product Review Work Flow
NER Vegetation Mapping Project Work Flow
NER NatureBib Data Management Plan
NER NatureBib Data Management and Data Entry Standards Manual
NER NPSpecies Data Management Plan
NER Species Certifier's Guide
NER NPSpecies Data Entry Standards
NER NPSpecies Museum Voucher Fields
NER NPSpecies Observation Form
NER GIS Specifications
NER GPS Metadata Form
MIDN File Naming Guidelines
NER I&M Product Specifications
NER Biological Metadata Tools



Acronyms

ANCS+: Automated Cataloging System	NERO: Northeast Regional Office
APCO: Appomattox Court House National Historical Park	NPS: National Park Service
ASCII: American Standard Code for Information Interchange	NRDT: Natural Resource Database Template
BIL: Band Interleaved by Line	NRPC: Natural Resource Program Center
BOD: Board of Directors	PETE: Petersburg National Battlefield
BOWA: Booker T. Washington National Monument	PHSO: Philadelphia Science Office
CSDGM: Content Standards for Digital Geospatial Metadata	QA: Quality Assurance
DMP: Data Management Plan	QC: Quality Control
EISE: Eisenhower National Historic Site	RICH: Richmond National Battlefield Park
ESRI: Environmental Systems Research Institute	SAC: Science Advisory Committee
FGDC: Federal Geographic Data Committee	SHEN: Shenandoah National Park
FOIA: Freedom of Information Act	SOP: Standard Operating Procedures
FRSP: Fredericksburg and Spotsylvania National Military Park	TAC: Technical Advisory Committee
FTSC: Field Technical Support Center	TIC: Technical Information Center
GETT: Gettysburg National Military Park	TIFF: Tagged Image File Format
GIS: Geographic Information System	USGS: U.S. Geological Survey
GMP: General Management Plan	VAFO: Valley Forge National Historical Park
GPS: Global Positioning System	XML: Extensible Markup Language
HOFU: Hopewell Furnace National Historic Site	
I&M: Inventory and Monitoring	
IMG: Imagine	
IT: Information Technology	
ITD: Information Technology Division	
JPEG: Joint Photographic Experts Group	
LAN: Local Area Network	
MIDN: Mid-Atlantic Network	
Mr. SID: multiresolution seamless image database	
NBII: National Biological Information Infrastructure	
NCSU: North Carolina State University	
NER: Northeast Region	



Executive Summary



Great spangled fritillary (*Speyeria cybele*) on milkweed (*Asclepias* spp.)

The central mission of the National Park Service (NPS) Inventory and Monitoring (I&M) Program is to provide timely and usable scientific information about the status and trends of park resources to park managers. To meet this challenge, a method for data management that can effectively produce, maintain, and distribute the products of scientific investigation conducted in our parks is required.

Good data management provides the means by which a thorough understanding of the value of scientific information about our natural resources can become a part of the National Park Service heritage. Data management refers to the framework by which data are acquired, maintained, and subsequently made available. Data management is not an end unto itself, but a means of maximizing and insuring the quality and utility of our natural resource information. A robust data management system is particularly important for long-term programs where the life span of a data set will span the careers of several scientists. Sound data management is vital to the success of any long-term research program.

The purpose of the Mid-Atlantic Network's data management plan is to provide the I&M Program and park staff with a conceptual framework for a method of data management that will ensure the production and dissemination of scientific information about the status and trends of park resources. The goals of the Mid-Atlantic Network (MIDN) data management program are to make data and information:

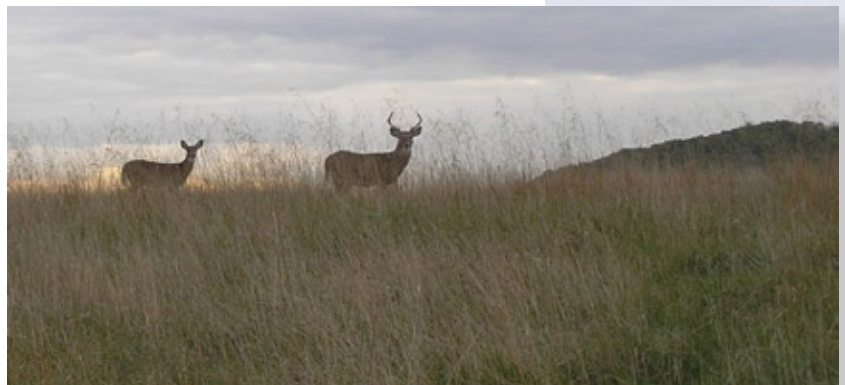
- Available - Data are discoverable, accessible, and easily located.
- Usable - Data are stored in a stable, reliable, and interpretable data retrieval system.

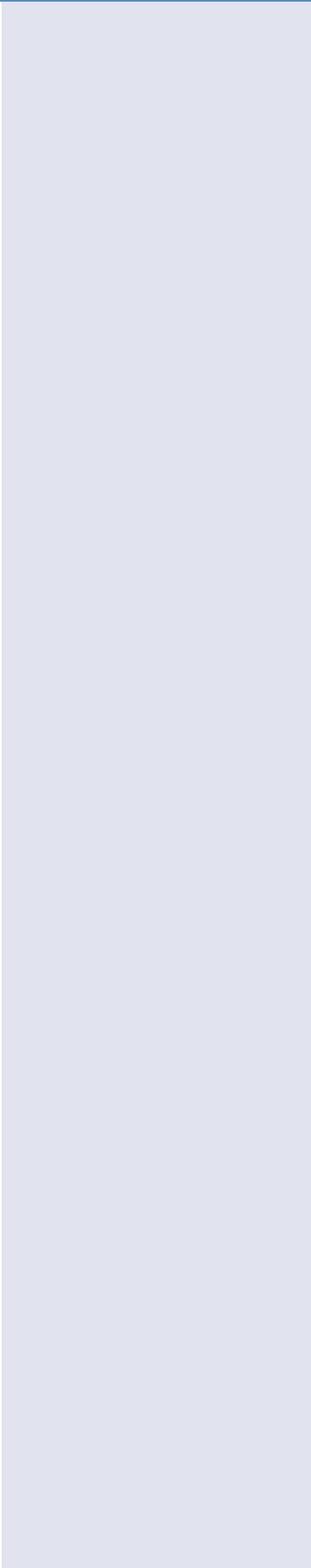
- Shareable - Data products are complete, certified for quality assurance, screened for sensitive information, formatted for use, and documented for interpretation by others.
- Integrated - Data products are consistent with data exchange standards, interoperable with related natural resource data sets, and collected and stored in a way that optimizes the tradeoff between meeting local needs and achieving compatibility with other agencies and partners.
- Interpreted - Data have been reviewed, summarized, and transformed into useful information.

This plan is to be used in conjunction with the Data Management Guidelines for Inventory and Monitoring Networks (NPS 2008). These guidelines are referred to as the national-level data management plan (DMP), which describes national policies and overall principles used across all I&M Program networks. The structure of the Data Management Plan for the Mid-Atlantic Network and much of its content is derived from the national DMP. This plan provides details for how data management activities will be conducted within the network and specifically describes how MIDN will:

- Support Inventory and Monitoring Program objectives
- Acquire and process data
- Assure data quality
- Document, analyze, summarize, and disseminate data and information
- Maintain nationally developed data management systems

White-tail deer (*Odocoileus virginianus*) at Valley Forge National Historical Park





Chapter 1: Introduction and Background

1.1 Inventory and Monitoring Program

The Inventory and Monitoring (I&M) Program represents a long-term commitment by the National Park Service (NPS) to assess and document the status and trends of park ecological resources. In 1998, the National Parks Omnibus Management Act (NPS 1998) established the framework for the I&M Program with a call to fully integrate natural resource monitoring and other scientific activities into the management processes of the National Park System.

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

To carry out this mission, the National Park Service initiated the service-wide natural resource Inventory and Monitoring Program, which encompasses 270 parks with significant natural resources. These 270 parks were grouped into 32 networks of ecologically similar sites. Each I&M Program network has been tasked with documenting the vertebrates and vascular plants found in network parks, developing a management-based ecological monitoring program with a written plan and protocols, crafting a data management plan that encompasses all aspects of the network program.

1.2 The Mid-Atlantic Network

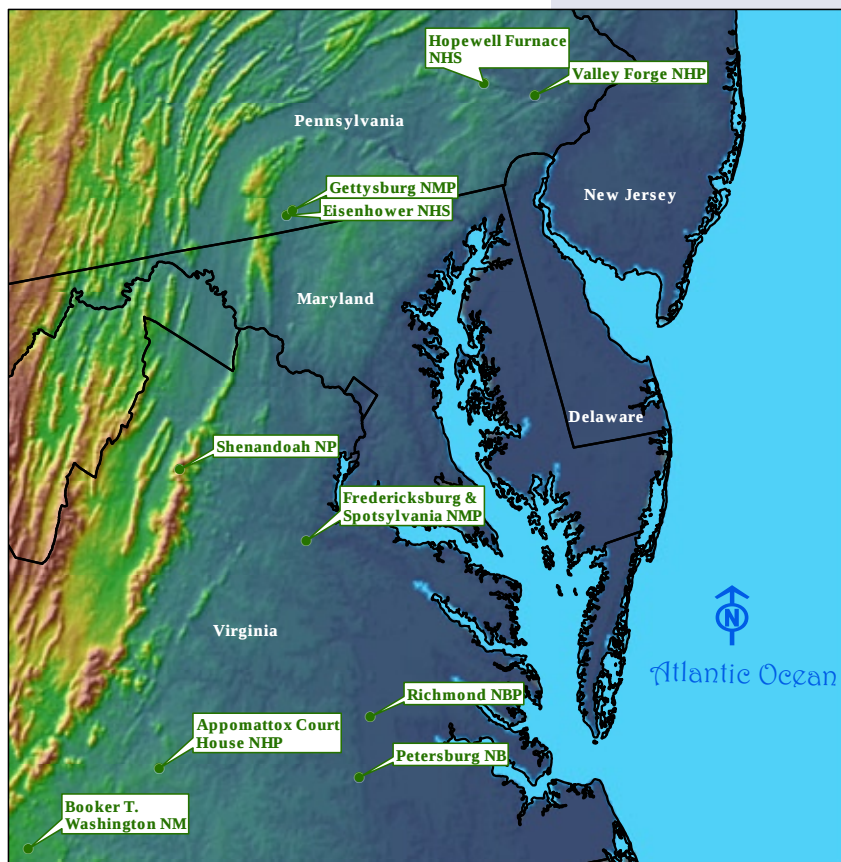
Funding for the 32 I&M Program networks was phased in over several years, with the first 12 networks receiving operating funds beginning in October 2000. The Mid-Atlantic Network (MIDN) was established in April 2003, part of the fourth and final group of funded networks.

The Mid-Atlantic Network consists of ten parks distributed from southern Pennsylvania to southern Virginia (Figure 1.1). The network includes several small, cultural parks with a limited history of natural resource monitoring, as well as, the comparatively large Shenandoah National Park, which had been designated a

prototype park for development of vital signs monitoring under the I&M Program.

The ten parks that make up the Mid-Atlantic Network are Appomattox Court House National Historical Park (APCO), Booker T. Washington National Monument (BOWA), Eisenhower National Historic Site (EISE), Fredericksburg & Spotsylvania National Military Park (FRSP), Gettysburg National Military Park (GETT), Hopewell Furnace National Historic Site (HOFU), Petersburg National Battlefield (PETE), Richmond National Battlefield Park (RICH), Shenandoah National Park (SHEN), and Valley Forge National Historical Park (VAFO).

Figure 1.1. National park units in southeastern Pennsylvania and Virginia that comprise the Mid-Atlantic Inventory and Monitoring Program Network



1.3 Data Management Overview

With the primary goal to “improve park management through greater reliance on scientific knowledge,” the Mid-Atlantic Network (MIDN) Inventory and Monitoring Program will identify, catalog, organize, structure, and archive high-quality natural resource data, and share it with other institutions and individuals. Researchers, managers, and administrators will rely on the network to secure and maintain

the quality of these data over a long period of time.

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

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

To meet I&M Program goals, and to ensure adequate context for primary data products, all of these categories require some level of management to ensure their quality and availability. Thus, the network will use a holistic process to generate, process, finalize, and provide natural resource data. All phases of data processing will be integrated, and information about each phase and the corresponding processes will be shared through high-quality documentation.

Table 1.1. Categories of data products and project deliverables

Category	Examples
Raw data	GPS rover files, raw field forms and notebooks, photographs and sound/video recordings, telemetry or remote-sensed data files, biological voucher specimens
Derived/compiled 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

1.4 MIDN Data Management Goals

The goals of data management in the Mid-Atlantic Network are to provide accurate, efficient, and effective information to support resource management and protection. These goals are not limited to data collected by the Network; we plan to serve as a repository for existing data sets, and we will work with parks in the network to manage data for a wide range of park resource management projects. To meet these goals, park managers, cooperators,

and other data users need to know what data are available from the network. They need to know the location, quality, timeliness, and uses of available data; how to incorporate these data into resource management decisions; and how the data will be managed over time.

The NPS Strategic Plan (NPS 2000), Mission Goal 1b, requires that “...management decisions about resources and visitors are based on adequate scholarly and scientific information...” In addition, long-term Goal #1b1 states that acquiring “...outstanding data sets...of basic natural resource inventories of all parks...” is a desired outcome. The objective of the I&M Program is to provide scientifically and statistically sound data for resource management, and to ensure that quality data are available for this task. To fulfill these objectives, the Mid Atlantic Network must:

- 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.
- Make certain network data are readily available, by implementing standard procedures for distributing data, while protecting sensitive data; and by designing a standardized filing system for organizing I&M Program information.
- 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.
- 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.

1.5 National Data Management Plan

The I&M Program’s national-level Data Management Plan (NPS 2008a) was published in 2008 and serves as a reference that guides and complements the specific data management

procedures that are developed by each network.

The Mid-Atlantic Network Data Management Plan does not attempt to revisit all the concepts described in the national plan. It merely attempts to identify how various components of data management will be implemented within the network. The introductions and objectives in each chapter of the present document are adopted from the national Data Management Plan and serve as a reminder of the detailed information described in the national plan.

1.6 Intended Audience and Layout of the Data Management Plan

The conceptual model in Figure 1.2 shows the layout of the MIDN Data Management Plan. There will be three tiers of document types, with multiple modules within each tier. Four types of documents are being developed as part of this plan:

1. MIDN Data Management Plan (the summary document)
2. Subject-specific guidance documents (e.g., GIS Specifications)
3. Specific database management plans (e.g., NPSpecies, NatureBib)
4. Standard operating procedures (SOPs)

The Mid-Atlantic Network developed this data management plan in modular format with the hope of increasing both its readability and its usability for a wider audience. This format allows readers to easily locate and access subdocuments pertaining to a particular element of data management guidance or process. Individual procedural documents within the plan can be provided as stand-alone documents to park and regional data management staff, cooperators, and project leaders. This modular format also allows for a more complete and technical review. Technical experts can be provided with applicable portions of the plan based on their expertise (i.e., GIS, databases, metadata, GPS, etc.).

Standard operating procedures (SOPs) developed as part of this plan detail work processes that are to be conducted or followed by network staff and cooperators. SOPs are useful because they document the way activities are to be performed, which facilitates consistent conformance to technical and quality system requirements and supports data quality. The development and use of SOPs promotes qual-

ity through consistent implementation of a process or procedure within the organization, even if there are temporary or permanent personnel changes. Supplementary documents (i.e., Guidance documents and SOPs) may be found online at http://science.nature.nps.gov/im/units/midn/dm_sop.cfm.

1.7 Revisions

The MIDN Data Management Plan covers I&M Program needs based on the most current information systems technology relevant through 2008. Revisions to this plan and associated data management documents (guidelines and SOPs) will be made as needed. Changes will be recorded in the Revision History Log that accompanies each document. Version numbers will increase incrementally by hundredths (e.g., version 1.01, 1.02,...) for minor changes and by whole number (e.g., version 2.0, 3.0,...) for major changes.

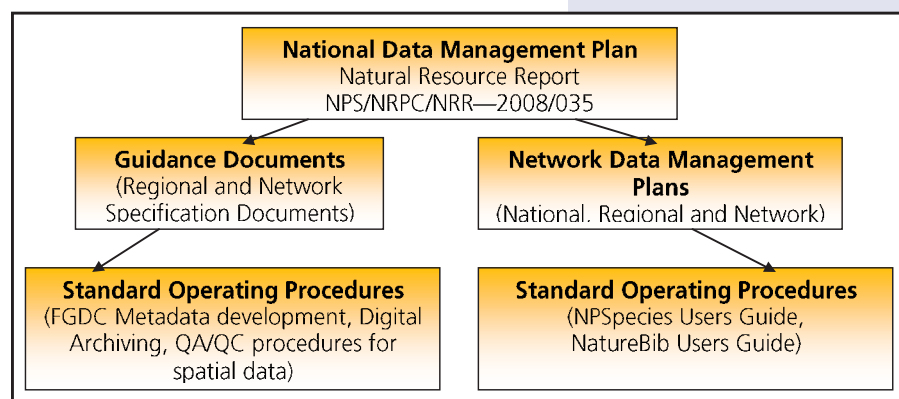
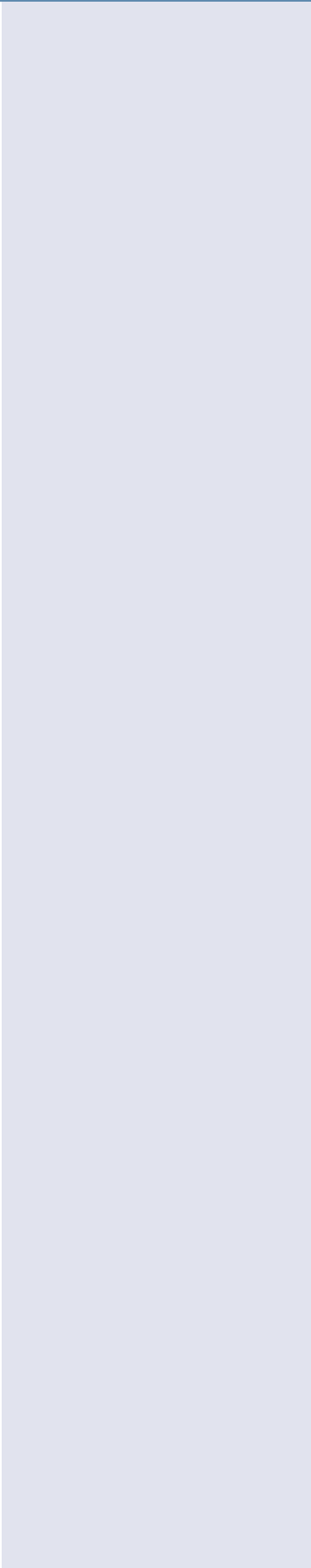


Figure 1.2. Conceptual Layout of the MIDN Data Management Plan



Chapter 2: Infrastructure and System Architecture

2.1 Introduction

Infrastructure refers to the computers and servers necessary to perform the daily tasks run by the Mid-Atlantic Network. Systems architecture refers to the applications, databases, repositories, and tools, supported by that infrastructure that the network uses to manage data. This chapter describes the computer resources housed at the Mid-Atlantic Network and the systems architecture used by the network to store, maintain, and distribute data products.

2.2 Objectives

The objectives of proper management, configuration, and maintenance of I&M Program network infrastructure and systems architecture are to:

- Ensure network staff has appropriate access to electronic files that are secure and protected from accidental or malicious loss
- Simplify software installation and management, and expedite access to and retrieval of data using automation tools
- Make available collaborative technologies and tools that enhance networking and communications at both intra- and interagency levels
- Develop and implement efficient means for storing, managing, searching, and disseminating electronic files

2.3 Computer Infrastructure

The Mid-Atlantic Network relies heavily on park, regional, and national information technology (IT) personnel and resources to maintain its computer infrastructure. These sources provide assistance including, but not limited to, hardware replacement, software installation and support, email administration, security updates, virus-protection, telecommunications,

networking, and backups of servers. Communication with park and regional IT specialists is essential to ensure the growth, stability, and security of the network's infrastructure.

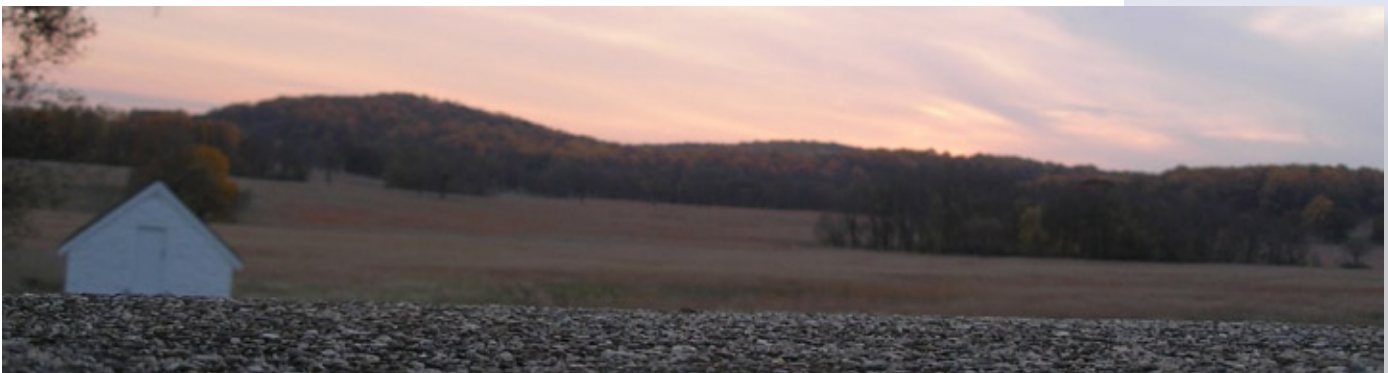
The MIDN computer infrastructure is composed of a networked system of computers and servers, which is the foundation for the information system. This infrastructure has four main components: servers maintained at the national level, park-based local area networks (LANs), a MIDN data server, and individual staff workstations. These components each host different parts of the NPS and MIDN information systems. The infrastructure is maintained by park, regional, and national IT specialists, who administer all aspects of system security and backups.

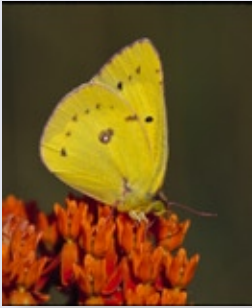
The Mid-Atlantic Network is based at Fredericksburg and Spotsylvania National Military Park (FRSP) and has been integrated into the park LAN. Network computer resources include laptops and docking stations for each of the staff members as well as one server.

The park IT specialist is responsible for the development and management of the local infrastructure with oversight and support from regional and national IT programs. Since the IT specialist already supports three parks (Fredericksburg and Spotsylvania NMP, George Washington Birthplace NM, and Thomas Stone NHS), the network makes every effort to be as independent as possible by working with regional and national IT specialists where feasible.

As technology relating to computer hardware advances, and the network's computer infrastructure evolves, modifications will be made to meet data management objectives. Communication and networking with park unit IT specialists is essential for ensuring adequate

Valley Forge National Historical Park



Orange Sulfur (*Colias eurytheme*)

resources and service continuity for the network's computer systems architecture.

2.4 Introducing Bob

The Mid-Atlantic Network has acquired a 600GB server and named him Bob. Bob is part of the Fredericksburg and Spotsylvania NMP LAN, and lives at the network's office within that park. Bob's primary duties are to act as a file server and provide space for backing up office laptops. Bob is regularly connected to a 1TB external hard drive for full monthly backups and weekly incremental backups. The drive is rotated once a month and stored at an off-site location.

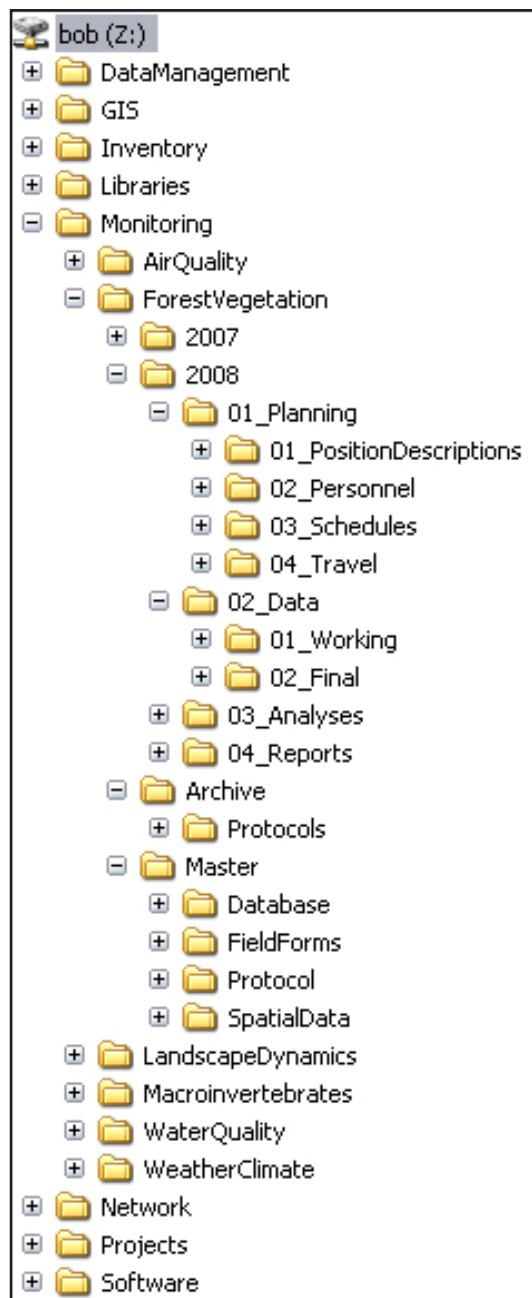


Figure 2.1. Screen capture of the MIDN data file folder structure.

2.5 Local Workstations

Network staff manage files on their personal computers, and place files that need to be shared on the MIDN server. Staff are responsible for regularly copying their workstation files to workspace on the MIDN server for backup.

2.6 Data File Folder Structure

All electronic files connected with the MIDN Inventory and Monitoring Program are stored with Bob. I&M Program project data should always be stored under the corresponding Inventory, Monitoring, or Projects folder. The specific storage location within these folders depends on the project to which the data relates as well as the status of the data (i.e., raw vs. archived). Only personnel employed by the MIDN I&M Program have access to these directories. Employees should not maintain any original I&M Program project data on their local hard drives.

The network has defined a data file folder structure (see Figure 2.1) for organizing files on the network server. The eight Level 1 folders, or directories, are defined as:

- **DataManagement** – This directory contains the tools and documents used for data management within the network, including the current version of this Data Management Plan and related standard operating procedures (SOPs), as well as additional documents from the national I&M Program.
- **GIS** – This directory contains an extensive library of spatial data for each of the network parks as well as adjacent parks with which the network works. Data are organized by four-letter park code and/or state and region
- **Inventory** – This directory contains files relating to each of the network's inventories. It contains an archive folder as well as an individual folder for each active project. When an inventory is completed and the final products have been moved to the Libraries folder, the folder will be moved into the archive folder of this directory.
- **Libraries** – This directory contains the network's libraries. Currently the network maintains four libraries: artwork, maps, photos, and publications. The last two are the most extensive. This folder is for final documents only.

- **Monitoring** – This directory contains a folder for each of the network’s monitoring protocols and will be the most extensive of the eight folders. The monitoring protocols all represent long-term projects. Each protocol’s folder includes a master folder for materials that define the protocol, a working folder by year, and an archive folder for previous years. Within the yearly folder, data are further separated by planning, data, analysis, and report files.
- **Network** – This directory contains files related to the network, but not to a specific short or long-term project. These files include budget, equipment, Board of Director and Science Advisory Committee presentations, newsletters, and plans such as the AARWP, Network Charter, and Phase Three Monitoring Plan. Note: For plans, this folder will contain the background and archived information. The current, final version of each of the plans is found in the Libraries folder under publications.
- **Projects** – This directory contains files relating to projects (other than inventory or monitoring projects) that the network participates in and/or funds. These include watershed condition assessments, vegetation maps, and fire fuel projects. This directory includes an archived folder as well as an individual folder for each active project. When a project is completed and the final products have been moved to the libraries, the folder will move into the archive of this directory.
- **Software** – This directory contains copies of software and drivers the network uses and relevant key code information. It is important to keep copies of old software for reading archived data.

Further details regarding the functional and structural requirements for organizing files on the network server may be found in the MIDN Directory Structure Guidance Document.

2.7 MIDN Websites

The Mid-Atlantic Network maintains two websites to present information about the network to interested parties. Both websites are hosted on a server at the Natural Resource Program Center (NRPC) in Fort Collins, Colorado. The websites are currently maintained using Macromedia Dreamweaver and ColdFusion. The websites follow a national design standard

in compliance with the National Park Service’s web design guidelines.

The **public website** is accessed at <http://science.nature.nps.gov/im/units/midn/>. Figure 2.2 shows a screen capture of the home page. The website includes information about the network, inventories, monitoring programs, and reports. The site is updated as new materials become available.

Figure 2.2. Screen capture of the public Mid-Atlantic Network website



The **internal website** is accessed at <http://www1.nrintra.nps.gov/im/units/midn/index.cfm>. Figure 2.3 shows a screen capture of the internal home page. The website includes links to most sections on the public website as well as additional information that may be preliminary or otherwise not suitable for public distribution.

2.8 Standard Operating Procedures

The following guidance document provides further detail on MIDN computer infrastructure and systems architecture.

MIDN Directory Structure Guidance Document

Figure 2.3. Screen capture of the internal Mid-Atlantic Network website



Chapter 3: Project Management and Data Life Cycle

3.1 Introduction

Data management within an I&M Program network must coordinate the broad phases of short- and long-term inventory or monitoring projects, as well as the detailed steps of data acquisition and processing. Establishing a structure, sequence, and steps for these phases is needed to ensure that all tasks are completed, responsibilities are clear, resources are available, and overall workflows are synchronized.

3.2 Objectives

The objectives of establishing and maintaining proper project management and a data management life cycle are to:

- Ensure that a sequence of steps is applied to all network projects, from project inception through project conclusion
- Ensure the necessary and appropriate participation of staff and cooperators at all stages of project development
- Ensure that all data resulting from projects are consistently processed, reviewed, and finalized.

3.3 Project Work Flow

There are two main types of projects from the perspective of managing workflow:

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

From a data management standpoint, the primary difference between short- and long-term projects is an increased need to adhere to and maintain standardizations in long-term projects. Maintaining standardization from year to year is necessary in order to ensure the comparability of data collected over an extended period of time. While the need to follow standards is still present for short-term projects, sometimes the cost of compliance outweighs the benefits due

to the scope, budget, and level of NPS influence over project details. Nevertheless, both short-term and long-term projects share many work flow characteristics, and both generate data products that must be managed and made available.

The workflow of a project can be broken down into five primary phases: planning and approval, design and testing, implementation, product integration, and evaluation and closure. Figure 3.1 outlines the data management activities carried out by network staff during different phases of a project.



Vegetation data collection at Richmond National Battlefield Park

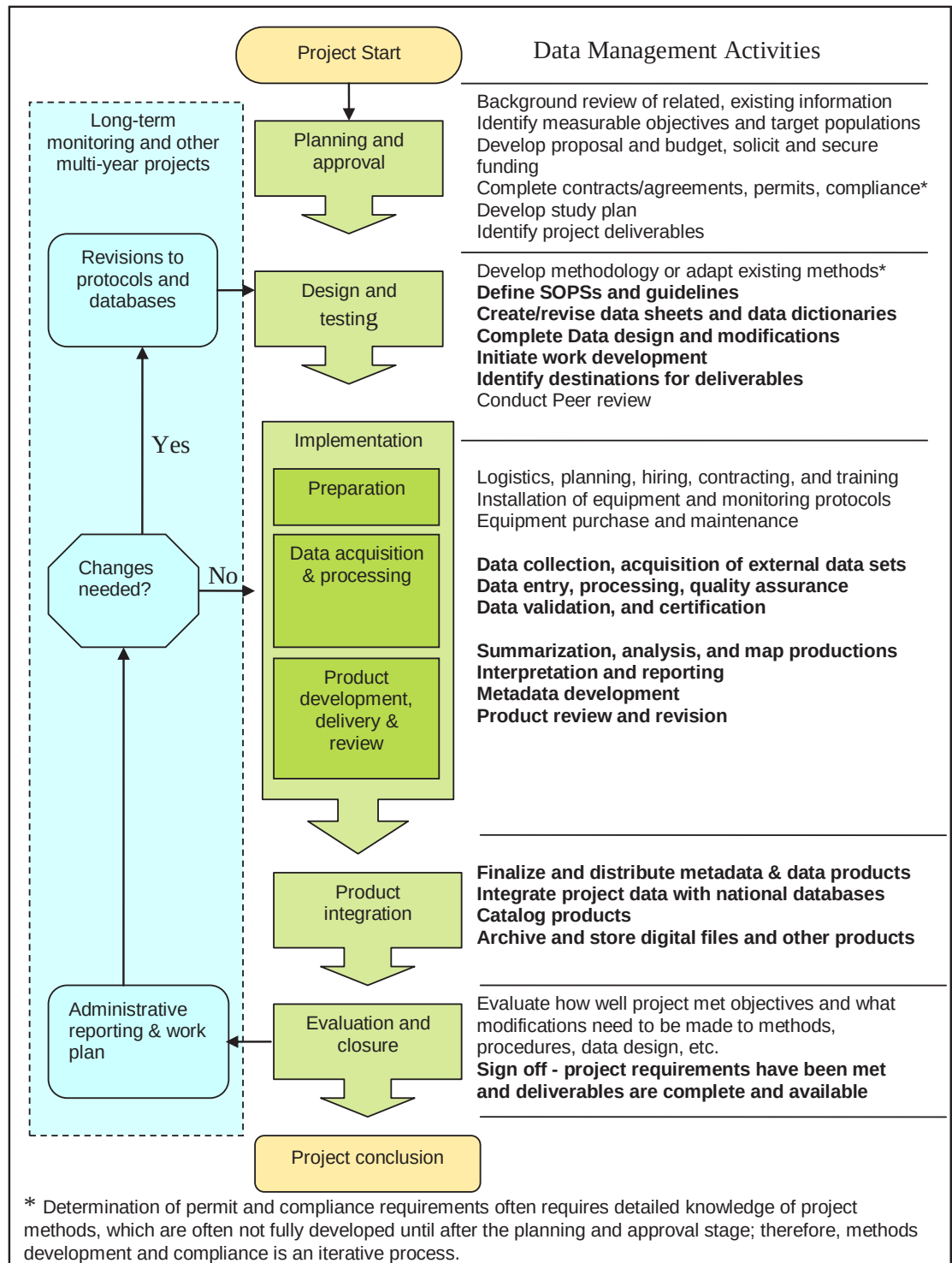
3.3.1 Planning and approval

During the initiation and approval phase, preliminary decisions are made regarding the scope of the project and its objectives. A scope of work may or may not be written for each project, but a proposal will be developed and funding sources, permits, and compliance must be addressed. A cooperative agreement or contract is developed and finalized. Although many of these responsibilities reside with the project leader and/or program administrators, data management staff must be involved in identifying project deliverables and assuring that each contract or agreement includes a clear list of these deliverables, with reference to either national, regional, or network information and data standards. All contracts, agreements, and permits should include standard language that describes the formats, specifications, and timelines for project deliverables.

3.3.2 Design and testing

Procedures for how data will be acquired, processed, documented, analyzed, and reported

Figure 3.1. Project and data workflow with core data management activities in bold



are developed during this phase. The project leader is responsible for the development and testing of project methodology as well as creating or modifying existing SOPs. The data manager will work with the project leader to create the data dictionary and to design, develop, and test the project database. See Chapter 5 for further details on database design for the Mid-Atlantic Network.

3.3.3 Implementation

The groundwork of the project is implemented during this phase. Field crews are hired, equipment is purchased and installed, data are collected, data containers are developed, and deliverables are identified. The project leader carries the primary responsibility during this phase, but the data manager acts as a facilitator by providing training and support for database or data processing applications, GIS, and GPS.

All raw and derived data products, metadata, reports, and other documentation should be delivered to the data manager. Additionally, all product deliverables should be reviewed by the data manager to ensure that proper quality assurance and quality control (QA/QC) procedures have been followed and that product specifications have been met. See Chapter 6 for further details on data acquisition, processing, and reporting and Chapter 7 for details on QA/QC procedures.

3.3.4 Product integration

Once the data and project deliverables have met all QA/QC requirements and been finalized, the data manager will upload data from the working database to the project's master database, merge data products and other deliverables into national and network databases, finalize and distribute metadata to clearinghouses, distribute products to their appropriate audiences, and archive all final products. Further details on product cataloging, documentation, data sharing, data dissemination, and archiving are found in Chapters 6 and 8-11.

3.3.5 Evaluation and closure

At the conclusion of a project or at the end of a field season, network staff will conduct a project review to determine the effectiveness of the existing program and to obtain feedback on project methods. Due to the relatively small size of the MIDN program, these sessions will likely be informal, but they will be documented and archived.

3.4 Project Data Life Cycle

Project data take different forms and are maintained in different places as they are acquired, processed, documented, and archived. These phases can be modeled as a sequence of events and tasks that involve interaction between each of the forms. Specifics regarding the data life-cycle can be found in the national DMP. Key highlights of this cycle include:

- Archiving all raw data in their native format
- Using working databases as the focal point of all modification, processing, and documentation of data collected for a given time period
- Archiving and posting or integrating data with national archives after all quality assurance, documentation, and certification are complete

- Uploading data for long-term monitoring projects into master databases
- Using certified data sets to develop reports and other data products (maps, checklists, etc.). These products are also archived and posted to appropriate national applications and repositories
- Documenting all subsequent changes to certified data sets in an edit log, which is distributed with the data

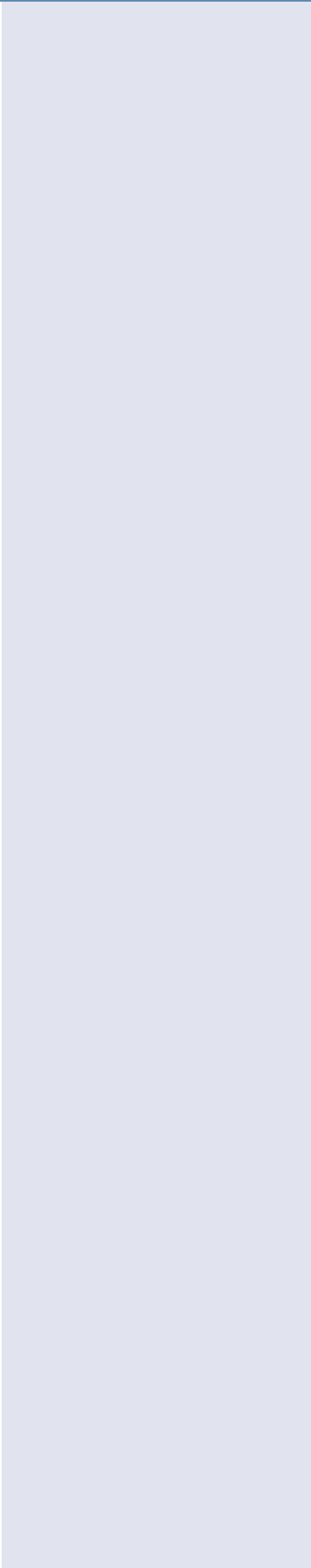
3.5 Standard Operating Procedures

The following Standard Operating Procedures provide further guidance for project work flow:

NER Inventory Product Review Work Flow

NER Vegetation Mapping Project Work Flow

Project Data Life Cycle Guidance Document (Under development as of January 2009)



Chapter 4: Roles and Responsibilities

4.1 Introduction

Data management is about individuals and organizations as much as it is about information technology, database practices, and applications. All staff participating in an I&M Program network project have a role in data stewardship, and project data sets and products reflect the roles of all who have contributed to them. In order to meet the data management goals and standards developed by the National Park Service and the I&M Program, all staff must understand their associated roles and responsibilities.

4.2 Objectives

The objectives of delineating data management roles and responsibilities are to:

- Clearly define roles associated with functions
- Establish data ownership throughout all phases of a project
- Instill data accountability
- Ensure that adequate, agreed-upon data quality and metadata metrics are maintained on a continuous basis.

4.3 National Level Staff and Coordination

Data managers have been hired for each of the 32 I&M Program networks nationwide. These network data managers regularly coordinate with each other and with national program staff via annual meetings, conference calls, workgroups, a listserv, websites, and informal communication. They share the workload by working closely with each other while developing Data Management Plans for their respective networks. This model of cooperation and communication has been very effective, and should be applied to other resource data management issues and implemented widely by administrators.

The Mid-Atlantic Network maintains an active role in promoting practical consistency among protocols and data sets involving other networks and organizations. GIS managers and data managers from all network parks are encouraged to participate in I&M Program development and activities. Park data managers are encouraged to use I&M Program resources, and to communicate with network and pro-



gram staff to share knowledge, expertise, and information.

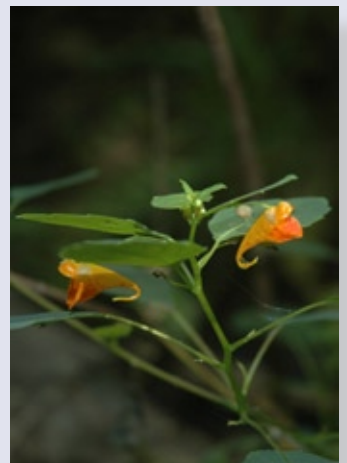
4.4 Northeast Region Data Management Staff

The Northeast Region Inventory and Monitoring Program falls under the Natural Resource and Stewardship Division of the Northeast Regional Office (NERO), and directly under one of two Northeast Region Chief Scientists. Natural Resource and Stewardship staff at the regional level will assist regional network I&M Program staff in:

- Identifying and prioritizing regional and individual park needs
- Identifying qualified researchers and potential cooperators
- Administering research contracts and cooperative agreements
- Providing scientific consultation as network Board of Directors (BOD) and Science Advisory Committee (SAC) members

At the regional office level, no over-arching natural resource data management positions existed prior to the development of the four I&M Program networks within the Northeast Region. Today, the Northeast region is staffed with four I&M network data managers, each housed at their respective network. These data managers work closely together to leverage resources and expertise to improve data management practices throughout the region.

Inventory and Monitoring program staff at Appamattox Court House National Historical Park



Jewelweed (*Impatiens capensis*)

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



MIDN staff work with seasonal biotechnicians and cooperators on I&M projects



4.5 Mid-Atlantic Network Staff

The MIDN I&M program is directed and supervised by the MIDN coordinator with oversight by the NER regional I&M Program coordinator. The core network staff currently consists of two permanent positions, the network coordinator and data manager. The core functions are currently augmented by expertise in botany and database support and via cooperative agreements with Pennsylvania State University. This botanist/database support position is shared with the Eastern Rivers and Mountains Network. Additional support is provided by a part-time science communicator through the NER Science Office. Seasonal staff members, shared with the Northeast Temperate Network, are hired for the forest vegetation monitoring program, and park staff provide as-

sistance when their time permits. A cooperative agreement also exists with the Center for Earth Observation at North Carolina State University, one of two Regional NPS Field Technical Support Centers (FTSC).

4.5.1 Coordinator/Ecologist

The network coordinator provides the overall direction and guidance for the MIDN I&M Program and also develops and oversees the implementation of the Monitoring Plan. The coordinator acts as the liaison between MIDN staff and network parks, Science Advisory Committee (SAC), Board of Directors (BOD), Technical Advisory Committee (TAC), and the NER Regional I&M coordinator. The coordinator supervises the MIDN professional level positions and provides general oversight and accountability for the network program. The coordinator also serves as the network ecologist and is responsible for overseeing the design, implementation, data analysis, and reporting of the monitoring protocols.

The network coordinator ensures that programmatic data and data management requirements are met as part of the overall network business. The coordinator communicates with network staff, park staff at all levels, and other appropriate audiences to support and emphasize data management as a critical aspect of network business. The coordinator also works with the data manager regarding data management policy and guidelines, budget, staffing, and training.

4.5.2 Data Manager/Biologist

The MIDN data manager is responsible for the development, management, coordination, and implementation of natural resource information systems, including databases, data archives, and Geographic Information Systems (GIS) technology. Responsibilities include creating new databases consistent with NPS standards, quality assurance/quality control (QA/QC) for new data, and generating metadata that are Federal Geographic Data Committee (FGDC) compliant. The data manager assists the network coordinator in making data accessible to parks by summarizing data and generating standard reports as identified by protocols, and sharing data through the Internet or other media. The data manager also serves as the network biologist, providing additional support for MIDN monitoring protocols.

The data manager supports the development, implementation, and maintenance of the data

infrastructure and standards for the network as a whole and serves as the point of contact for NatureBib for all of the parks within the network. The data manager also provides coordination to ensure I&M Program products are gathered and maintained at the network office and are accessible for submission to the national I&M Program office.

4.5.3 Seasonal Biotechnicians

Seasonal-biotechnician positions typically last four months, with data collection as their primary duty. It is anticipated that two seasonal biotechnicians will be needed for the vegetation monitoring program. The biotechnicians are directed by the network coordinator and follow existing protocols to gather data; record, verify and correct data values; and to perform regular data transfers and backups. These biotechnicians also assist with dataset and procedural documentation, and are responsible for documenting any deviations from protocols or study plans.

4.5.4 North Carolina State University Field Technical Support Center (NCSU FTSC)

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

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

NCSU's experience in GIS spans 20 years and includes a long association with public resource management agencies. The Center for Earth Observation is part of the College of Natural

Resources and its GIS research program has focused on the development of appropriate sized GIS programs for non-centralized resource management (e.g., smaller National Park Service units, U.S. Forest Service ranger districts, state park regions, etc.) The NCSU facilities include a full complement of GIS and related software packages, a separate remote sensing research program, a large array of computer hardware configurations within two laboratories, and over 10 professional staff.

4.6 Park Level Data Management Staff

In general the parks within the Mid-Atlantic Network are understaffed in the areas of data management and GIS. Network staff and resources are available to the parks for technical assistance on any issues related to data management.

Shenandoah National Park is the only MIDN park with a permanent data manager. Shenandoah NP was identified as a prototype park for the development of long-term ecological monitoring efforts within the mid-Atlantic region. As a prototype park and in partnership with the U.S. Geological Survey (USGS), the park was charged with developing and refining long-term monitoring protocols that could be of utility to other parks, in addition to supporting management of Shenandoah's natural resources. The Shenandoah NP data manager has the same basic duties and responsibilities as the MIDN data manager, but also has the responsibility for mentoring and training others and developing and testing new approaches to data analysis, synthesis, and reporting of monitoring results.

Shenandoah National Park



4.7 Data Management Coordination

The network data manager works with national I&M Program data management staff and regional resource data management personnel to maintain a high level of involvement with service-wide and regional databases and data management policy. The network data manager works locally with network personnel, park staff, and cooperators to promote and develop workable standards and procedures that result in the integration and availability of data sets.

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

Chapter 5: Databases

5.1 Introduction

The quantity and complexity of data acquired as part of I&M Program network inventory and vital signs monitoring projects require a well-planned and consistent approach to database design, development, and management. Databases are the cornerstone of an I&M Program network, and are the primary repositories of data from long-term studies that need to be accessible and applicable for decades to come. In addition, network databases must accommodate a wide diversity of data types, enforce data integrity, support a variety of queries and output, and help substantiate resource management decisions.

5.2 Objectives

The objectives of effective database design, development, and management within I&M Program networks are to:

- Ensure the long-term stability, security, and accessibility of data
- Encourage common practices, standards, and sharing of database development efforts among networks
- Encourage data exchange and multidisciplinary applications of data
- Ensure databases are designed and built using NPS- and industry-accepted best practices

5.3 Project Database Standards

Project database standards are necessary for ensuring compatibility among data sets and across networks, allowing managers and researchers to compare data and analyze trends across parks or across protocols. Standards also help to encourage sound database design and facilitate interpretability of data sets. The MIDN-developed databases for park and network projects will contain the following main components: core tables, protocol-specific field data tables, common lookup tables, and a user interface. The standard template supported by the national I&M Program is the Natural Resource Database Template, which has been adopted by the MIDN. A SOP for Database Specifications to be developed by the MIDN will describe naming conventions, data dictionaries, data verification methods, and documentation.

5.4 Natural Resource Database Template

The Natural Resource Database Template (NRDT) is a flexible, relational database in Microsoft Access for storing inventory and monitoring data (including raw data collected during field studies). This relational database can be used as a stand-alone database or in conjunction with GIS software (e.g., ArcGIS) to enter, store, retrieve, and otherwise manage natural resource information. The template has a core database structure that can be modified and extended depending on the components and specific sampling protocols in use. The NRDT adds an element of compatibility to databases across networks but does not ensure it. Field methods must also be coordinated in order to be sure that data within databases can reasonably be rolled-up or compared. The NRDT provides a good core set of tables and relationships on which to build a personal-level database. However, as the National Park Service moves toward a more enterprise-level data solution, Access will likely be replaced by another database application.

A description of the NRDT, a data dictionary, general database guidelines and example implementations are located on the NRDT website: <http://science.nature.nps.gov/im/apps/template/index.cfm>.

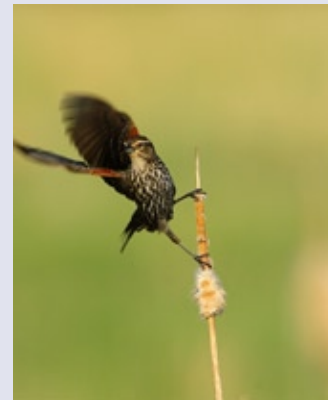
5.5 Natural Resource Monitoring Protocols Clearinghouse

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

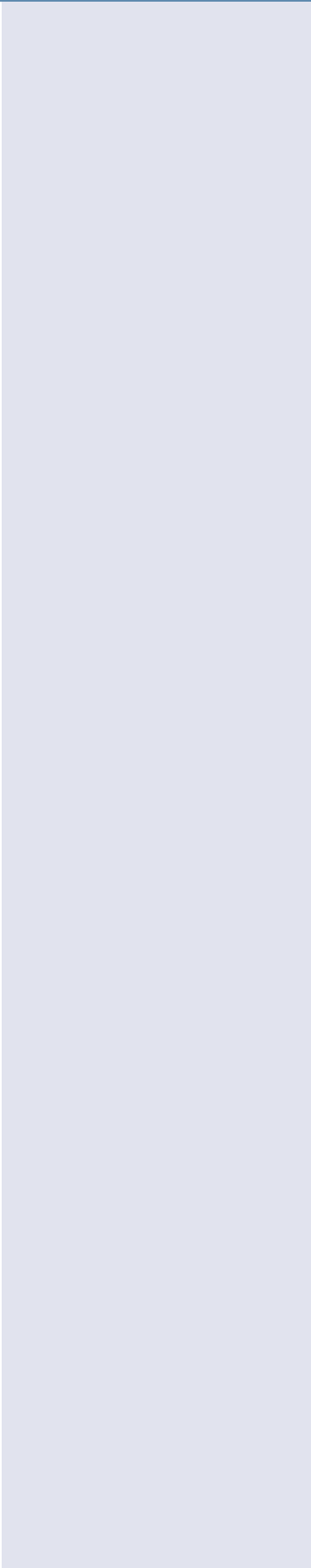
5.6 Standard Operating Procedures

The following standard operating procedures provide guidance for designing databases:

Database Specifications (in development as of January 2009)



Red-winged blackbird (*Agelaius phoeniceus*)



Chapter 6: Data Acquisition, Processing, and Reporting

6.1 Introduction

Interdisciplinary or multi-scale natural resource programs and projects increasingly rely on data gathered from multiple sources. This chapter describes the general steps involved with acquiring, processing, and reporting data to meet standards established by the NPS I&M Program, although these steps could apply to other NPS park units or programs. Data acquisition also encompasses physical objects (e.g., photographs, voucher specimens), which are often collected as part of resource management, inventory and monitoring, and other research projects.

6.2 Objectives

Acquiring, processing, and reporting on natural resource-related data are core responsibilities of I&M Program networks. To effectively meet these responsibilities networks must:

- Identify, evaluate, and acquire, through a variety of sources, data needed by networks to properly carry out inventory or vital signs monitoring projects within parks
- Determine the most reliable and efficient means or tools for collecting data
- Establish routines and procedures for data processing
- Identify and meet the information needs of a broad range of end users including scientists, park managers, park interpreters, and the general public.

6.3 Types of Data

The National Park Service's Inventory and Monitoring (I&M) Program is responsible for acquiring information to complement existing park data and to assist park managers to properly manage and maintain the natural resources of their park. To successfully accomplish this task, information from multiple sources is collected by the MIDN I&M Program and processed to ensure that it meets the data standards established by the MIDN. This chapter describes the general procedures the MIDN follows for acquiring and processing natural resource-related data. Procedures will vary depending on the data source, which can be placed into three general categories:

- **Mid-Atlantic Network Data:** Programmatic data resulting from projects that are initiated, sponsored, or funded by the MIDN.
- **Other National Park Service Data:** Non-programmatic data resulting from projects that are initiated, sponsored, or funded by park units, or by regional or national NPS programs.
- **External Data:** Non-programmatic data produced or managed by agencies, organizations, or individuals other than the National Park Service.

The following sections outline in more detail the manner in which the different data types described above are acquired and processed.

6.4 Mid-Atlantic Network Data

The collection of programmatic data under the purview of the MIDN I&M Program is connected to either natural resources inventories or to vital signs monitoring projects.

6.4.1 Inventory Projects

The acquisition and processing of network inventory data is a straightforward process due to the high level of network control over the inventory projects. These projects are completed by NPS staff or cooperators working under a contract or agreement that has been written specifically by the network for the project. In these agreements the MIDN itemizes data deliverables and formats that are required from the cooperator. The data manager works closely with the principal investigator and the project leader to develop these data criteria, and, when possible, to develop a Microsoft Access database modeled on the Natural Resource Database Template (NRDT) and tailored to the specific data needs of the project. Additional data specifications usually included in agreements are procedures for acquiring and documenting project photographs, collecting specimens, using global positioning systems (GPS), and delivering geospatial data.

Data processing typically consists of reviewing the project deliverables, integrating the project into the various MIDN cataloging systems, and distributing project results. Principal investigators are responsible for verifying and validating the data they submit; however, the MIDN



Green heron (*Butorides virescens*)

Table 6.1. Examples of park-based natural resource information that is processed or documented by the Network

Data Description	Processing
Gray literature, unpublished reports, research summaries	Review and document in NatureBib. Evaluate for species data and link to NPSpecies. Obtain copies for network files (preferably electronic) as appropriate.
Published reports, journal articles	Review and document in NatureBib. Evaluate for species data and link to NPSpecies. Obtain copies for network files (preferably electronic) as appropriate.
Investigator Annual Reports	If the IAR is the only reference to a project, document in NatureBib. If a more comprehensive report has resulted from the project, create a NatureBib record for the report and reference the IAR in the NatureBib record.
Photographs	If photos are part of an assemblage accompanied by documentation, enter into Dataset Catalog. Evaluate for NPSpecies data (photo vouchers).
Databases, spreadsheets, other electronic data assemblages — undocumented	Review for associated documentation. Locate project leader, if possible, for further explanations. Create Dataset Catalog record and document to extent possible.
Databases, spreadsheets, other electronic data assemblages — documented	Create Dataset Catalog record. Evaluate for NPSpecies data. Obtain copies for network files as appropriate.
GIS data — documented and undocumented	Retain copy at MIDN office as applicable. Create meta-data using ArcCatalog and NPS Metadata Tools and Editor.
Species lists	Create NatureBib record if author and date are included with list. Evaluate for NPSpecies data certification
Assemblages of field forms	Create Dataset Catalog record. Evaluate for NPSpecies data.
Data collected by NPS regional or national programs (e.g., air quality, sound, night sky)	Create Dataset Catalog record.
Data collected and managed by other agencies (e.g., stream gauging stations, climate, breeding bird data, other monitoring programs)	Create Dataset Catalog record. Evaluate for NPSpecies data.
ANCS+ natural history specimen data	Evaluate for inclusion in voucher table of NPSpecies.
Park library records	Evaluate for inclusion in NatureBib.
Park archives	Evaluate for inclusion in NatureBib, Dataset Catalog, and NPSpecies.

performs additional quality checks on all data sets. Data management staff are usually in contact with the principal investigator after the submission of final products in order to resolve discrepancies or answer questions that inevitably arise as staff work with the data.

(QA), quality control (QC), reporting, analysis, distribution, and archiving. Depending on the protocol, data can be acquired by park staff, network staff, cooperators, or other agencies or organizations.

Monitoring data need to go through processing and integration steps that are similar to inventory data; however, many of the steps are repeated annually or on a schedule corresponding to sampling frequency. Each monitoring project will have a processing and integration checklist that is specific to the project to ensure that all steps are completed.

6.5 National Park Service Data

6.5.1 Data Mining

A fundamental step in developing and maintaining the information base for the MIDN is locating, evaluating, and documenting, on an ongoing basis, park-related natural resource information. The term “data mining” refers to this process, which requires regular visits to network parks and establishing data mining procedures specific to each park. The range of materials that require documentation is broad, from historical inventories, to complex databases, to remote-sensing data.

As a general rule, the goal of network data mining is not to acquire and store at the network copies of all data residing at parks. Instead, cataloging tools are used to describe and document data sources so that potential users can find them. Data that are an exception to this rule are documents or data sets that are of specific need or interest to network staff (e.g., GIS data, monitoring data), or that complement network projects. Also, as NatureBib expands its capability to store and serve full-text versions of documents, the MIDN is working with network parks to identify and obtain electronic copies of documents for full-text posting on the Internet where they can become widely accessible.

Parks within the network vary widely in how natural resource information is organized and stored. While some parks have developed efficient and standardized methods of cataloging new information as it arrives, other parks have no formalized approach for organizing their natural resource materials.

The MIDN shared a NatureBib data manager with three other regional networks for a number of years. The regional staff member was hired to mine natural resource information

6.4.2 Monitoring Data

Data acquisition for network monitoring projects is guided by established protocols that include step-by-step procedures for all aspects of data collection, data entry, quality assessment

and data for all Northeast Region I&M Program parks. The NatureBib data manager has performed searches of each park's library and staff files; staff interviews; and searches of local universities, Web-based library catalogs, and other agency files.

While the network no longer has a dedicated NatureBib data manager position, the online database continues to be updated as new references are added to the network library.

Table 6.1 summarizes the range of park-based data that the network is cataloging, and the associated processing steps. Specific details on processing steps are provided in the guidance documents listed at the end of this chapter.

MIDN data management staff complete most of these cataloging tasks, and have been the primary users of NatureBib, NPSpecies, and Dataset Catalog. Most parks have had no extra staff available to perform this work or to devote to using the databases. The MIDN places a high priority on maintaining and building these databases, and will continue to perform the work as needed. Data management staff will provide training to park staff interested in learning to use the databases, and will provide reports from the databases in formats that are useful and easily accessible to park staff (e.g., hard-copy printouts, Web-based species lists).

6.6 External Data

Organizations and agencies outside the National Park Service have been an ongoing source of natural resource data for the network and are included in the network's data mining activities. Processing external data follows the same steps and procedures as processing park data: information is evaluated, entered into one or more cataloging databases, and is then posted or made available as appropriate.

Museums and Herbaria

One of the more challenging data mining tasks has been acquiring information on specimens that were collected within park units, but that are housed in a variety of repositories across the country. The Inventory and Monitoring Program at the Washington Office has obtained specimen data from major museums and herbaria that have their holdings in electronic format. The MIDN is examining these records for inclusion in NPSpecies. Smaller regional and state repositories will be checked as time and budget allows. Some of the biological inventories have assisted in checking for museum

records – e.g. Joe Mitchell's Herpetological Inventory.

Frequently the associated and required Automated National Cataloging System (ANCS+) records for these specimens do not exist at the parks. The MIDN works with park museum curators when receiving and managing specimen data from external sources, and provides them with formatted data for uploading into ANCS+, if appropriate.

6.7 GIS Data

Numerous geospatial data sets correspond to park units and surrounding areas in the MIDN. These data sets are acquired by park staff, through cooperative national or regional NPS projects, and from multiple external organizations and agencies.

Spatial data acquired by the network either directly or through contractors must meet established standards and product specifications. Spatial data gathered using global positioning system (GPS) units must also meet established standards.

6.8 Standard Operating Procedures

The following standard operating procedures provide specific guidance for data acquisition, processing, and reporting:

NER NatureBib Data Management Plan

NER NatureBib Data Management and Data Entry Standards Manual

NER NPSpecies Data Management Plan

NER Species Certifier's Guide

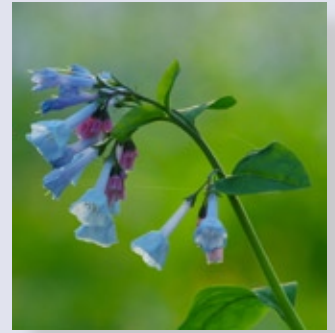
NER NPSpecies Data Entry Standards

NER NPSpecies Museum Voucher Fields

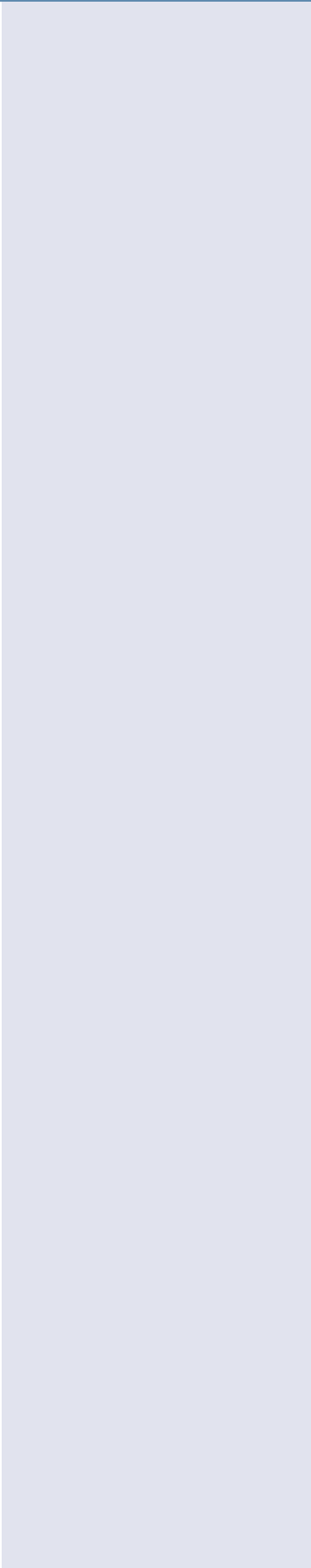
NER NPSpecies Observation Form

NER GIS Specifications

NER GPS Metadata Form



Virginia bluebell (*Mertensia virginica*)



Chapter 7: Quality Assurance and Quality Control

7.1 Introduction

The success of the Inventory and Monitoring Program is dependent on the quality of the data it collects, manages, and disseminates. Analyses performed to detect ecological trends or patterns require data that are recorded properly and have acceptable precision and accuracy, and minimal bias. Poor-quality data can limit detection of subtle changes in ecosystem patterns and processes, lead to incorrect interpretations and conclusions, and greatly compromise the credibility of the program managing it.

Quality assurance (QA) can be defined 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. Quality control (QC) is a system of technical activities that measure the attributes and performance of a process, item, or service relative to defined standards (Palmer 2003). While QA procedures maintain quality throughout all stages of data development, QC procedures monitor or evaluate the resulting data products.

7.2 Objectives

The objectives of I&M Program network quality assurance and quality control are to:

- Ensure that inventory and vital signs monitoring projects produce high-quality and credible data that can be confidently used by managers, researchers, and the public
- Design, document, and implement standard quality assurance and quality control procedures that minimize or eliminate errors at every project stage.

7.3 File Naming

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

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

- Do not use spaces or special characters within a file name
- Use underscores or periods as delimiters instead of spaces where necessary
- Include a date, in yyyyymmdd format (e.g., 20061205 for December 6, 2006), to help with version control and save drafts regularly
- Include an indication of the QA/QC status (raw, draft, final, or verified etc.) of the file in the name element. For example: MIDN_20061205_DMP_Draft.doc.

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

7.4 Standards and Review Procedures

The NPS requires QA/QC review prior to communicating or disseminating data and information. Only data and information that adhere to NPS quality standards may be released. Data and information disseminated to the public



Cardinal flower (*Lobelia cardinalis*)



Wilderness Battlefield at
Fredericksburg and Spotsylvania
National Military Park

must be approved by the appropriate reviewing officials and programs. Documentation of the QA/QC standards used in producing the information and that substantiate the quality of the information must be formally certified and distributed with related data and information.

7.5 Monitoring Protocol Development

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

Although QA/QC procedures will depend upon the individual vital signs being monitored and must be specified in the monitoring protocols for each vital sign, some general concepts apply to all network projects. These general quality assurance guidelines will ensure that all data collected are checked for integrity before being integrated into monitoring program databases. As the MIDN monitoring program develops, standard QA/QC procedures will be tested and documented in relevant SOPs.

7.6 Standard Operating Procedures

The network has developed a number of SOPs, guidance documents, and review procedures for addressing quality assurance and quality control:

MIDN File Naming Guidelines

NER I&M Product Specifications

NER Inventory Product Review Work Flow

NER Vegetation Mapping Project Work Flow

Relevant Reports:

Northeast Region, National Park Service. November 2004 (revised November 2005). Natural and Social Science Study Proposal and Deliverable Guidelines. Natural Resources Report NPS/NER/NRR--2004/001. National Park Service. Philadelphia, PA.

Valley Forge National Historical Park



Chapter 8: Dataset Documentation

8.1 Introduction

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

The term “metadata” is defined as information about the content, context, quality, structure, accessibility, and other characteristics of data. In addition to ensuring data longevity, metadata increases the possibility of data sharing and reuse for multiple purposes. Creating and maintaining comprehensive metadata is neither a simple nor a quick process: it requires an up-front time investment for planning and organization, and an ongoing investment to keep it current.

8.2 Objectives

The objectives of data documentation are to:

- Ensure the longevity of data and their reuse for multiple purposes
- Ensure that data users understand the content, context, and limitations of data sets
- Facilitate discovery of data sets
- Facilitate interoperability of data sets and data exchange

8.3 Network Standards

The MIDN has adopted the following data documentation standards for all projects (i.e. inventory, monitoring, or other):

- Projects will be documented and organized in their appropriate folder using standard folder organization (see Appendix Q. MIDN Directory Structure Guidance Document).
- Spatial data files will be documented in compliance with Federal Geographic Data Committee (FGDC) standards using the Content Standards for Digital Geospatial Metadata (CSDGM; FGDC 1994).
- Biological databases will be documented in compliance with FGDC standards using the National Biological Information Infrastructure (NBII) biological metadata profile (NBII 2008).



Black bear cub (*Ursus americanus*)

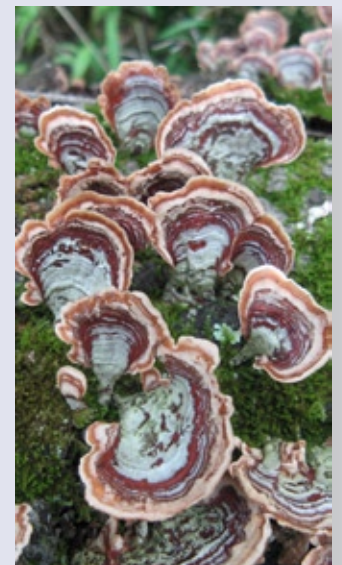
8.4 Spatial Data Documentation

The following are guidelines for spatial metadata records created by or for the MIDN:

- FGDC compliant metadata is developed for all spatial data files using the CSDGM.
- The metadata file should parse with no errors, however this is sometimes more difficult to achieve than is worth the effort.
- Generally fewer than 5 errors, none of which are associated with missing content, is what the network has treated as acceptable.

In addition to FGDC standards and the NPS Metadata Profile (including the standard NPS liability statement; NPS 2008b), the following NPS-specific information must be included in MIDN metadata records:

- Section 1: Identification Information - The primary originator should be “NPS NER Science Office,” and the secondary originator should be the cooperator.
- Section 6: Distribution Information - The distributor should be the MIDN data manager.
- For sensitive data - the Standard Order Process in the Distribution Information needs to be revised to state that the data may be requested by contacting the MIDN data manager.
- Section 7: Metadata Reference Information - The author of the metadata should be identified (usually either the cooperator, MIDN data manager, or NCSU Center for Earth Observation).
- Information on GPS settings and estimated accuracies are provided in the Data Quality element if GPS units were used to collect field information.



Bracket fungus (*Polyporus* spp.)

8.5 Tabular Database Documentation

A required element of any MIDN database is a data dictionary. The data dictionary documents descriptions of each field in the words of its creator. It also includes fully written out abbreviations, acronyms, codes, or classes for the universe of possible entries used in the associated field (even if the code appears in more than one field). The data dictionary may also include ranges of values if these are known. Data dictionaries are created either in the field description portion of a Microsoft Access database when viewed in Design View, in a Microsoft Excel table, or a Microsoft Word document. When possible, a user guide should be provided to facilitate training for data entry and database maintenance.

In addition to the above, FGDC-compliant biological metadata is developed for all biological projects using the NBII biological metadata profile. This is what the MIDN uses to document tabular biological project databases. It is also what the MIDN posts to the NPS Data Store and becomes the publicly searchable record for that project. In the case of projects with sensitive information, this is the only element that gets posted to the NPS Data Store. For other projects FGDC metadata records are zipped up along with their associated spatial data and posted for download as a zip file along with the other project data files.

8.6 Vital Signs Protocol Documentation

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

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

8.7 Standard Operating Procedures

The following Standard Operating Procedures provide guidance on data documentation:

MIDN Directory Structure Guidance Document

NER Biological Metadata Tools

Forest vegetation monitoring microplot



Chapter 9: Data Ownership and Sharing

9.1 Introduction

For the purposes of releasing data, the term “sharing” means providing data to an individual or entity upon request or as part of an agreement to do work that benefits the parks. The term “dissemination” includes releases to the public via publications or other means of delivery such as the NPS Data Store. Chapter 10 provides detailed standards and guidelines for data dissemination. Guidelines for data sharing, including sensitive data procedures, are described in this chapter.

9.2 Objectives

- The objectives of data ownership and sharing guidelines are to:
- Establish a clear understanding of the ownership of data and any associated responsibilities
- Protect sensitive data from unauthorized access and inappropriate use

9.3 Data Ownership

The MIDN relies heavily on partnerships, contracts, and agreements to collect information about natural resources. When working with multiple agencies, cooperators, and other organizations to collect and disseminate information, it is important to define the aspects of ownership prior to implementing a project. Deciding who the owner(s) of the data will be allows the project planners to determine where the information will be stored and duties will be placed, such as who will backup and archive the data, who has the overall responsibility of maintaining the data, and how scholarly publication goals can be accommodated.

All cooperative and interagency agreements, as well as contracts, should include clear provisions for data ownership and sharing as defined by the National Park Service:

- All data and materials collected or generated using National Park Service personnel and funds become the property of the National Park Service.
- Any important findings from research and educational activities should be promptly submitted for publication. Authorship must accurately reflect the contributions of those involved.

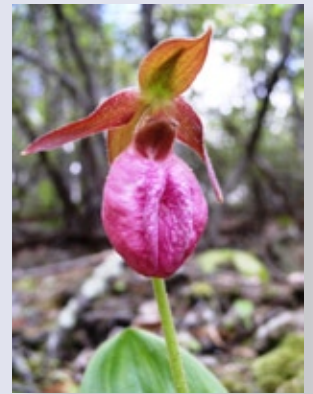
- Investigators must share collections, data, results, and supporting materials with other researchers whenever possible. In exceptional cases, where collections or data are sensitive or fragile, access may be limited.

To ensure that proper ownership, format, and development of network products is maintained, all cooperative or interagency work must be conducted as part of a signed collaborative agreement.

9.4 Data Sensitivity

The National Park Service is directed to protect information about the nature and location of sensitive park resources. Information that could result in harm to natural resources, including endangered or threatened species, can be classified as “protected” or “sensitive” and may be withheld from public release. Examples of sensitive information include the nature and/or specific locations of threatened or endangered species and habitats, minerals or paleontological objects, or significant caves.

All data and associated information from I&M Program activities will 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. The Mid-Atlantic Network staff, with input from park managers, will carefully identify any information that is considered sensitive and manage which data need access restrictions and which can be made publicly available.



Pink lady's slipper (*Cypripedium acaule*)



MIDN staff work with park staff and other cooperators on inventory and monitoring projects

Classification of sensitive I&M Program data will be a shared responsibility that includes network staff, park resource management staff, park superintendents, and investigators working on individual projects. The MIDN will complete a decision tree process that will help standardize the designation of sensitive data, and ensure that comparable guidelines are used for all parks. The decision tree should also reduce the potential for unwarranted personal bias in the classification process, and provide documentation for the rationale behind sensitive status designations.

The network staff is also responsible for identifying all potentially sensitive materials and information from principal investigators working on a project. The investigators, whether network staff, park staff, or partners, will develop procedures to flag all sensitive materials in project products, including documents, maps, databases, and metadata. When submitting products or results, investigators should specifically identify records and other references as being sensitive. Cooperators should consult with network staff to determine whether information is classified as sensitive or protected before releasing it in a public forum.

9.5 Access Restrictions to Sensitive Data

For each project, the MIDN data manager will provide a list of potentially sensitive information to the park resource manager for review.

The resource manager then determines which information should be protected. Resource managers will be responsible for working with park superintendents in this process.

Network staff will remove sensitive information from public versions of documents or other media. The network staff ensures that all references to protected information are removed or obscured in reports, publications, maps, or other public forum.

When preparing or uploading information into any network or national database, network staff ensures that all protected information is properly identified and marked. The main distribution applications and repositories developed by the program are maintained on both secure and public servers, and all records that are marked “sensitive” during uploading will only be available on the secure servers. Abridged records will be available on public servers.

9.6 Standard Operating Procedures

The following standard operating procedures provide guidance on data sharing:

Data Sensitivity and Dissemination (in development as of January 2009)

Young Virginia opossum
(*Didelphis virginiana*)



Chapter 10: Data Dissemination

10.1 Introduction

Providing well-documented data in a timely manner is a primary goal of the I&M Program, and is essential to program success. This chapter describes the methods by which quality natural resource data and information collected by an I&M Program network are made available to park managers, researchers, educators, and the general public.

10.2 Objectives

The objectives of data dissemination are to:

- Ensure non-sensitive data are easily discoverable and obtainable
- Ensure data that have not yet been subjected to full quality control are not released to the public, unless necessary in response to a FOIA request
- Distribute data with complete and accurate metadata that clearly identify who collected the data; what data were collected; and where, when, and how the data were collected
- Identify and protect sensitive data from unauthorized access

10.3 Data Distribution

The Mid-Atlantic Network's main mechanism for distribution of inventory and monitoring data will be the Internet. Use of the Internet allows data and information to reach a broad community of users. As part of the NPS I&M Program, Web-based applications and repositories have been developed to store a variety of park natural resource information. The network will use the following applications and repositories to distribute data developed by the I&M Program:

10.3.1 NatureBib

NatureBib is a master web-based database housing natural resource bibliographic data for I&M Program parks and gives users the ability to upload, and download full-text PDF documents.

The MIDN converts all reports to a format that meets the NER Technical Reporting Guidelines. Full-text versions of reports are posted to NatureBib and hard and digital copies are mailed to the Technical Information Center (TIC) in Denver, Colorado.

Documents found as a result of network data mining efforts are posted to NatureBib where copyright and other publisher agreements allow for this distribution

Sensitive data: The MIDN has produced duplicate reports for projects with large amounts of sensitive information. The public version of these reports can be posted to NatureBib, labeled under "Sensitivity Level" as "Public" for both the Citation and Abstract and made available as full-text downloads. A separate version of the report will have a NatureBib record labeled under "Sensitivity Level" as "NPS Only" and will not be searchable or viewable by the public. It is also possible to identify a citation as "Public" and its abstract as "NPS Only".



MIDN produces many reports and newsletters available to network parks and the general public

10.3.2 NPSpecies

NPSpecies is a web-based master database to store, manage, and disseminate scientific information on the biodiversity of all organisms in all I&M Program national park units.

Species records are mined from new reports and added to NPSpecies as appropriate.

Sensitive Data: At the moment the NPS Data Store does not have any built-in tools or functionality to handle sensitive data products. The MIDN approach to this has been to only post metadata (not the actual data) for the projects that have sensitive data associated with them. In these cases, the Standard Order Process section of the FGDC Distribution Information metadata element must be filled out with the MIDN data manager's contact information and instructions as to how someone would request a copy of these data.

NPS Data Store
National Park Service
U.S. Department of the Interior

Application Home
Login

Search
Select Search Type:
Online Data
and NPS Unit:
All NPS Units
Category:
All Area Categories
Word Search:
Keywords
Title
Abstract
Search
Advanced Search

Help
Instructions
About
News
Data Server
More Info
Data Store FAQs
Data Standards FAQs
Metadata Standards
NPS Metadata Profile

Links
Metadata Tools & Editor
GIS/Metadata Tools
NPS GIS Clearinghouse
(search by map)
Legacy NR FTP Search
NPSFocus
NPS IMN Program
NPS Research Permits
NPS Park Planning

Dataset Table

ID	DOI/ID	DATE/TIME	CATEGORY	SUBJECT
1	20010004	20020716	Ornithology	
2	20010004	20020827	Water Quality	
3	20010004	20010804	Water Quality	
4	20010004	20020716	Water Quality	
5	20010004	20020716	Zoology	
6	20010004	20020716	Ornithology	
7	20010004	20020716	Entomology	
8	20010004	20020716	Entomology	
9	20010004	20020716	Biology	
10	20010004	20020716	Delelatuschil	
11	20010004	20010804	Water Quality	
12	20010004	20010804	Water Quality	
13	20010004	20020716	Mammalogy	
14	20010004	20020716	Mammalogy	
15	20010004	20020716	Ornithology	

about NPS Data Store | feedback | disclaimer | privacy policy

Natural Resource GIS | NPS GIS

- If there is no known sensitive information associated with the project, perform an NPSpecies search for “Management Priority.”
- Email the park contact for that project with the results of the NPSpecies search and ask for permission to post to the NPS Data Store.
- Document park staff’s response giving permission to post in the “Projects” folder.
- Upload project products to the NPS Data Store.

NPS Focus is an agency-wide search engine and image management server being developed by the NPS Information and Telecommunications Center (ITC) that is linked to a number of other NPS repositories including the NPS Data Store and Electronic Technical Information Center (see next mechanism below). Individuals can steward their own data on this server, which has the capability to handle very large datasets.

NPS Data Store (Figure 10.1) is an agency-wide online repository for metadata and associated products that spans and is supported by many NPS programs. The NPS Data Store is the primary repository for static MIDN products.

10.3.5 Electronic Technical Information Center (ETIC)

Sensitive Data: TIC sometimes posts full-text, downloadable copies of reports, so it's important that TIC staff be notified of sensitivity levels. They also have the ability to post to InsideNPS only.

10.3.6 Northeast Region (NER) Philadelphia Science Office (PHSO)

PHSO posts full-text, downloadable versions of reports produced by the NER Science program to its website. Currently, a Pennsylvania State University cooperator, maintains this site under the direction of the NER chief scientist. At the moment this site is not searchable. If reports are posted on the network website, a link to the site will appear on the PHSO site so there is no duplication.

10.3.7 U.S. Geological Survey (USGS)-NPS Vegetation Mapping Website

North Carolina State University (NCSU), a MIDN cooperator, is working with the USGS Vegetation Mapping Program to post final vegetation mapping data to a USGS website in USGS-approved formats.

10.3.8 Mid-Atlantic Network Website

The network website provides detailed information about the MIDN and the I&M Pro-

gram. Originally envisioned as a place to post network data, the MIDN website is no longer being developed in this capacity. The MIDN website has information on MIDN projects, cooperators and partners, the Board of Directors (BOD) and Science Advisory Committee (SAC), photos of network activities, and downloadable network and programmatic administrative documents.

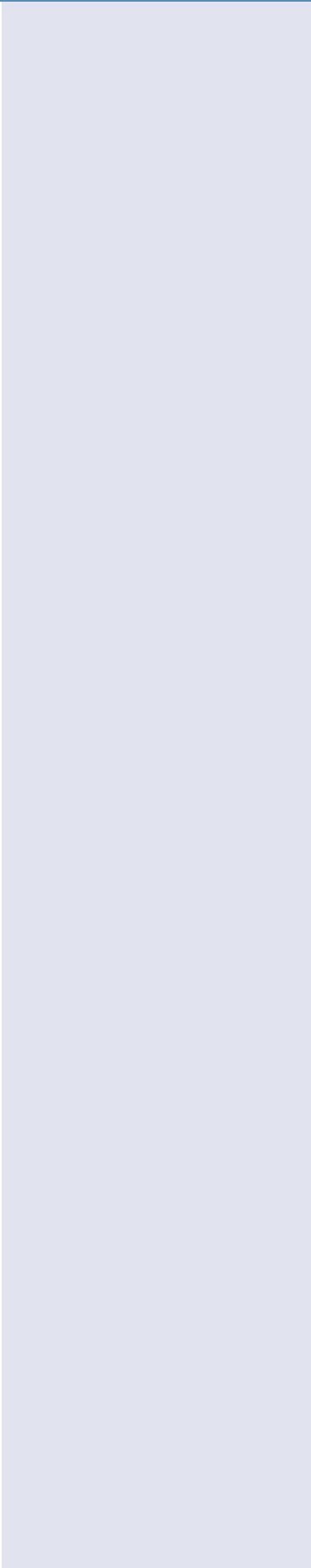
Sensitive Data: The MIDN plans to include the “NPS Only” version of reports on the intranet website in the future.

10.4 Feedback Mechanisms

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



Valley Forge National Historical
Park



Chapter 11: Records Management and Object Curation

11.1 Introduction

I&M Program networks are responsible for managing the documents, specimens, photographs, digital files, and other products resulting from network projects, administration, and activities. Much of the information generated by I&M Program networks will only increase in value and applicability over time. The potential for loss of data, documents, or objects can come from a variety of sources, including catastrophic events (e.g., fire, flood, and earthquake), user error, hardware failure, software failure or data corruption, theft, and intentional acts of vandalism. A fundamental responsibility of network data management is to establish procedures that will guarantee that I&M Program information will be accessible and usable for future generations.

11.2 Objectives

The objective of I&M Program network records management, data maintenance, and archiving procedures is to:

- Ensure that I&M Program network information can be easily found, obtained, shared, and properly interpreted by a broad range of users in perpetuity.

11.3 Digital Data Maintenance

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

The MIDN has entered into a cooperative agreement with the North Carolina State University NPS Field Technical Support Center (NCSU FTSC) for the maintenance, storage, and archiving of data sets produced as part of the network's inventory, vegetation mapping, and vital signs monitoring programs. The maintenance and storage of digital data developed by network and regional I&M Program staff stationed at Fredericksburg and Spotsylvania NMP are housed on the network server. These files include the network's administrative and working data files.



Jack o'lantern mushrooms
(*Omphalotus illudens*)

Digital data maintained in the MIDN data archive at NCSU FTSC are expected to result from one of two types of projects:

1. Short-term projects, after data collection and modification have been completed (i.e., biological inventory and vegetation mapping projects)
2. Long-term monitoring projects, where data acquisition and entry will continue indefinitely

Following the lead of the National Park Service and the national I&M Program, the MIDN has adopted Microsoft Access as its database standard and ESRI ArcGIS as its spatial data management standard. The MIDN will update its versions of these software programs as necessary to remain current and compatible with National Park Service and national I&M Program standards.

Technological obsolescence is a significant cause of information loss, and data can quickly become inaccessible to users if they are stored in out-of-date software programs or on out-moded media. The network does maintain a folder on the network server to archive software. Effective maintenance of digital files depends on the proper management of a continuously changing 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). Thus, MIDN archiving pro-



Chatham Manor, Fredericksburg and Spotsylvania National Military Park

cedures include saving data sets in both their native format (typically MS-Access database or Excel spreadsheet) and as sets of ASCII text files. As a platform- and software-independent format, ASCII text files ensure future usability of the data in a wide range of applications and platforms. In addition, data sets will periodically be converted to upgraded versions of their native formats.

11.3.1 Short-Term Data Sets

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

When a project is completed, NCSU FTSC staff prepare the tabular data for archiving by creating:

- A set of ASCII comma-delimited text files for the tabular data files and tables comprising the data set
- An XML file that preserves relationships between tables for each MSAccess database
- A readme.txt file that explains the contents of each ASCII file, file relationships, and field definitions

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

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

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



Rough green snake (*Opheodrys aestivus*)

11.3.2 Long-term Monitoring of Data Sets

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

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

Network staff are responsible for preparing tabular data from the network long-term monitoring projects for archival following completion of standard QA/QC procedures. The procedures listed earlier under “Short-Term Data Sets” above will also be used for long-term data sets.

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

Table 11.1 describes the fields required when network staff submit digital biological data sets or products to NCSU FTSC for archiving. This simple tracking system will be made available on the Web for all network staff.

11.3.3 Spatial Data

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

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

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

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

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

Table 11.1. Digital Data Tracking Database.

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

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

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

11.4 Storage and Archiving Procedures for Digital Data

The archiving procedures described below are designed to ensure both the security of

and ready access to MIDN data in perpetuity. These procedures apply to both short-term and long-term monitoring project data sets (tabular and spatial data) and to digital aerial imagery acquired by the MIDN.

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

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

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

For digital aerial photography, this information includes:

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

Upon receipt, NCSU FTSC staff will:

1. Check the digital media for physical damage

2. Count and record the number of media received
3. Count and record the number of data/image files received
4. Check data file integrity (i.e., make sure that each file opens and displays appropriate data)

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

NCSU FTSC staff will enter the information about short-term and long-term monitoring project data into an Excel spreadsheet, referred to as the Digital Data Tracking Database (see Table 11.1). Information about sets of digital aerial photography should be entered into a similar spreadsheet, the Aerial Imagery Tracking Database (see Table 11.2).

Copies of the Digital Data Tracking Database (Table 11.1) and the Aerial Imagery Tracking Database (Table 11.2) will be sent to the network once a year or as requested. In the future, these databases will be posted on the Internet for further accessibility to Network staff. These databases will be backed up following the procedures used for backing up archived data described below.

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

Tagged and marked trees photographed during vegetation monitoring



11.4.1 Storage Facilities for Electronic Archives

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

11.4.2 Directory Structure for Electronic Archives

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

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

- Programmatic documents including protocols, procedures, and supporting documents
- Interim data sets or milestones
- Data sets submitted by contractors
- Data sets—in ASCII format
- Conceptual or statistical models used for data interpretation
- Final and NPS technical reports
- Read-me files—including an explanation of directory contents, project metadata (including a data set catalog report), and version documentation

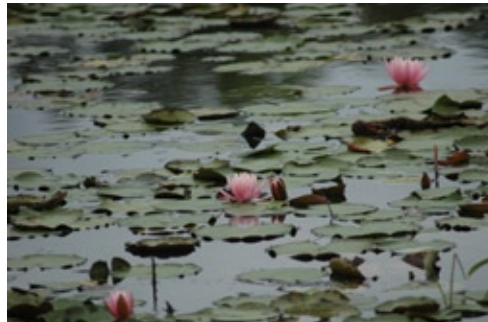
After final data and reports have been submitted, draft products will not need to be maintained.

11.4.3 Backup Procedures for Digital Data

The risk of data loss can come from a variety of sources, including catastrophic events (e.g., fire, flood), user error, hardware failure, software failure or corruption, and security breaches and vandalism. Performing regular backups of data and arranging for off site storage of backup

sets are the most important safeguards against data loss.

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



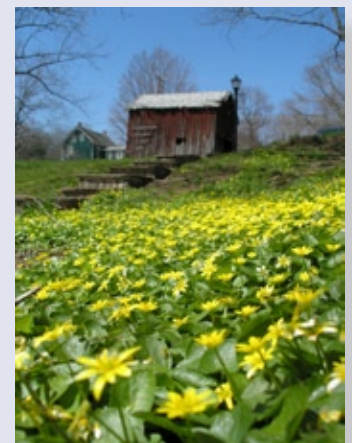
Water lilies (*Nympha* spp.)

Data on the network server stationed at Fredericksburg and Spotsylvania NMP is backed up using server software, BackupAssist Version 5.1. Full backups are performed monthly and incremental backups weekly. Backups are stored on two portable external hard drives which are rotated offsite every month. Network data and information produced by staff on workstations are backed up by individual staff members. Data are stored both on site on the network server and at off site locations using portable, external drives.

11.4.4 Data and Network Security

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

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



Valley Forge National Historical Park

Aerial photograph of City Point at Petersburg National Battlefield



Eastern fence lizard (*Sceloporus undulatus*)

11.5 MIDN Storage of Hard-Copy and Digital Reports

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

1. MIDN central files, Natural Resource Office, Fredericksburg, Virginia

These files contain project files, copies of administrative documents, and non-record copies of documents. Examples include: meeting minutes, correspondence, memoranda of understanding, contracts and agreements, research permits, and interim and selected final reports produced by the program or under its auspices. The MIDN will use acid-free paper and folders for all permanent records in the central files. In addition to maintaining these paper records, the MIDN will maintain electronic versions when possible. These will be housed and backed up on a regular schedule on the network server.

2. Technical Information Center (TIC), Denver, Colorado

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

11.6 Specimens

The network will provide specimens collected under the auspices of its activities to the park in which they were collected for inclusion in a park museum, or to a repository approved by the park. Specimens at non-NPS repositories are considered on loan and remain property of the National Park Service. MIDN will provide park curators with associated data required for cataloging each specimen in the Automated National Catalog System (ANSC+). Currently, legacy data are only entered into the NPSpecies database. All specimen information collected as part of the network's I&M Program is entered into NPSpecies.

11.7 Photographs

The MIDN prefers all images (photographs and scanned documents) be delivered as TIFF or Joint Photographic Experts Group (JPEG) files, as is specified in NPS Focus.

Though the MIDN doesn't currently have an image management system in place, it will eventually develop and adopt such as system,

preferably one that allows search ability of image subject and project and tracks the location of image files on local machines or servers.

Project photographs are stored and archived with the project data. Part of the archiving process is identifying photos for the network library. These photos are duplicated and maintained in the library of photos on the network server.

All digital photos relating to I&M Program projects will be stored in a sub-directory under that project's directory. They will remain in the project's active directory until they have been added to the network's photo management system and received metadata documentation at which time they can then be archived.



11.7.1 Aerial Photographs

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

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

Then NCSU staff will proceed as follows:

1. Upon receipt, NCSU FTSC staff count the aerial photos received and check each photo for physical damage. If the number of photos does not match the transmitted number, or if any photos are physically damaged (e.g. torn, water stained, etc.), FTSC staff will notify the network.
2. FTSC enters information about the set of aerial photos (as appropriate and as it becomes known) into an Excel

spreadsheet, referred to as the Air Photo Tracking Database (see Table 11.2)

3. FTSC sends a copy of the Air Photo Tracking Database (Table 11.2) to the regional I&M coordinator and network coordinator every January. In the future, this database will be posted on the Internet.
4. Aerial photographs are then stored in plastic sleeves, lying flat, in a fireproof cabinet, in a dark, cool, temperature-controlled environment.

11.8 Role of Curators in Storage and Archiving Procedures

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

11.9 Standard Operating Procedures

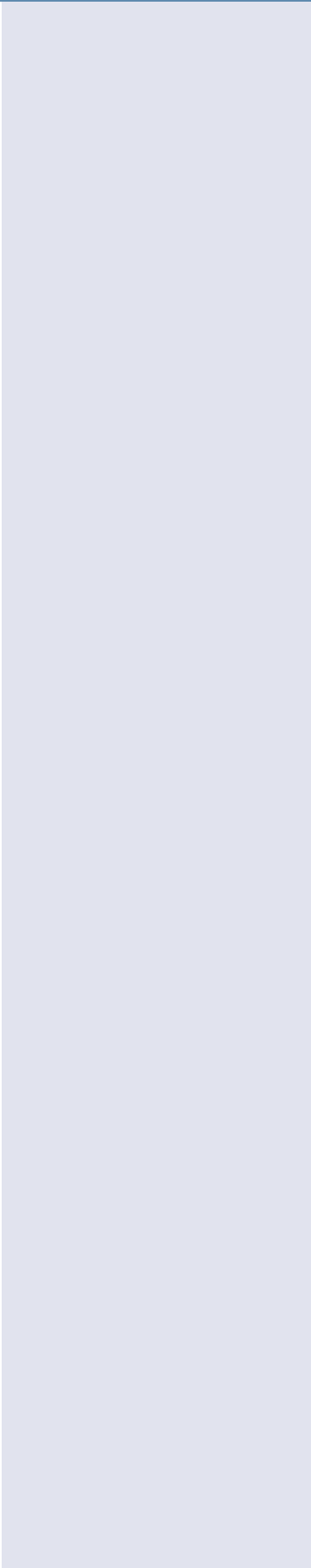
The network has developed the following SOP for records management:

NER NPSpecies Museum Voucher Fields

Photo Management Guidelines (Under development as of January 2009)

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

Table 11.2. Air Photo Tracking Database



Chapter 12: Project Tracking and Documentation

12.1 Introduction

I&M Program networks face the challenge of managing multiple projects, contracts, and agreements; budgets; timelines and deliverables; and permanent, term and temporary staff. As networks develop and become fully staffed, many have found that a project tracking and documentation system is necessary to function effectively.

12.2 Objectives

The objective of project tracking and documentation procedures are to:

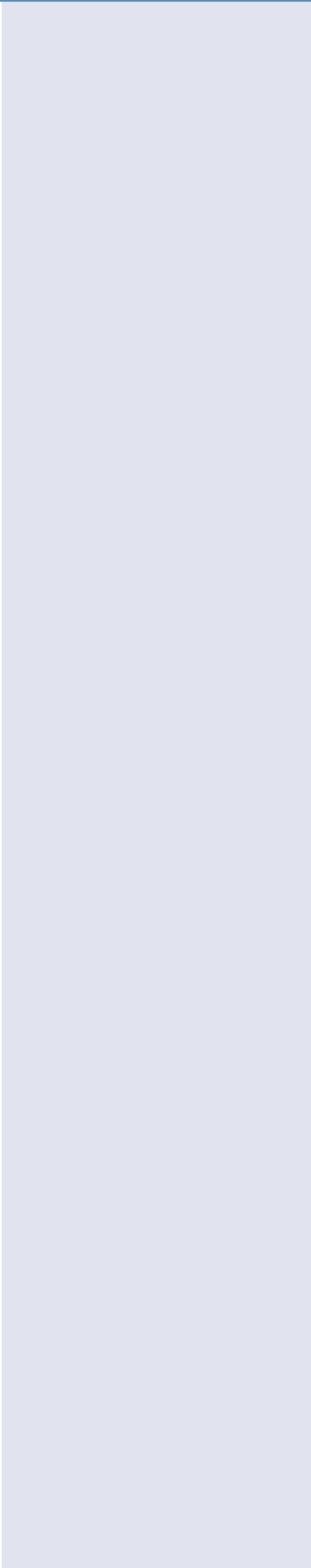
- Efficiently manage and coordinate contracts, permits, agreements, and associated timelines, schedules, and budgets
- Ensure project products are delivered on time and per specifications
- Manage information needed for annual reporting and accountability
- Improve accountability for I&M Program network natural resource inventory and monitoring efforts and provide a means of tracking accomplishments

12.3 Project Tracking

The Mid-Atlantic Network currently maintains a list of projects to track the status of projects and project deliverables. This information is available on the shared network drive so that both the network coordinator and data manager have access.



Wild turkey chick (*Meleagris gallopavo*)



Chapter 13: Implementation

13.1 Introduction

The Inventory and Monitoring Program data management plans contain practices that may be new to many NPS staff and cooperators. While this document and the accompanying SOPs offer practical guidelines and solutions, successful implementation will require continuous formal and informal education of staff and cooperators, and a long-term and disciplined commitment on the part of data managers and network staff to make it all work.

13.2 Objectives

The objectives for a plan to implement I&M Program network data management practices and procedures are to:

- Ensure that the procedures outlined in this plan and related SOPs are put into practice
- Provide a framework to realistically assess related network procedures and adjust them as needed
- Establish benchmarks so that implementation progress is measured and evaluated

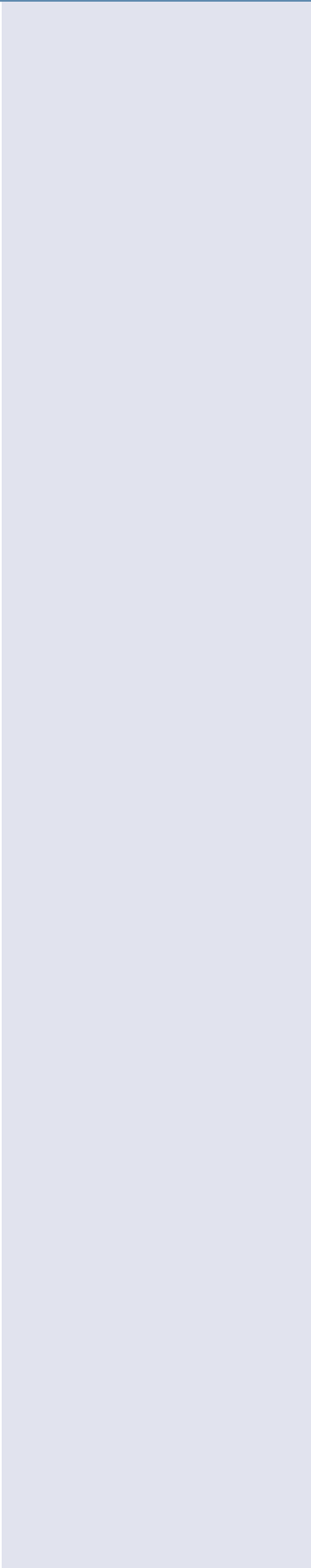
13.3 Network Implementation

The national data management plan and this document serve as the foundation for good data management practices within the network. All network staff are encouraged to read and follow the guidelines outlined in this plan.

If you have questions and/or concerns, please contact the network data manager. Good network data management practices are a collaborative effort.

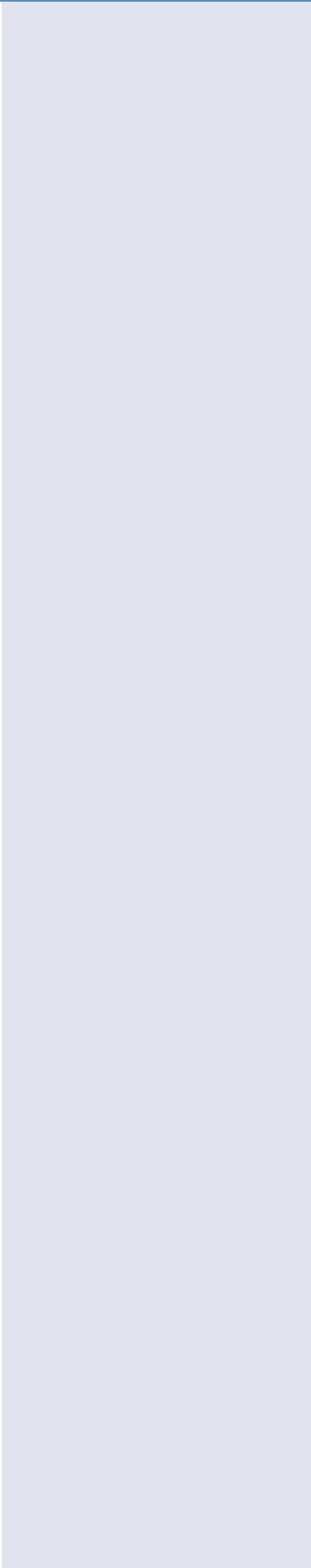


Great egret (*Ardea alba*)



Literature Cited

- Federal Geographic Data Committee, 1994. Content standards for digital spatial metadata (June 8 draft). Federal Geographic Data Committee. Washington D.C.
- Michener, W. K. 2000. Metadata. Pages 92-116 in Michener, W. K. and J.W. Brunt, editors. Ecological data: design, management, and processing. Blackwell Science Ltd. Methods in Ecology Series, Malden, Massachusetts
- National Biological Information Infrastructure (NBII). 2008. NBII Metadata Activities. (<http://metadata.nbii.gov/portal/server.pt>). Accessed 16 January 2008.
- National Park Service. 1998. National Parks Omnibus Act of 1998. Public Law 105-391. (<http://www.nps.gov/legal/laws/105laws.htm>)
- National Park Service. 2000. National Park Service Strategic Plan FY 2001-2005. NPS D-1383/August 2000. (http://planning.nps.gov/document/NPS_strategic_plan.pdf)
- National Park Service. 2008a. Data management guidelines for inventory and monitoring networks. Natural Resource Report NPS/NRPC/NRR—2008/035. National Park Service, Fort Collins, Colorado.
- National Park Service. 2008b. NPS Metadata Profile. (<http://science.nature.nps.gov/nrdata/docs/npsprofile.cfm>). Accessed 16 January 2008.
- Palmer, C. J. 2003. Approaches to quality assurance and information management for regional ecological monitoring programs. Pages 211-225 in Busch, D. E. and J. C. Trexler, editors. 2003, Monitoring Ecosystems: interdisciplinary approaches for evaluating ecoregional initiatives. Island Press, Washington, D.C.



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

NPS D-68, December 2008

National Park Service
U.S. Department of the Interior



Natural Resources Program Center

1201 Oakridge Drive, Suite 200
Fort Collins, Colorado 80525

www.nps.gov