Data Management Plan
Northeast Temperate Network

Northeast Temperate Network
Marsh-Billings-Rockefeller National Historic Park
54 Elm Street
Woodstock, Vt 05091
802-457-3368

By: Fred Dieffenbach
Date: September 24, 2006
Version: 1.18

Northeast Temperate Network Signatories

_____

Elizabeth Johnson, Regional I&M Coordinator Date
Northeast Region

_____

Brian R. Mitchell, Network Coordinator Date
Northeast Temperate Network

_____

Fred W. Dieffenbach, Data Manager Date
Northeast Temperate Network

Acknowledgements

The Northeast Temperate Network Data Management Plan is a product of the efforts of many, not one person in particular. While the approach to the Northeast Temperate Network plan may appear to differ substantially from the direction taken by other Networks, our decision to take this path was based upon the many conference calls, template documents, and general conversations from the last 2 ½ years. We have borrowed liberally from existing resources, in some cases literally capturing passages of text from previous efforts because we found it impossible to improve upon the way certain ideas were presented. In other cases, we have borrowed from the ideas expressed by others and have put them into words that represent the way the Northeast Temperate Network intends to solve the problem. Finally, in some cases we learned the most from ideas with which we differed.

The Network team that has contributed in various ways to this plan includes Greg Shriver, Theresa Moore, and Brian Mitchell, as well as the data management staff from the other offices in the Northeast Region. The Northeast Region Data Managers who deserve special recognition include: Sara Stevens, Nate Piekielek, Kristina Callahan, Velma Decker and in particular Bill Gawley at Acadia National Park.

Thanks also go to the following individuals from outside the Northeast Region for their input and assistance: John Boetsch, Gordon Dicus, Brent Frakes, Lisa Nelson, Dorothy Mortenson, and Nicole Trancreto.

I would be remiss if I failed to acknowledge the assistance offered by Rob Daley from the Greater Yellowstone Network who volunteered to review a draft of this plan in advance of the submission date in 2005. The information Rob has presented at our national conferences has always impressed me, and I was particularly pleased when I learned that he offered to review the NETN plan. Rob's comments on the plan were as insightful and valuable as the information he presents at national I&M meetings. I am also grateful to the anonymous data managers who reviewed the NETN plan following its official submission.

Finally, Margaret Beer has helped this Network and every other Network in more ways than I can count. Since making to move from the Southern Colorado Plateau Network Data Manager to the National Program Data Manager she has helped bring organization and structure to a program that needed direction. When I asked her to take a look at the draft plan I was developing, she did; when I asked her if it would be possible to assemble a preliminary review team in the fall of 2005 she did that too. I think I speak for all Network Data Managers when I say, Thanks Margaret!

Fred Dieffenbach
Data Manager/Biologist
Northeast Temperate Network

Table of Contents

Acknowledgements ... 3
Executive Summary ... 7
Objects and Goals ... 7
Capabilities and Organizational Infrastructure ... 7
Roles and Responsibilities ... 8
Project Management ... 8
Database Design ... 8
Data Acquisition and Management ... 9
Quality Assurance and Quality Control ... 9
Documentation ... 9
Data Analysis and Reporting ... 10
Data Dissemination ... 10
Records Management and Archiving ... 10
1. Introduction ... 11
Overview of Parks and Natural Resources ... 11
2. Capabilities and Organizational Infrastructure ... 13
Background ... 13
Capabilities ... 13
Geographic Information System (GIS) Support ... 13
Relational Database Support ... 14
Document Preparation ... 14
Data Integration ... 15
Data Acquisition ... 15
Organizational Infrastructure ... 15
Directory Structure ... 16
File Naming ... 17
System Back-up ... 17
Budget Tracking ... 17
Hardware and Equipment ... 17
Computers and related hardware ... 17
Global Position System (GPS) Receivers ... 18
Digital Cameras ... 18
Water Quality Sampling Equipment ... 18
3. Roles and Responsibilities ... 19
Overview ... 19

Roles and Responsibilities 20
Coordination 20
Communication 22
Awareness and Accountability 22
Data Stewardship 22
Documentation 22
Network Implementation 22
4. Project Management 24
Background 24
Communication 24
Planning and approval 24
Project Tracking 25
Project Budget 26
Project Design 27
Project Testing 27
Project Implementation 28
Preparation 28
Data acquisition and processing 28
Product delivery and review 28
Product Integration 28
Evaluation and closure 29
5. Data Acquisition and Processing 30
Background 30
Acquiring and Processing Field Data 30
Programmatic Data 30
Non-Programmatic Data 31
Species Data 32
Bibliographic Data 32
Specimen Collection data 32
Long-term monitoring 32
Water Quality Data Integration 32
6. Quality Assurance and Quality Control 33
Background 33
NPS Standards 33
Quality Control 33
Quality Assurance 34
Verification 34

Validation 34
Version Control 34
Data Quality Process Review and Communication 35
7. Documentation 36
Metadata Tools 36
ArcCatalog: 37
Metadata Parser: 37
Flow Charts 37
Documentation Process 38
Project Documentation 38
Data Types 39
8. Data Dissemination 41
Background 41
Data Disclosure 42
Data Ownership 42
Guidelines and Statutes 42
Data Release Policy 42
Internet Accessibility 43
Natural Resource and GIS Data Store 43
Feedback 43
9. Records Management and Archiving 44
NPS Standards 44
Network Standards 44
Curatorial Resources 45
10. Glossary 46
11. References 49

Executive Summary

Objects and Goals

The goals of Northeast Temperate Network data management are to provide accurate, efficient, and effective information and support for 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 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 where data is stored; its quality, timeliness, and uses of the data; how to incorporate these data into resource management decisions; and, how the data will be managed over time.

The NPS Strategic Plan, 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 park*s…" is a desired outcome. The objective of the NPS I&M Program is to provide scientifically and statistically sound data for resource management, and to ensure that quality data is available for this task. These objectives establish needs to:

- Develop metadata for all significant spatial and non-spatial data;
- Ensure that all significant data are of "very high quality";
- Develop and maintain all essential data;
- Ensure that data are logically organized and retrievable by staff, cooperators, and the public;
- Identify sensitive data and protect it from unauthorized access and inappropriate use;
- Optimize data sharing, development, and analyses; and,
- Ensure that all Network held digital and non-digital information (i.e. data sheets, documents, published and unpublished reports, manuscripts, photographs, maps, metadata, etc.) are archived and protected in accordance with recognized and required archival standards.

Capabilities and Organizational Infrastructure

In the context of information technology, infrastructure refers to the utilities, hardware, software, user training and support systems that keep the information system running. Accordingly, this section describes the systems, programs, policies, and capabilities that the Network has established to provide the data management services and support to Parks, and to cooperators working at Parks within the Network.

The Network has identified five distinct data management capabilities it will offer to parks within the Network: Geographic Information System support, relational database support, document preparation support, data integration, and data acquisition. In addition, the network will work with parks to manage any datasets they may possess, and will assist them with all data management needs including issues relating to data collection, storage, and stewardship.

The Network has also established a series of standards and policies that relate to the organization of Network and Park data holdings and to the long-term security of Network data. For example, the Network has established a naming convention, a directory structure, a comprehensive data storage procedure, a data back-up process, and a budget tracking system.

Finally, the Network has acquired computer hardware and equipment to complete its mission. This includes Global Positioning System (GPS) receivers, water quality sampling equipment, and digital cameras. Additionally, the Network has a number of desktop, laptop, and hand-held computers and digital storage devices. A complete listing of computers and related equipment can be found in the equipment Appendix or by linking to the Network equipment database (not presently linkable).

Roles and Responsibilities

The Network's staff is the 'eyes and ears' of the Network. A knowledgeable staff that knows what to do and when to do it are vital to the success of the Network.

In January 2003, the Northeast Temperate Network hired a Data Manager to oversee issues related to data acquisition, organization, security, access, dissemination, and documentation. Beyond data stewardship, the Data Manager works with cooperators and park staff on database design and standardization issues, is responsible for determining whether data sets are complete enough for inclusion into master NPS data systems, and evaluates field data forms and data entry modules. The Network Data Manager is also primarily responsible for determining the data management roles and responsibilities for every project.

To help the Network team coalesce, the Network has adopted a framework that identifies key data tasks and the primary person who must ensure that each task has been completed. The underlying philosophy behind the various Roles and Responsibilities identified by the Network is shared responsibility and cooperation. The Network believes that all staff, from field technician to Network Coordinator, is equally responsible for ensuring that data collected by the Network are scientifically and statistically sound.

Project Management

Data management begins with the conception and design of a project and continues until the desired end product is made available to the intended audience. The value of good data management is fully realized when data is readily accessible to a broad audience, and when that data fulfills the objectives of the project. To achieve this level of performance, the Network has established guidelines for the project management process, from inception to completion. The guidelines stress the importance of clearly defining the purpose and objectives of a project. Without these fundamental building blocks, it is neither possible to evaluate the success of the project nor is it possible to determine the utility of the data because the purpose of the project is unknown. The Network also stresses the importance of tracking each project's progress, and on performing a post project-completion evaluation

The key project management elements identified by the Network that must be addressed with every project include:

- Planning and Approval
- Project Tracking
- Project Budget
- Project Design
- Project Testing
- Project Implementation
- Preparation
- Data Acquisition and processing
- Product Delivery and Review
- Product Integration
- Evaluation and Closure

Database Design

Consistency and compatibility are two important keys to ensuring high quality data. If data collected by the Network are intended to be used by park mangers, network staff, the public, and/or the scientific community, the data the Network collects must be high quality. The task of ensuring high quality data is made more difficult (if not impossible) if the Network does not implement rigorous database standards. While database standards alone will not solve all possible problems, standards promote compatibility among data sets, making it easier to aggregate and summarize data in the future.

Designing an appropriate database is more dependent on communication than it is on database programming acuity. Accordingly, the Network stresses the importance of remaining involved with each database development project instead of establishing a prescriptive step-by-step process that must be followed during the development process. This philosophy notwithstanding, defining the purpose for a database is one step that cannot be overlooked, and must be established at the outset of a database design project.

With respect to standards that do exist, the National Park Service's Inventory and Monitoring Program has developed the Natural Resource Database Template (NRDT). The Network will use the NRDT as the preferred framework for all future natural resource database development projects, while standards and expectations established by the Network are detailed in the Database Specification SOP.

Data Acquisition and Management

The Northeast Temperate Network intends to acquire and maintain a complete record of natural resource data for all parks within the Network. The Network may also acquire data that is not associated with parks, but is regionally focused or related to park activities.

Digital data shall be stored by the Network and made available to cooperators, park and/or Network staff, and others in compliance with established data distribution policies. Data that is properly documented with metadata and that is free of data distribution restrictions will be posted to the NRGIS-Data Store where it can be accessed by the broadest audience. Data that is not documented with metadata (or that has data distribution restrictions) will also be acquired by the Network, but the Network will not distribute inadequately or improperly documented data, or data that has distribution restrictions. Historic data, in formats other than digital will also be obtained, when available, and scanned into digital format. These data will then be made available to cooperators, park and/or Network staff, and others in compliance with established data distribution policies.

Data that is generated through Network activities will be permanently stored and archived along with all other project-related information. Data that is not generated through Network activities will generally not be permanently archived by the Network.

Quality Assurance and Quality Control

Data collected through monitoring activities must be uniform, consistent, and accurate if they are to serve the needs of the Inventory and Monitoring program and resource managers. If data do not meet these requirements, analysis and decisions based upon these data may be flawed, and could produce unwanted results and promote poor decisions. To ensure that data quality problems do not produce these undesirable consequences, the Network has established a program to ensure that data generated through Network activities are of known quality. The Network quality assurance and quality control (QA/QC) program relies on the following to deliver high quality data:

- Thoroughly evaluated scientific measurement protocols;
- Standard operating procedures;
- Verification, validation, and editing procedures;
- Data documentation and metadata standards;
- Version control; and,
- Data quality process review and communication.

Documentation

Documentation brings a project to completion by fully describing the process, limitations, application, and restrictions that might apply to a project or dataset. It makes it possible to repeat a project, and thorough documentation includes guidance on how to appropriately use a dataset.

While documentary requirements may vary depending on whether it applies to a dataset, a database, an application, or a project, it will in all instances provide a roadmap to proper usage and understanding.

Beyond the obvious reasons for documenting a project, Executive Order 12906 (April 1994) mandates that federal agencies create metadata, or "information about data," for all geospatial data. The Network intends to comply with the requirements of this Executive Order, and will ensure that all projects administered by the Network, including those that do not generate geospatial data, are fully documented with metadata and appropriate guidance.

Data Analysis and Reporting

Presenting meaningful information in a manner that is beneficial to managers and scientists is a fundamental objective of the Inventory and Monitoring Program. For the Network to achieve this objective, the intended analytical and reporting requirements must be clearly defined by the project leaders and described in the appropriate project standard operating procedure documents. The Network data management program will support this objective by ensuring that data necessary for the specified analysis are properly formatted and compatible with applicable statistical software applications.

Data Dissemination

Data collected, maintained, and/or stored by the Northeast Temperate Network will be entered into the appropriate National Park Service "national" data system. This may be any combination of the following systems: NPSpecies, NatureBIB, Dataset Catalog, ANCS+, NPSTORET, and, the NRGIS Data Store. Data may also be presented through the Network web page or other means by special request.

Prior to disseminating any data, the Northeast Temperate Network will work with cooperating agencies, organizations, and individuals to protect the security of any and all sensitive data. The Network will implement the Regional Freedom of Information Act (FOIA) policies, and will place special emphasis on procedures for handling sensitive data.

Records Management and Archiving

The Network is responsible for maintaining and archiving documents such as final reports prepared by staff and/or contractors, program administrative documents, contracts and agreements, memoranda of agreements, and other documents related to Network administration, activities, and projects. The Network must also manage and archive physical items such as natural history specimens, photographs, and audio tapes. Finally, the Network must permanently archive all data obtained during Network activities. A complete discussion of the Network's intentions regarding records management and collections is outlined in the Network Scope of Collections statement. All Network data will be archived on CD, DVD, tape, or other appropriate media and stored at Acadia National Park.

Storage for many of the aforementioned items is prescribed in NPS Director's Order 19: Records Management and associated appendices. However, for things such as data that may be software dependent, proper procedures for long-term archiving do not currently exist. In these instances, the Network will work with the curator at Acadia National Park to develop the best long-term solution to the data archiving problem.

1. Introduction

The goals of the Northeast Temperate Network data management plan are to provide accurate, efficient, and effective information and support for 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 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, where it is stored, the quality, timeliness, and uses of the data, how to incorporate data into resource management decisions, and how the data will be managed over time.

The National Park Service (NPS) Strategic Plan, 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 NPS Inventory and Monitoring Program is to provide scientifically and statistically sound data for resource management, and to ensure that quality data is available for this task. These objectives establish a need to:

- Create metadata for all significant spatial and non-spatial data;
- Guarantee that all significant data are of "very high quality";
- Develop and maintain all essential datasets;
- Ensure that data are logically organized and retrievable by staff, cooperators, and the public;
- Identify sensitive data and protect it from unauthorized access and inappropriate use;
- Optimize data sharing, development, and analyses; and,
- Ensure that all Network held digital and non-digital information (i.e. data sheets, documents, published and unpublished reports, manuscripts, photographs, maps, metadata, etc.) are archived and protected in accordance with recognized archival standards.

The following plan is intended to establish a framework that will guide Network data management planning and implementation. The plan is intended to be 'stand-alone' because it addresses subject matter that is not included in the Network ecological monitoring plan. However, the data management plan is not intended to 'stand-alone' outside the context of the Network ecological monitoring program -- the plan is complementary to the Network monitoring plan, not independent of it. The data management plan specifically excludes most details regarding Network resources and park composition since those elements are detailed in the ecological monitoring plan.

The Data Management Plan is intended to be conceptual, providing a general guide to the fundamentals of data management, and identifying the primary data management issues and concerns. It is not intended to provide step-by-step solutions to specific data management issues. Specific solutions to data management issues are addressed in focused appendices and Standard Operating Procedure (SOP) documents that are, by definition, targeted toward a particular subject area. For example, the data management plan will identify the general types of software resources used by the Network, but will not incorporate a detailed listing of all software applications used by the Network. A software inventory is too dynamic to effectively present in the plan, and is more appropriately listed in an appendix that is easier to develop and keep up-to-date than the data management plan itself. Similarly, the Network data management plan will discuss the fundamental principals behind data Quality Assurance and Quality Control (QA/QC), but the stepwise process that Network staff will follow to assure high quality data will be relegated to project specific SOP's that can be updated more immediately in response to evolving technology, regulatory changes or policy mandate, changes to data collection procedures or standards, or any other factor that dictates a change in QA/QC procedures.

Overview of Parks and Natural Resources

The Network contains 10 parks (Table 1.1), and is coordinating Inventory and Monitoring activities for the five networks through which the Appalachian National Scenic Trail (NST) passes.

The Appalachian Trail is on a different schedule and implementation plan from the rest of the Network, and is being treated as a separate monitoring entity. In contrast to the other parks, monitoring along the Appalachian Trail will primarily involve analyzing information from existing sources and identifying information gaps.

The 10 Network parks contain diverse cultural and natural resources and span two ecological divisions (Laurentian / Acadian and Central Interior & Appalachian). Parks within the network range geographically from Acadia NP in coastal Maine to Morristown NHP in central New Jersey (Table 1.1), and range in size from 4 hectares (9 acres) at Saugus Iron Works to more than 19,000 hectares (47,000 acres) at Acadia.

Table 1.1. Parks included in the Northeast Temperate Network.

Park Name (state)	Park Acronym	Acres	Hectares
Acadia NP (ME)	ACAD	47,498	19,229
Appalachian NST (GA-ME)	APPA	227,001	91,903
Boston Harbor Islands NPA (MA)	BOHA	1,465	593
Marsh-Billings-Rockefeller NHP (VT)	MABI	643	260
Minute Man NHP (MA)	MIMA	967	391
Morristown NHP (NJ)	MORR	1,707	691
Roosevelt-Vanderbilt NHS (NY)	ROVA	778	315
Saint-Gaudens NHS (NH)	SAGA	150	61
Saratoga NHP (NY)	SARA	3,392	1,373
Saugus Iron Works NHS (MA)	SAIR	9	4
Weir Farm NHS (CT)	WEFA	74	30

These parks include the beginning and end of the Revolutionary War (Minute Man and Saratoga respectively) and a strategic military location for General George Washington (Morristown). Two National Historic Sites commemorate the lives of artists (Saint-Gaudens and Weir Farm) and Roosevelt-Vanderbilt celebrates the “Gilded Age.” Marsh-Billings-Rockefeller and Boston Harbor Islands are both new to the NPS and unique in their establishment and mandates. Marsh-Billings-Rockefeller is the only national park to focus on conservation history and the evolving nature of land stewardship. Boston Harbor Islands, established in 1996, is a diverse collection of 34 drowned drumlins (elongated whale-shaped hills formed by glacial action) in the Massachusetts Bay. Boston Harbor is also unique in that the National Park Service does not own any of the land on which the park is located, but presides over a 13-member management partnership. Saugus Iron Works marks the site of the first integrated iron works in North America, which gave rise to the industrial revolution. Acadia is the only National Park in the Network and hosts a diverse array of cultural, natural, and geologic resources. The Appalachian Trail, which crosses some of the most diverse ecological communities in the Northeast, is managed by a unique partnership between the NPS and the Appalachian Trail Conservancy, and provides an exciting opportunity for ecological monitoring across 2,100 miles of habitat representative of the entire east coast of the US.

2. Capabilities and Organizational Infrastructure

Background

Infrastructure is broadly defined to mean the basic structure or features of a system or organization. In the context of information technology, infrastructure typically refers to the utilities, hardware, software, user training and support systems that keep the information system running. For the purpose of this document, organizational infrastructure refers to the combination of equipment and capabilities maintained by the Northeast Temperate Network Data Management Program and the services the network provides to its constituent parks.

With respect to the infrastructure requirements incorporated in the 'information technology' focused part of the definition, those elements are defined and managed by the Northeast Region Office of Information Technology, which in turn ensures that computing systems in the northeast region of the NPS are in compliance with all Agency and Departmental information technology requirements. NPS IT requirements are documented on the following intranet site:

http://inside.nps.gov/waso/waso.cfm?prg=308&lv=3.

The support services provided by the Regional Information Technology (RIT) office include, but are not limited to: computer, server and other related hardware support; software installation and support; email administration; security updates; virus-protection; telecommunications; and computer networking.

Because the aforementioned services are maintained by the RIT, they will not be discussed in any detail in this document. Rather, the focus of this chapter is on the programs and policies that the NETN does control and/or has developed to steward the natural resource data the Network possesses, and that enable the Network to manage the Natural Resource Inventory and Monitoring program. Accordingly, this chapter will describe the capabilities that the NETN maintains, and the structure and systems that the NETN has developed to provide and maintain the services it offers to parks, the public and other audiences.

Capabilities

The Northeast Temperate Network offers a number of capabilities that are either missing, or that are underdeveloped in many of the small parks that comprise the network. These range from technical capabilities like GIS development and database design that may not exist at the park level, to technical assistance dealing with issues those parks or cooperators may not have time to complete. For example, the Network is positioned to offer technical guidance on database development, data collection and stewardship requirements, and will be ready to store and distribute data on behalf of Network parks by posting it to the NR-GIS DataStore.

Geographic Information System (GIS) Support

As part of the Inventory and Monitoring Program mission, the Network develops, maintains, analyzes, and distributes GIS datasets. GIS is an accepted and expanding technology, and much of the data the Network currently possesses and will be collecting has a distinct spatial component. The Network participates in the NPS GIS enterprise licensing program and possesses a full complement of software available through that program from Environmental Systems Research Institute (ESRI). A complete listing of the GIS software currently held by the Network can be found in the Software Inventory Appendix. In addition to using GIS to analyze data the Network possesses, the Network can provide GIS related assistance to parks, and cooperators working for the Network or Network parks.

Relational Database Support

The majority of monitoring data to be collected by the network will be stored and distributed using relational databases. Even data sets that are commonly thought to be purely spatial (i.e., GIS) may also be available in tabular format. Because of the Network's reliance on databases to store, manage and distribute data, promoting database standards throughout the network is increasingly important. Furthermore, as the number of data sets increase, and as NPS personnel come and go, information and data that are not part of a well designed database will become harder to efficiently locate, and may eventually be lost. While a database may never completely replace an organization's 'institutional memory,' well designed databases are a necessary part of the modern working environment. Accordingly, the Network stands ready to work with Parks to address their database needs, and in support of this objective has developed the Network database specifications SOP.

Managing data that exists in a variety of disparate formats is inefficient at best and ineffective at worst. NPSpecies, a system developed by NPS to inventory and report species found at parks, and NatureBIB, the NPS bibliographic program are good examples of two relational database systems developed and supported by NPS to organize park-based data and information. At the network level, NETN has developed a number of databases to support ongoing projects, and has worked with parks to import data from other formats into databases. For example, the network developed a forest monitoring database that is currently being used as the network pilots its forest monitoring protocol, and following receipt of four spreadsheets of loosely organized field observations, NETN worked with staff at the Appalachian Trail Park Office to develop a database that both stores data and provides basic search and reporting capabilities. In both of these instances, the Network worked with end users to develop a system that ensures data security while making the data easier to use.

The Network relies upon Microsoft Access as the primary tool to manage and work with tabular data. In addition to working with existing data sets, the Network uses the NPS Natural Resources Database Template (NRDT) format to develop new Microsoft Access databases. Maintaining a strong commitment to support Network parks and cooperators with relational database issues is among the highest NETN priorities. NETN extends an open invitation to help all Network parks and cooperators with any database problems they may be experiencing; help them to define database requirements; will help to consolidate data from multiple sources; and will work with them to develop new databases as required.

A complete listing of relational database related software, utilities, and applications developed and maintained by the Network can be found in the Software Inventory Appendix

Document Preparation

The Northeast Temperate Network requires that all future project reports be submitted in a format that is consistent with the NPS Northeast Region technical series printing standards. To support this requirement, the network has hired a Science Communication Specialist whose duties include assuring that all park and network sponsored project reports conform to Northeast Region technical series requirements. A notable example of a document the network prepared using the Northeast Region technical series guidelines is the Appalachian Trail Vital Signs report issued in November 2005.

In addition to future submittals, the Network has worked with parks and cooperators to reformat a number of significant existing studies. The purpose for doing this is to ensure that studies are prepared and presented in a manner commensurate with the standards expected by the public for work commissioned by the National Park Service. A good example of an existing report that the Network reformatted after it was originally submitted is an avian inventory completed by the Marsh-Billings-Rockefeller NHP entitled: A biological Inventory of Breeding Birds at the Marsh-Billings-Rockefeller National Historical Park and Adjacent Lands, Woodstock, Vermont (Faccio, 2003). The Network has used the Adobe suite of software products to produce more than a dozen of its own documents, several for another Network, and is ready to work with Network

parks and cooperators in the preparation of future documents. Refer to the Document Preparation SOP for instructions on initial document preparation, and to the Software Inventory Appendix for a complete listing of document processing software maintained by the Network.

Data Integration

Data integration is a term that means bringing together data from various sources and in differing formats and presenting that information in a form that is useful to the target audience. Making acquired data meaningful to our audience is among the most fundamental of all Network responsibilities. To do this, the Network Science Communication Specialist packages and disseminates Network information in reports and newsletters, and meets with park staff and the public to describe network activities. To ensure that the network delivers information that is targeted to the audience, the Science Communication Specialist works closely with Parks and cooperators to develop applications tailored to their requirements, combining disparate information into a single user interface. For example, the Network is developing a plan to enhance the NETN 'web' presence by making the existing Network web page more engaging and meaningful to our audience by integrating information from various sources. For more details on the Network's data integration strategy, see Chapter 11.

Data Acquisition

The Northeast Temperate Network seeks to acquire and maintain a record of natural resource data for all parks within the Network, as well as data that is not directly associated with parks but is regionally focused and related to park activities. The value of these acquisitions is many fold. Information contained in these datasets may offer insight into the resources that are in and around network parks; they may offer a comparison between current and historic park resource condition; and, they may help to improve our ability to understand work commissioned by the network or by a park.

Digital data shall be stored by the Network and made available to cooperators, park and/or Network staff, and others in compliance with established data distribution policies. Data that is properly documented with metadata and that is free of data distribution restrictions will be posted to the NRGIS-Data Store where it can be accessed by the broadest audience. To promote the distribution of Network data that is stored on the NRGIS-DataStore, the Network worked with a University of Massachusetts Cooperator to develop a search utility that streamlines the process of locating network data resources. This utility may be accessed at the following link: http://www1.nature.nps.gov/im/units/netn/data/data.cfm

Data that are not documented with metadata, or that have data distribution restrictions, will also be acquired by the Network. However, the Network will not distribute inadequately or improperly documented data, or data that have distribution restrictions. Historic data, in formats other than digital, will also be obtained when available, and scanned into digital format and made available to cooperators, park and/or Network staff, and others in compliance with established data distribution policies.

Data that is generated through Network activities will be permanently stored and archived along with all other project related information. Data that is not generated through Network activities will generally not be permanently archived by the Network. For additional details on the Network's plan to acquire key datasets, please see Chapter 6, Data Acquisition.

Organizational Infrastructure

Establishing and maintaining a consistent data structure is a critical first step in developing a robust Data Management program. The following elements describe the fundamental structure upon which Network Staff depend.

Directory Structure

The NETN directory structure is modeled after the recommended Inventory and Monitoring Program directory layout (http://science.nature.nps.gov/im/datamgmt/docs/datastr5.doc). The Northeast Temperate Network rendition of the recommended directory layout groups data holdings into six 'master' groupings: Parks; Program; Projects; Shared; Users; and Version Control. Below the master grouping level, directory configurations closely match the recommended layout. Refer to the File and Folder Naming Conventions Appendix for additional details on the NETN directory structure.

Parks

Data contained in the park master group is related primarily to a single park, and consists primarily of GIS data. The park master group may also contain correspondence specific to the park in question. Working data is typically contained in park directories. In addition to park related day, this master group also includes a NETN directory that contains GIS data applicable to the broader Network region.

Programs

The Program master group is subdivided into NETN, NERO and WASO sub-sections. Data found in the program master group relates to the operation of the program in question. Data contained in the WASO sub-branch is related to the National Inventory and Monitoring Program, while data contained in the NETN sub-branch may include correspondence, reports, database projects, and anything relating to the operation and development of the Northeast Temperate Network. Datasets found in the Programs directory structure are generally working or test versions of real datasets, but in some instances original datasets are also maintained under the Program directory structure. For example, the WASO program directory contains original versions of databases for NPSpecies and NatureBIB, whereas the Databases subdirectory in the NETN program directory only contains working and development versions of databases developed and maintained by the network.

Projects

The project master group contains data and documentation specific to Network projects, both ongoing and complete. The intent of this master group is to provide a comprehensive filing structure intended to document all aspects of a project from inception to completion. Project directories contain original data as well as versions of datasets that are preliminary or that are undergoing review.

Shared

Data contained in this master grouping is broadly shared by multiple users, but is not directly associated with a park, program, or project. For example, GIS projects that may relate to one or more parks are stored under the shared grouping. Software used by the Network is also documented in this master group. Shared directories are not used to store datasets either temporarily or permanently.

Users

Data contained in the user master group is generally temporary or transitional, and may be of interest only to a specific system user. Development versions of datasets, databases or documents may be found in a User subdirectory prior to permanent storage.

Version Control

The version control master group contains two sub-groupings – archives and reference – both of which are accessed exclusively by the version control system the Network has implemented. Files contained in the version control master group are software application dependent, and cannot be used or accessed independent of the version control application. No user accessible data are stored in the Version Control directory.

File Naming

The Northeast Temperate Network has developed a file naming convention that concatenates a string of elements that relate to a file's topic or purpose. In addition to developing the file naming convention, the Network has also developed a utility that automatically generates a file name by prompting the user to select elements from a series of pick lists. The resulting file name complies with the Network file naming convention. The File and Folder Naming Conventions Appendix contains a detailed description of the NETN naming convention and guidelines on the use of the Network file naming utility.

System Back-up

Ensuring the stability and security of all data acquired, created, and maintained by the Northeast Temperate Network is a necessity. To achieve this requirement, the Network has implemented a multi-stage, automated digital file back-up routine. A complete description of the data back-up system can be reviewed in the Network Backup Appendix.

Budget Tracking

The Northeast Temperate Network, like every other Inventory and Monitoring Network, balances its funding against the NPS Administrative Financial System (AFS) and the Inventory and Monitoring Program Annual Administrative Report and Workplan (AARWP) database. Funding tracked in AFS is strictly linked to funding source code, whereas in the AARWP funding from multiple sources is combined and distributed among projects and activities. NETN has developed a 'front-end' database that bridges the gap between these two systems, giving staff the ability to balance expenditures and track the Network budget throughout the fiscal year. The system incorporates a series of features that are designed to show staff where funding balance discrepancies exist, and gives staff the ability to reconcile Network expenditures against AFS in a way that is similar to how a bank customer reconciles their checking account. The user manual for the NETN budget tracking database is available in the Network Budget Tracker Appendix.

Hardware and Equipment

As discussed in the introduction to this chapter, a lengthy description of the hardware used by the Network, and the systems through which the Network connects to other park, regional, and national computers is beyond the scope of this document. Many of these systems are administered by other NPS elements, and many are transitional in nature. The following elements are provided to offer a more complete understanding of the Northeast Temperate Network Inventory and Monitoring program.

Computers and related hardware

The Network possesses a number of computers, both laptop and tower systems, pocket PC's, external storage devices, and scanning and printing equipment. Because the equipment list is subject to change based upon the needs of this developing program, equipment is tracked in the Network equipment inventory database. Refer to the equipment Appendix or to the Network equipment database (not currently linkable) for an up-to-date listing of equipment.

Global Position System (GPS) Receivers

Establishing a spatial reference for sampling locations and significant natural resources is an important part of all work being done by the Network. To support this requirement, the Network has acquired two GPS systems that are available for Network projects, may be borrowed by Network parks, and may be used by Network cooperators. Each system consists of the following:

- GARMIN V GPS Receiver
- MBX-3S Dual Channel Beacon Receiver with High Gain MGL-3 antenna
- GARMIN MapSource and GPS Photo-Link Software

Refer to the equipment Appendix or to the Network equipment database for an updated list of GPS equipment available from the Network, links to specifications lists, and to the Regional GPS Operations Guide for directions on the proper use of this equipment.

Digital Cameras

The Network has, and can make available for Network projects, entry level digital photographic equipment. Currently, the network has a 4.1 megapixel Minolta DiMage S414, a 4.1 megapixel Casio QV-4000 and a 7.1 megapixel Canon Powershot A620 with a waterproof case. Refer to the equipment Appendix or to the Network equipment database for an updated list of digital photographic equipment available from the Network and links to detailed specifications lists.

Water Quality Sampling Equipment

To support the Network's water quality sampling program, the Network has acquired YSI 600XL Sonde's equipped to measure depth, temperature, conductivity, pH, and dissolved oxygen. Two units are configured for shallow (0-30 ft) operations while the third unit is configured for medium depth operations (0-200 ft). Refer to the equipment Appendix or to the Network equipment database for an updated list of water quality sampling equipment available from the Network and links to specifications lists.

3. Roles and Responsibilities

Overview

George Miller (1960) described information processing in the human mind in the following way:

"…Like the computer, the human mind takes in information, performs operations on it to change its form and content, stores and locates it and generates responses to i…"

What we are attempting to accomplish in the Inventory and Monitoring Program is akin to what each of us does every day of our lives – acquire and store data, generate information, and make that data and information available at a later time. Like a living organism that relies on its senses to acquire data, an Inventory and Monitoring Network relies on many different individuals to acquire project data, organize it, store it for later use and then retrieve it when necessary. In the end, the team of field personnel, ecologists, resource managers, data managers, project leaders, and others must combine their efforts and function as a single 'organism' to collect, organize store and synthesize all acquired data if a project (or the program in general) is to be successful. If one segment of the Inventory and Monitoring 'organism' fails to perform their part of the project, the resulting data may exclude the pieces necessary to yield meaningful information.

Fundamentally, data management is as much about people and organizations as it is about information technology and database theory and application. Every person in an organization manages data and information at some level, and their understanding of how they fit into the organization and how their actions affect the quality and utility of the resulting data and information is a critical element of the data management process. For the Network program to be most effective, every individual involved with a Network project must share stewardship responsibilities in the production, analysis, management, and/or end use of the data (Figure 3.1 and Table 3.1). By communicating and working with all the responsible individuals, creating and

Table 3.1 Categories of data stewardship involving all Network personnel (adapted from Greater Yellowstone Network, Daley 2005).

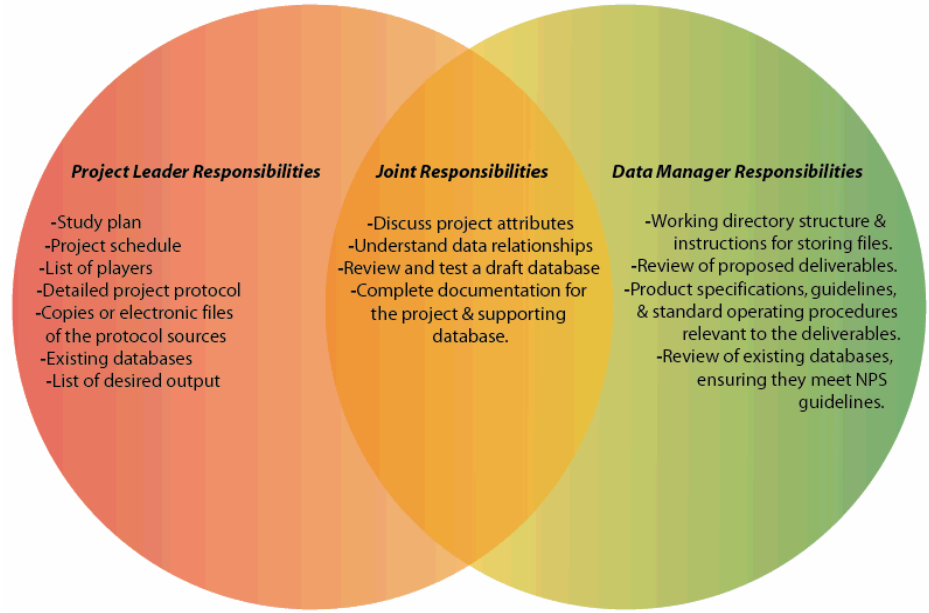
Project Stage	General Description	Responsible Individual
Data Acquisition and Organization	Acquire data or information from original and derived sources; record field measurements and observations; digitize source maps; convert and process data; and, prepare summary tables, maps, charts, and reports.	Field Staff Data Technicians Data Manager
Data Analysis	Use data to predict, qualify, and quantify ecosystem elements.	Network Coordinator Cooperating Ecologists
Data Storage	Prepare and execute policies, procedures, and activities to keep data and information resources organized, available, useful, compliant, and safe.	Network Data Manager
Data Retrieval and Usage	Post data to NPS NRGIS Data Store	Network Data Manager

maintaining data infrastructure and standards, and understanding program and project requirements, the data manager supports and guides people and activities for the purpose of meeting the Network's data management objectives.

Roles and Responsibilities

The role an individual plays in a project is the primary project-related function or functions that the person serves. A person can have single or multiple roles on a given project. Responsibilities

Figure 3.1 Shared Responsibilities. (From Southwest Alaska Network)



are those project related actions that a person in a given role is expected to perform. Table 3.2 identifies eight specific project-related roles, ranging from field crew to Network Coordinator and a series of responsibilities expected of someone performing each of the roles. Some responsibilities transcend those that are linked to a particular role and deserve specific attention. Such responsibilities, including coordination, communication, and documentation, can be considered Shared Responsibilities.

Coordination

The Natural Resource Challenge encourages collaboration between the National Park Service, other public agencies, universities, and non-governmental organizations to effectively acquire, analyze, and disseminate the scientific knowledge gained in parks. To make this vision a reality, the Network Data Manager as well as other project staff will work locally with Network personnel, park staff, and cooperators to promote and develop workable standards and procedures that result in integration and availability of datasets that are consistent with service-wide and regional databases and data management policy and guidance.

Table 3.2 Summary of Project Level Roles and Responsibilities (from GRYN)

Role	Primary responsibilities related to data management
ALL	• Understand the objectives of the project, the resulting data, and its scientific and management relevance; • Data stewardship.
Field Crew	• Adhere to sampling protocol; • Record and verify measurements and observations; • Document problems with protocol, recommend changes, and document anomalies.

Table 3.2 Summary of Project Level Roles and Responsibilities (from GRYN)	
Role	Primary responsibilities related to data management
Crew Leader	• Supervise field crew; • Ensure sampling protocol is followed; • Review data entries for accuracy; • Document all procedural deviations; • Ensures that computer is routinely backed-up.
Data/GIS Specialist or Technician	• Enter and transform data; • Recommend ways to improve data quality and stewardship procedures; • Develop procedures for spatial data acquisition; • Display, analyze, and create maps from spatial data; • Properly document data in compliance with applicable metadata standards.
Project Leader	• Oversee and direct Network projects; • Maintain communication between project staff, Network staff, and resource specialists; • Document the who, what, where, when, why and how of a project; • Develop, document and implement standard procedures for data collection and handling; • Oversee field operations, including staff training, equipment calibration, species identification, and data collection; • Coordinate changes to the field data forms and the user interface for the project database; • Identify sensitive information that requires special consideration prior to distribution; • Define how project data will be transformed from raw data into meaningful information.
Resource Mgr.	• Make decisions about data validity, utility, sensitivity, and availability; • Describe, publish, release, and discuss data and associated information products.
GIS Manager	• Establish GIS data and analysis needs for the project; • Develop and enforce GPS standard procedures; • Coordinate and integrate local GIS and resource information management with Network, regional and National standards and guidelines.
Network Data Manager	• Provide planning, training, and operational support for data and information management activities, including people, information needs, data, software, and hardware; • Serve as Point of Contact for National Park Service database applications; • Coordinate internal and external data management activities; • Oversee metadata creation, project documentation, and project data management; • Create and maintain project databases in accordance with best practices and current program standards; • Develop quality assurance and quality control procedures; • Develop an archival process to preserve all project documentation, original field data, databases, reports and summaries, and other products from the project; • Establish and implement procedures to protect sensitive data; • Ensure that project documentation is complete, complies with metadata requirements, and enhances the interpretability and longevity of the project data; • Inform project staff of changes and advances in data management practices.
Network Coordinator	• Ensure project objectives fulfill programmatic data and information management requirements.

A good example of where the Network has benefited from coordinating its interests and needs with another agency is the development of ecological monitoring protocols. Specifically, the Northeast Temperate Network has worked closely with the U.S. Geologic Survey on the development of aquatic resource protocols. Most or all Network monitoring protocols have benefited either directly or indirectly from coordination within the Inventory and Monitoring Program and from contributions made by external partners. Examples of organizations and

agencies outside of the National Park Service, and projects with which they are associated, are listed in the Adjacent Monitoring Appendix of the NETN Vital Signs Monitoring Plan.

Communication

Communication should exist at all levels – project, Network, program level and beyond. Field staff members are encouraged to identify and relay issues they encounter during work on a project just as staff in one Network should be encouraged to discuss issues with staff in another Network. A good example of communication can be seen in the data management community where Data Managers at Networks throughout the Inventory and Monitoring Program interact at annual meetings, through conference calls, in workgroups, on a listserv, through a web site, and through direct communication. This approach to collaboration using available resources and communication can be productively applied as the NETN encounters data management issues and/or requires assistance. Open communication is the key to success!

Awareness and Accountability

Keeping track of data from the time of acquisition until it is properly archived is the shared responsibility of everyone involved with a project. Every level, from producer, to analyst, to manager, to end user bears some stewardship responsibilities. The importance of data stewardship cannot be understated, and has a direct bearing on the quality and utility of products developed by the Network.

Data Stewardship

Data stewardship is defined to mean the proper 'handling' of data, and applies through every stage from data acquisition through storage and dissemination. In short, data stewardship is necessary to ensure that data resulting from a project is organized, complete, reliable, and available. Successful data stewardship requires that people involved in Network activities learn and understand the expectations for continuous data management. This is equally important for Network staff, park employees, and contractors or cooperators. All project participants receive training, briefings, materials, and additional communication about data stewardship from supervisors, project managers, and data managers. The purpose is to promote the appropriate level of understanding about how their efforts relate to park and Network management objectives, National Park Service and Department of Interior policies, and other federal government requirements.

Documentation

If a single shared responsibility stands out in importance and value, it is to document data sets, the data source(s), and the methodology by which the data were collected or acquired. This establishes the basis for the appropriate use of the data in resulting analysis and products, both short term and long term. Network monitoring protocols contain key elements of data documentation.

Network Implementation

Roles and responsibilities exist in two broad categories: program oriented and project oriented. Program oriented roles and responsibilities belong primarily to the Network Coordinator and the Data Manager, establishing the umbrella that covers all activities performed by the Network. For example, in the Northeast Temperate Network the Data Manager is responsible for ensuring that all project files are regularly backed-up to minimize the risk of data loss in the event of system failure. This is a program level role and responsibility. Project roles and responsibilities are project specific, and typically involve more people. For example, field crew leaders working on the Northeast Temperate Network forest monitoring project are responsible for ensuring that their

equipment is accounted for, that all data has been recorded, and that the field computer is backed-up.

Because the focus of this document is on the way the Network stewards the data and information it acquires, and how projects are planned and how data is acquired and delivered, this document is project oriented.

The Northeast Temperate Network is comprised mostly of small parks with relatively few natural resource staff. Most parks in the Network have at least one natural resource manager, one park has a part-time data manager (ACAD), and two parks have GIS specialists (ACAD & APPA). None of the Network parks have a dedicated database developer or an information technology specialist. As Network projects are conducted, the work will be performed by a combination of staff from cooperating agencies and organizations, and staff assigned to Network parks or the Network itself. Overlap in responsibilities is likely to be the norm, not the exception. Because of this overlap, it is more important to identify likely roles and responsibilities according to what function a person is performing on a given project than to strictly assign roles and responsibilities according to the job title a person holds. For example, when the Network Coordinator assists in the collection of field data for a mammal survey, the coordinator's actions would be governed by the obligations and responsibilities expected of other field staff.

4. Project Management

Background

"*Routinely, projects are late, over-budget, or fail in some way to satisfy the client*." This sobering assessment of the project management process by Macomber (2004) clearly suggests that time spent planning for projects is worthwhile, and that procedural improvements aimed at addressing project deficiencies are justified. Multiple factors contribute to project failures, including poor planning; lack of standards; undefined timeline; unstated objectives; missing contact information or unstated responsible parties; and a failure to track project progress and monitor critical milestones. With these concerns in mind, a task that confronts Northeast Temperate Network staff is organizing and becoming familiar with all the ongoing and planned activities occurring at the affiliated parks.

Communication

Though communication is not ordinarily identified as a discrete stage of the project management process, it would be impossible to complete a multiple component project without some form of communication between those responsible for the different pieces. For this reason, communication deserves full consideration at the outset of a project's design. Roles and responsibilities must be discussed and established (reference Chapter 3), the objectives must be set, and the various deliverables due at the conclusion of the project must be identified. The Northeast Temperate Network has found that communication is critically important in collaborative document writing. For example, during the development of the Network Vital Signs Monitoring Plan, the Network worked with a number of cooperators to collaboratively generate a single planning document. Several of the pieces had different origins, but they needed to fit together. To ensure that the process of fitting the pieces together went smoothly, the Network developed a Document Coordination SOP. The appendix broadly covers a number of topics intended to promote the use of standards, and minimize problems associated with merging disparate documents at the end of the collaborative process.

Planning and approval

During this initial phase, Northeast Temperate Network projects are conceptually defined; goals, objectives, scope, and purpose are identified; and potential funding sources, permits and other administrative requirements are considered, obtained, and addressed. This stage establishes a foundation for all remaining stages, making it the most important stage in the project development process. At the conclusion of this process, the project should have a well defined need, objective, and purpose that complies with the recommendations outlined in Oakley et al (2003). Though these recommendations may seem self-evident, the consequences of not following them can lead to increased project costs, reduced efficiency, and

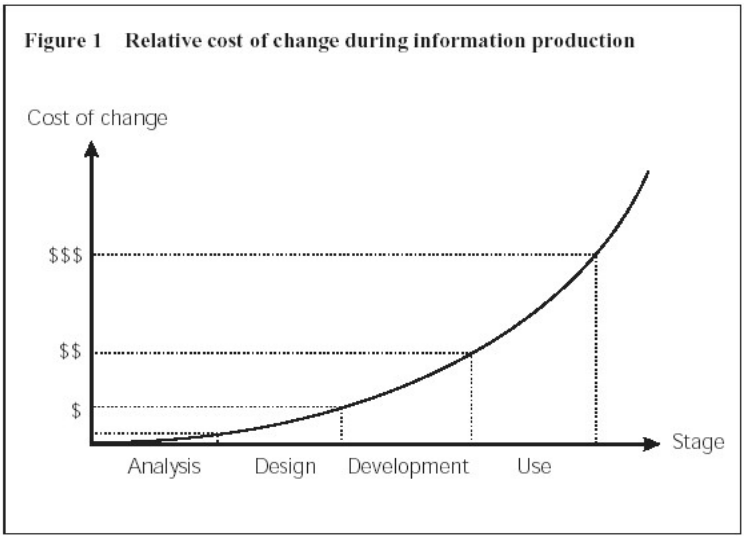


Figure 4.1. From Volume 2 of the WCMC Handbook (2004)

data with limited utility (World Conservation Monitoring Centre, 1998). Figure 4.1 illustrates the relationship between cost and the timing of project changes.

While this phase does not consist of activities such as archiving, data verification and validation, or any other traditional data management elements, data management is an important aspect of the planning and approval stage because this is when data management requirements are first identified. Contracts, agreements and permits will be developed; likely data formats are identified, and specifications are established; timelines are proposed and the final destination for permanent storage of project deliverables is identified. Data management takes time and costs money. For this reason, data management requirements need to be explicitly stated in agreements and contracts, and factored into time lines to ensure they are not overlooked.

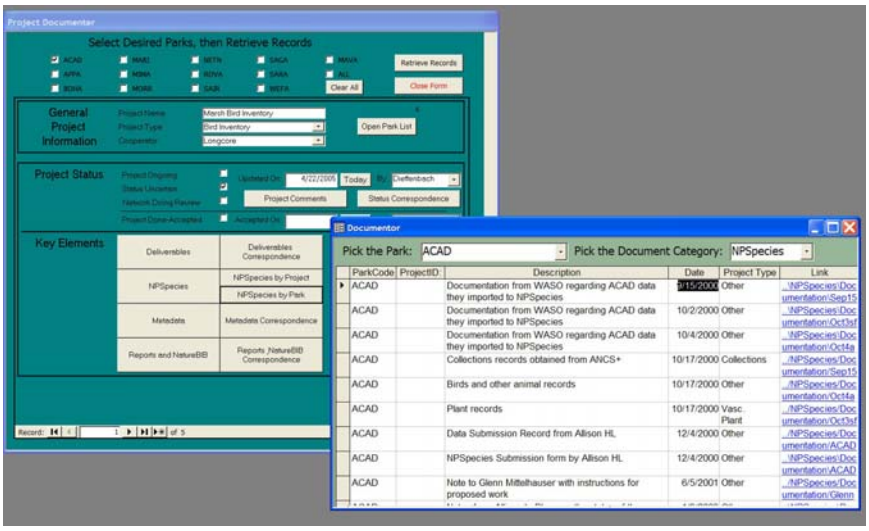


Figure 4.2. Custom Project Tracking Database

Project Tracking

Project tracking can be as simple as assembling a list of Network sponsored projects in a spreadsheet to implementing a dedicated application like Microsoft Project. The Northeast Temperate Network relies on both a custom designed project management solution (figure 4.2) as well as Microsoft Project (figure 4.3). The custom designed application enables the network to organize and store information that is specific to various Inventory and Monitoring program activities that would be difficult or impossible to incorporate into a Microsoft Project file. For

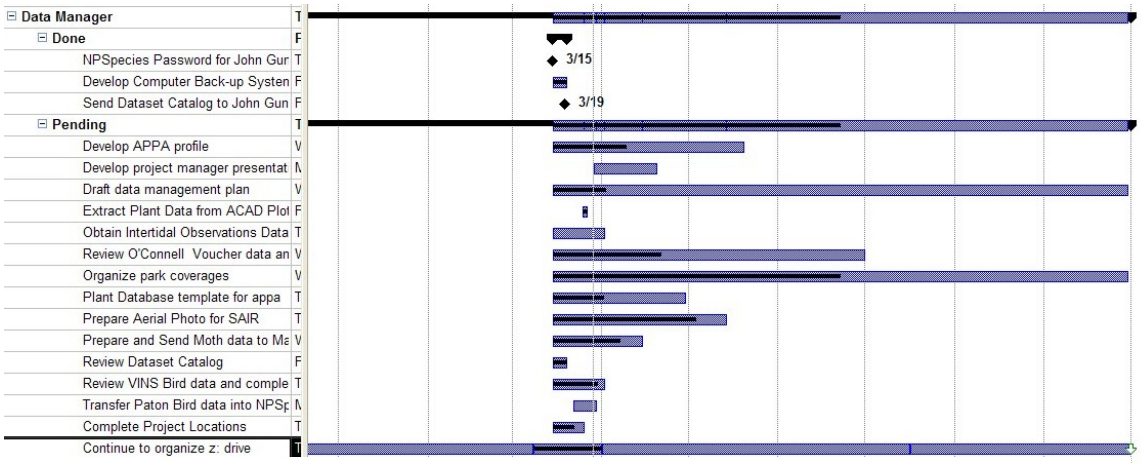


Figure 4.3. Gantt Chart created in Microsoft Project

example, the application incorporates a system that records NPSpecies certification elements on

a park by park basis and organizes correspondence related to NPSpecies – or any other project – on a park-by-park basis, thus permitting ready access to this information. Prior to implementing this system, Network staff struggled to manage the volume of correspondence that pertained to different activities. With the custom application, the entire historical record for a project is readily available. The Network uses Microsoft Project as a planning tool to estimate proposed activities. Though Project is capable of many things, the Network utilizes it in a limited way primarily during planning to identify due dates and time lines, and display these key data in a variety of ways. Figure 4.3 shows an example of a Gantt chart in Microsoft Project similar to what the Northeast Temperate Network data manager used to visualize proposed activities during Fiscal Year 2004.

Project Budget

Network budgets are broken into three segments: agency; program; and, project demands. The National Park Service funding and expenditure process (Agency) is tracked by AFS; the Inventory and Monitoring Program (Program) tracks Network financial activities through the Annual Administrative Report and Workplan (AARWP) database (Figure 4.4); and, Networks must devise their own way to track expenditures on a "daily" or project (Project demand) specific basis.

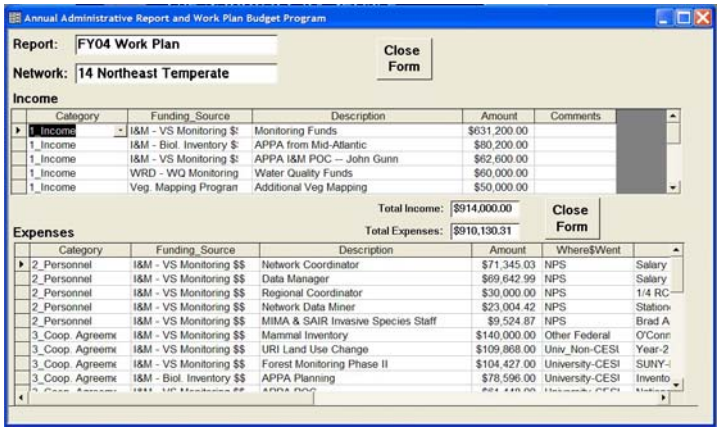


Figure 4.4. Standard AARWP Data Entry Screen

The Northeast Temperate Network (NETN) has designed an "add-on" for the AARWP database. While the AARWP database gives Networks the ability to summarize their income and

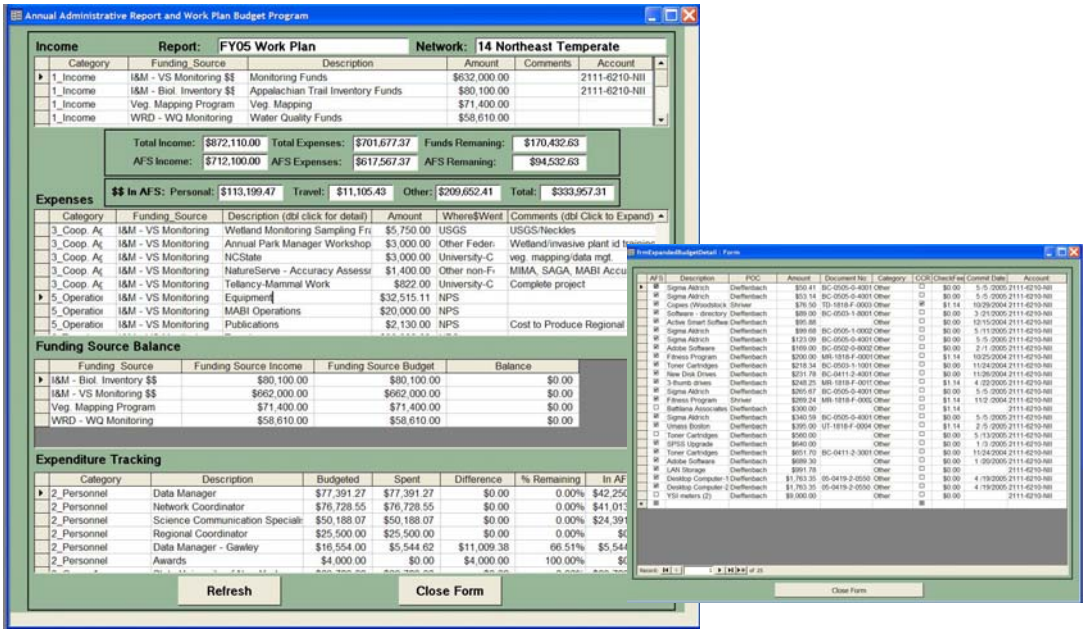


Figure 4.5. NETN Budget Tracker with Expanded Detail Inset

expenses to meet "Program" requirements, it does not give Networks the ability to easily manage "Agency" level activities or to track individual project specific or operational expenditures. The NETN add-on operates as a 'front-end' to the AARWP, taking advantage of the table structures and requirements built into the AARWP database. It then adds a series of modules that permit Network staff to closely monitor expenses throughout the year. The system functions similarly to a check register that is periodically 'reconciled' against the official National Park Service financial management system, AFS. By designing a system that is periodically reconciled against AFS, and that inherently works with the AARWP, the Network satisfies its internal budgetary requirements while fulfilling its AFS and AARWP budgetary obligations. Figures 4.4 and 4.5 show the standard AARWP data entry screen and the NETN Budget Tracker expanded entry screen, respectively. Specific Network budgetary details are discussed in the Network Budget Tracker Appendix.

Project Design

The Institute of Museum and Library Services (IMLS) states in their 2001 Framework of Guidance for Building Good Digital Collections that ". . .*projects are initiatives of finite duration, designed to accomplish a specific goal*…" and that ". . .*it is important to distinguish between the project, which is transient, and the collection, which in most cases should persist*…" While the objective of IMLS is to collect and catalog digital products and not to create a record of the work that generated the products, the Northeast Temperate Network embraces this philosophy by stressing the importance of high quality deliverables and proper documentation. For the purposes of making the information and data available, the products associated with a project must be documented, and for the purposes of documenting the Network's activities we must also document and archive project specific information, including contracts, agreements, proposals, reports, correspondence, and other products. Details regarding the Network's curatorial and archival strategy may be found in the Natural History Curation and Archiving Appendix (pending).

IMLS identifies three distinct project related principals:

Principle 1: A good project has a substantial design component.

- Project planning, from processing workflow to the final product;
- A realistic assessment of the functional requirements of users.

Principle 2: A good project has an evaluation plan.

- Standards used to evaluate the success of a project.

Principle 3: A good project produces a project report.

- A project assessment providing a detailed description and analysis of accomplished work (After Action Review);
- A report describing the project, data, and deliverables.

After a project is defined, processes, procedures, and methods will be described in a draft SOP. All data elements to be collected in the field are defined in a data dictionary. After the project methods and data dictionary have been developed, reviewed, and tested, the Network will develop a preliminary database design to meet project requirements. See the Northeast Temperate Network database specifications SOP. The resulting design will minimally include table structures, data entry forms, and data output requirements. The Network data manager will oversee the database development, with the actual development work being done 'in-house' or under agreement with a Network cooperator.

Project Testing

During this phase, the project methodology is tested, and SOP's are reviewed for usability, accuracy, and completeness. Specific procedures related to data acquisition, processing, analysis, and quality control are evaluated.

Network data management staff will evaluate existing databases developed by other Inventory and Monitoring Networks and/or cooperators for functionality overlap and/or data design. Every opportunity to utilize existing databases will be sought.

By the close of this phase all anticipated aspects of the project should be thoroughly tested, and the testing and evaluation process will be thoroughly documented. Trial data collection, processing, analysis, and QA/QC tests should be complete and deemed successful. Any parts of the project that seem uncertain should be addressed, and any outstanding questions should be answered. Once Network data management personnel are satisfied with the methodology and associated data collection system, the project metadata record can be initiated.

Project Implementation

During this stage, the project transitions from testing to actual use. Data are acquired, processed, error-checked and documented. Reports, maps, GIS layers, and other products are delivered. Data management staff will provide training and support for database applications, GIS, GPS and other data processing applications; will facilitate data summarization and validation; and will provide technical assistance with product delivery.

There are three main parts of the implementation phase: preparation; acquisition and processing; and product delivery and review.

Preparation

During the preparation stage, all logistical aspects of the project occur that are not set during earlier stages, including:

- Staffing;
- Staging for access to remote sites;
- Training; and,
- Equipment procurement and installation.

Data acquisition and processing

This stage represents the heart of the project's implementation. Everything considered and tested during earlier phases is put into action. Data will be acquired and quality assurance and quality control measures are put into place. Network staff should communicate periodically with those doing the actual data acquisition to ensure that the SOP's are being followed, and that any problems are quickly addressed. Any deviation from the protocols specified in the SOP's should be carefully documented, and quality assurance issues should be carefully described.

Product delivery and review

Project staffs provide the deliverables specified in the project contract, agreement or permit. Administrative records should be delivered to appropriate park and Network staff as specified for storage. All raw and derived data products, metadata, reports and other documentation should be delivered to the data manager. The Network data manager will review all delivered products to ensure that they meet product specifications, and have been subjected to all applicable Quality Assurance/Quality Control procedures specified in the Quality Assurance/Quality Control Appendix. Products that do not meet program requirements will be returned for revision.

Product Integration

The data manager is primarily responsible for coordinating product integration. The objective is to merge data products and other deliverables into national and Network databases; to ensure that metadata records are finalized and posted to clearinghouses; and to make sure that products are distributed or otherwise made available to their intended audience. Items that belong in collections or document archives are accessioned and catalogued; reports and other documents

are entered into NatureBIB; images and documents are uploaded to the appropriate repositories; metadata records are posted; and NPSpecies is updated with information derived from the project. Additional details relating to product integration may be found in the data dissemination chapter.

Evaluation and closure

At the conclusion of a project, or at the end of a project's field season, Network staff should conduct a project review. The review should be seen as an opportunity for program administrators, project managers, data managers, and other involved individuals to evaluate the effectiveness of the existing program or project and to implement changes. The review should provide a forum for all involved individuals to provide open and honest 'constructive' remarks about the project. The format for the review may be informal or formal, depending on the number of individuals and size of the project, but regardless of the formality the results should be documented and archived.

5. Data Acquisition and Processing

Background

The National Park Service Management Policies (2001) specifically direct the Service to inventory and monitor natural systems:

"Natural systems in the national park system, and the human influences upon them, will be monitored to detect change. The Service will use the results of monitoring and research to understand the detected change and to develop appropriate management actions".

Further, the 2001 Policy specific to Natural Resource Management directs the NPS and the Inventory and Monitoring Program to:

- *"Identify, acquire, and interpret needed inventory, monitoring, and research, including applicable traditional knowledge, to obtain information and data that will help park managers accomplish park management objectives provided for in law and planning documents";*
- *"Define, assemble, and synthesize comprehensive baseline inventory data describing the natural resources under its stewardship, and identify the processes that influence those resources";*
- *"Use qualitative and quantitative techniques to monitor key aspects of resources and processes at regular intervals";*
- *"Analyze the resulting information to detect or predict changes, including interrelationships with visitor carrying capacities, that may require management intervention, and to provide reference points for comparison with other environments and time frames";* and,
- *"Use the resulting information to maintain-and, where necessary, restore-the integrity of natural systems".*

To fulfill these objectives, the Network will collect data and information from multiple sources ranging from inventory and monitoring data (programmatic data) to other NPS and non-NPS sources (non-programmatic data). After acquiring these data, the Network will ensure that the data meet the standards set by the Inventory and Monitoring Program. This chapter describes the different kinds of data sources and the procedures used to obtain and manage them.

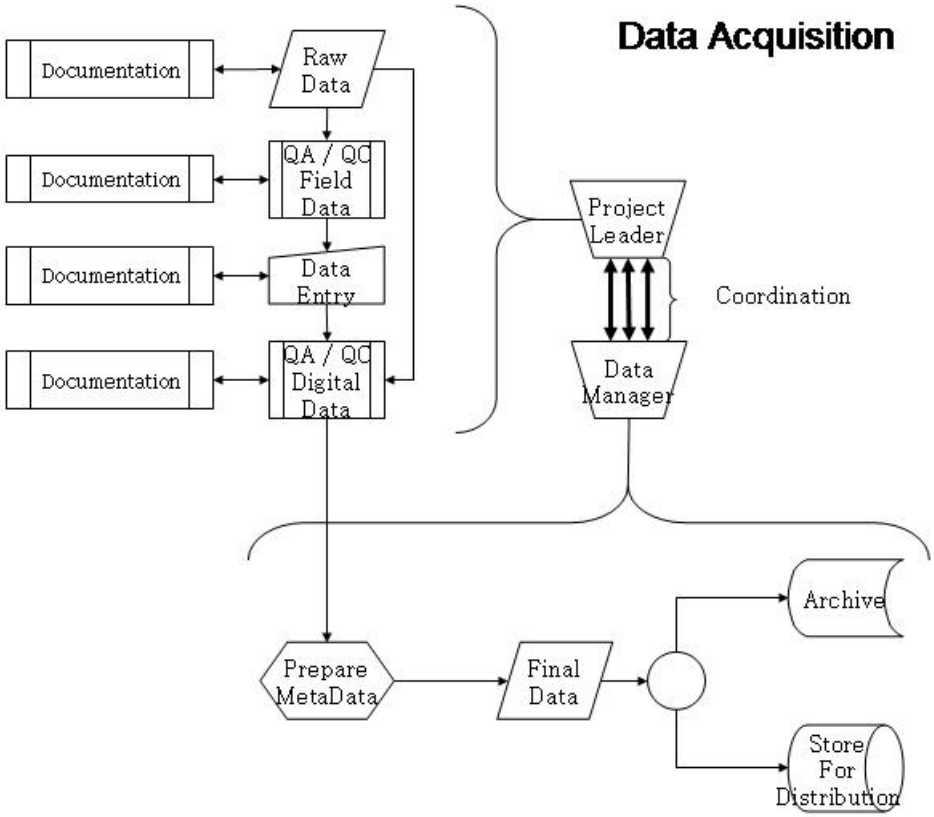
Acquiring and Processing Field Data

The following section provides an overview of how the Northeast Temperate Network intends to acquire and process data up through the production of archived datasets and the distribution of these data. The steps are written in a general format. While data acquisition and processing may vary depending on specific project needs and objectives, most data collected by the Network will follow this path. Figure 5.1 is derived from the linear model presented by Palmer (2003). In the NETN model, the same general steps appear as presented by Palmer, but documentation and coordination throughout the various stages is emphasized. In addition, instead of following a strictly stepwise process, the NETN model acknowledges that it may be necessary to return to earlier acquisition stages to verify or validate portions of the data.

Programmatic Data

Programmatic data originate from individual parks within the Network, or through work directed by the Network at one or more parks. If the Network has either paid for the work, or was actively involved with a project that was paid for by a park within the Network, it is considered to be a Network-based project. Examples of work already completed or currently underway in the Network include vegetation maps and species inventories, which are two of the twelve basic Inventory and Monitoring Program biological inventory datasets.

Figure 5.1. Data Flow Model-data acquisition to final products.



Non-Programmatic Data

Data from sources 'external' to the Network are defined by Northeast Temperate Network as non-programmatic. Sources include parks, universities, other agencies, and other NPS programs. The defining criteria are that these data offer important natural resource information, and that they provide park managers with information they need to make better-informed decisions. Non-programmatic data comes in two general forms:

- *Current or ongoing datasets*-These datasets may be park-based or from external sources. The distinguishing feature about these datasets is that they are independent of Network activities and are likely to continue regardless of whether they are adopted by the Network or not. Weather data is an example of an external dataset that Networks may use to help understand monitoring data.
- *Legacy datasets*-These datasets are typically found and acquired during the data mining process. Examples include vertebrate and vascular plant species data, other important natural resource inventory data, specimen or voucher data, bibliographic data, and existing monitoring datasets.

Identifying and obtaining legacy data are part of the ongoing Network data mining and compilation segments of the Network data management program. These efforts are directed at identifying:

- species records to be entered into NPSpecies;
- natural resource bibliographic data for entry into NatureBIB;
- long-term monitoring data; and

- spatial data

Several of the major legacy data efforts and project activities conducted to date for the Northeast Temperate Network are described below.

Species Data

Initial data mining efforts within the Network were directed by the Northeast regional inventory and monitoring office at the University of Rhode Island. After preliminary efforts to identify datasets and populate NPSpecies were completed, the Network Data Manager assumed the responsibility for managing the Network NPSpecies program. Ongoing NPSpecies management activities including QA/QC, Certification, uploading data and handling data requests, are done in compliance with the Network NPSpecies implementation plan.

Bibliographic Data

NETN has supported a NatureBIB Database Management position with the three other northeast region Networks. The regional NatureBIB database manager was hired to identify and enter bibliographic park references (data mine) for natural resource information associated with all Northeast Region Inventory and Monitoring Program parks. For more information see the I&M Program NatureBIB Program site, the Northeast Region NatureBib Database Management Plan Appendix, and the Northeast Region NatureBIB Users Guide Appendix.

Specimen Collection data

The Northeast Region Inventory and Monitoring Program initiated a study to locate voucher specimens from select national and international institutions with computer-accessible collections. The objective was to produce a complete list of institutions that possess vertebrate and vascular plant specimens collected on Northeast Region NPS lands. Data from this project were compiled and added to NPSpecies. More information is available from the technical report or the poster presentation.

Long-term monitoring

To ensure that Network activities are applicable to and consistent with existing monitoring programs, the Network initiated a multi-pronged approach to gather existing monitoring program information, summarize monitoring protocols, and identify ongoing monitoring locations relative to the NPS units in the Network. The work described under this project was accomplished through literature review, web-based queries, contact with NPS resource managers, and through interviews with regional professionals involved in monitoring natural resources. The information gathered through this effort is stored in a MS-Access database managed by the NETN data manager, and is incorporated into the Network Vital Signs Monitoring Plan as an appendix to that plan.

Water Quality Data Integration

The water quality component of the Natural Resource Challenge requires that Networks archive all physical, chemical, and biological data collected with water quality funds in the National Park Service's NPSTORET database. The NPS Water Resources Division manages the database and has developed a series of Microsoft Access-based templates to facilitate this process. NETN will enter and send their data from NPSTORET to the Water Resource Division in accordance with guidance provided by the NPS Water Resource Program on an annual basis, where it will be reviewed and uploaded into the NPS copy of STORET and the Environmental Protection Agency's (EPA) STORET National Data Warehouse.

6. Quality Assurance and Quality Control

Background

"*Information is useless unless it is accurate*…" (White, 2004). While this statement was originally directed toward a corporate audience, the fundamental truth it contains is as applicable to data collected by the Northeast Temperate Network as it is for a large corporation. Data collected through monitoring activities must be uniform, consistent, and accurate if they are to serve the needs of the Inventory and Monitoring program and resource managers. If collected data do not meet these requirements, analysis and decisions based upon these data may be flawed, and could produce invalid results and promote poor decisions. Consequently, Network data management programs must proactively establish standards to ensure that data collected through inventory and monitoring activities meet these basic requirements.

NPS Standards

The National Park Service has developed broad guidance in the form of Director's Orders to promote the acquisition and dissemination of high quality data.

Director's Order #11B, "Ensuring Quality of Information Disseminated by the National Park Service," defines 'quality' as incorporating three key components: objectivity, utility, and integrity.

- *Objectivity* consists of: 1) presentation, which focuses on whether disseminated information is being portrayed in an accurate, clear, complete, and unbiased manner within a proper context; and 2) substance, which focuses on ensuring accurate, usable, and reliable information.
- *Utility* refers to the usefulness of the information to its intended users.
- *Integrity* refers to the security of information; e.g., protection from unauthorized access or revision to ensure that the information is not compromised through corruption or falsification.

Director's Order #11B specifies that information must be based on reliable data sources that are accurate, timely, and representative of the most current information available. These standards apply not only to National Park Service (NPS) generated information, but also to information provided by other parties to the NPS if the NPS disseminates or relies upon this information.

Promoting standards goes beyond the criteria identified in the aforementioned Director's Order. From a data management perspective, standards enable people and information systems to perform common tasks; they make it easier to move data from one application to another; and they make it less likely that data will lose utility as technologies change (Biodiversity Conservation Information System (BCIS), 2000). In summary, a well designed data management program begins by establishing a foundation in the form of standards that are universally applied to all projects performed by the Network and to datasets maintained and distributed by the Network.

Quality Control

Eckerson (2003) contends that two of the most difficult problems organizations must overcome when dealing with acquired datasets is understanding the source data and ensuring data quality. Rectifying problems with data is costly, and it may not be possible to fully rehabilitate a poor quality dataset (i.e., unless rehabilitation is completely successful, the utility of a poor quality dataset may be compromised). Reasons for these problems are many, but common explanations include differing study objectives yielding incompatible data, an absence of standardization, and poor documentation. Consequently, when Networks attempt to use data from 'external' sources they are frequently confronted with questions that are, at best, difficult to answer. Given that Networks cannot easily control the quality of external data, Networks must strive to eliminate similar problems with data collected at the behest of the Network. Anticipating problems before they occur is a logical first step. Networks can accomplish this by identifying and defining all required dataset elements, and establishing standards in the form of a Standard Operating Procedure (SOP) manual to which Network staff adheres during data collection, entry, and

validation. Refer to the Network QA/QC Appendix for Network specific details. The SOP's will include, but may not be limited to, documentation for the entire array of data to be collected by a project, variable definitions, variable measurement units, and acceptable measurement ranges. The SOP's will describe the complete data collection process, include copies of field data forms, and thoroughly describe the project specific data verification and validation process.

Quality Assurance

Verification

Verification is the process of ensuring that a value recorded on a field form or keyed into a database is correctly entered (i.e., the entered value is the 'intended' value).

As mentioned previously, evaluating data post-acquisition is a potentially difficult task, but is necessary regardless of the data source. Even if data originates through a Network sponsored project that has been controlled by well designed data acquisition and entry controls, it must be verified and validated prior to being analyzed, interpreted, or distributed.

The following represent logical steps in the verification process:

- *Visual review at data entry.* The data entry technician checks (i.e., verifies) each record after input and immediately corrects any errors;
- *Visual review after data entry.* Data entered into a database is compared visually to the original record, discrepancies are identified and reconciled;
- *Summary queries and tallies.* Error detection queries are used to detect duplicate, omitted, or unassociated records. Similarly, queries can be used to establish the range, mean, median, and other basic statistical parameters to locate problem entries; and,
- *Visual review of spatial data*. The data can be viewed through a Geographic Information System (GIS) to verify the spatial accuracy of entered data.

Validation

Although data may have been correctly transcribed, they may be inaccurate or illogical. For example, pH measurements that exceed 14.0 or that are less than 0 are outside the possible range of values, thus, any pH records that lie beyond this range could be viewed as invalid. Likewise, a stream temperature of 95°C would seem unlikely unless the measurement was taken from within an active geothermal zone.

Invalid data commonly consist of misspelled words, codes, incorrect dates, or out-of-range values for parameters with well-defined limits (e.g., pH, as mentioned above). Logical errors include unreasonable metrics or impossible associations. Histograms, line plots, and descriptive statistics are useful logic and range error detection tools.

Data validators, typically subject matter experts selected for their knowledge of the type of data being reviewed, should not see the elimination of unusually high or low values (outliers) as the primary objective of the quality assurance process. Rather, the purpose of data verification is to determine if odd or unusual values result from data 'contamination.' If data contamination is not apparent, the suspect values may be unusual but not necessarily invalid.

Version Control

Networks possess files originating from a multitude of sources, comprising many formats and frequently including many iterations of a given product. Some of the files are complete, some are works-in-progress, and for many the status cannot be determined. If determining the status of a single file is difficult, determining the status of a series of similarly named files quickly becomes impossible. To address this issue, the Northeast Temperate Network has implemented a version control solution that differentiates earlier versions of a file from later versions.

The Network solution relies on a software application, QVCS, to track and manage different versions of electronic files and applications. The application creates an archive of all changes, assigns version numbers, and incorporates comments and remarks to each file.

Version control may be viewed as a responsibility primarily held by the Network Data Manager, and is thoroughly documented in the Northeast Temperate Network Version Control Appendix.

Data Quality Process Review and Communication

Network data quality techniques must not be viewed as static, inflexible procedures. Rather, data quality techniques are only as good as their ability to adapt to changing requirements, respond to the needs of users, and the ability of the Network to communicate the importance of data quality. Networks should ensure that each dataset complies with FGDC and National Park Service metadata standards; that the data management program has a mechanism to perform periodic audits to help maintain and improve data quality; and that the program is periodically reviewed to evaluate data quality procedures, facilitate the correction of any deficiencies, and promote improvement of the quality assurance / quality control process.

7. Documentation

"Don't find yourself not following these guidelines"
- Anonymous

Documentation "…*helps people who need information resources to find what they need and determine how best to use them*…" (BCIS, 2000). Documentation brings a project to completion by fully describing the process, limitations, application, and restrictions that might apply to a project or dataset. Documentation makes it possible to repeat a project; it helps someone use an application or database; it makes it possible to appropriately use a dataset; and it helps make a project more readily available to potential users. Documentation requirements will vary depending on whether they apply to a dataset, a database, an application, or a project, but in all instances documentation should produce a roadmap to proper usage and understanding. Metadata is a term that is often used synonymously with documentation, but the two are not the same. Metadata is literally defined as data about data. While metadata can be used to describe a project, a system, database, or anything else, it may not adequately explain the steps someone must follow to use an application, or the way a project was conducted. Metadata is but one part of the documentation process – a necessary part because it helps someone find the resource and decide whether the resource is useful. Other documentary elements include, but are not limited to Standard Operating Procedure (SOP) manuals, flow charts, project justification statements, and a record of changes to a sampling protocol. For a complete discussion on the importance of documentation, though with a more specific focus on metadata, see volume 5 of the BCIS manual.

The fundamental purpose of the Inventory and Monitoring Program is to "…*obtain information and data that will help park managers accomplish park management objectives provided for in law and planning documents*…" For the information and data available from the Network to be useful and usable, and for the Inventory and Monitoring Program to fulfill its basic purpose, data and information available from the program must be documented.

- Executive Order 12906, signed by President William Jefferson Clinton in 1994, mandates federal agencies to "*...document all new geospatial data it collects or produces, either directly or indirectly...*" using the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM). In addition, EO 12906 directs agencies to plan for legacy data documentation and provide metadata and data to the public.
- The FGDC Biological Data Profile contains all the elements of the CSDGM and includes additional elements for describing biological data sets. Metadata created in compliance with the Biological Data Profile can be added to the National Biological Information Infrastructure (NBII) Clearinghouse.
- The NPS GIS Committee requires all GIS data layers be described with FGDC standards and the NPS Metadata Profile.

Metadata Tools

There are a number of metadata utilities designed to aid the process of generating FGDC compliant metadata. Among the many, there are program specific applications like Dataset Catalog, developed by the Inventory and Monitoring Program, commercial off–the-shelf (COTS) metadata products like ArcCatalog and the Spatial Metadata Management System (SMMS), and error checking utilities like the Metadata Parser.

Dataset Catalog (http://science.nature.nps.gov/im/apps/datacat/index.htm) was developed to provide a way for Networks to inventory, organize, and maintain information about their dataset holdings, and to offer a way to develop basic metadata. The Northeast Temperate Network made an early commitment to Dataset Catalog by entering 975 records into the system. However, in practice Dataset Catalog did not become the 'day-to-day' utility for managing NETN digital resources and developing metadata that was initially anticipated. ArcCatalog and the NPS Metadata Tools and Editor have effectively supplanted Dataset Catalog as the preferred NETN metadata development utilities, and the NRGIS-Data Store replaces the cataloging function of

Dataset Catalog, at least as far as distributable datasets are concerned. In fact, the Network believes that the value of the NRGIS-Data Store is so great that we enlisted the services of a cooperator to develop a web based utility that links the official Network web-site directly to the NRGIS-Data Store, effectively streamlining the process of searching for Network data. Visit the Network data page to see this utility in action.

ArcCatalog:

ArcCatalog is a file management, or cataloging system that gives users the ability to browse, manage, create, and organize GIS data. In addition, ArcCatalog comes with support for several popular metadata standards making to possible to create, edit, and view information about the data in a variety of formats. Metadata developed using ArcCatalog is stored exclusively as Extensible Markup Language (XML) files. The NPS Integrated Metadata System Plan recommends ArcCatalog for gathering GIS-integrated geospatial metadata.

An optional and highly recommended extension for ArcCatalog is the NPS Metadata Tools and Editor developed by the NPS Midwest Region GIS Technical Support Center that is now supported by the Washington Support Office (WASO). The extension addresses a number of NPS specific issues and incorporates the ability to develop metadata for non-spatial datasets.

The Northeast Temperate Network uses ArcCatalog and the NPS Metadata Tools and Editor extension as its primary metadata generation and editing utility.

Metadata Parser:

The MetaParser (MP) program is used to validate metadata records. It generates a text report indicating errors in the metadata, primarily in the structure, but also in the values of some of the scalar elements where values are restricted by established standards.

Flow Charts

Flow charts are tools that help describe a project by creating a visual representation of the various stages. The example of a flow chart shown in figure 7.1 was created to describe a hypothetical team ski racing program. The fact that figure 7.1 effectively describes a process that is outside the focus of an inventory and monitoring program illustrates the value of a flow chart as a descriptive tool. The same tools and processes used to develop the ski racing example can be used to document development of any project, and are strongly recommended by the Northeast Temperate Network.

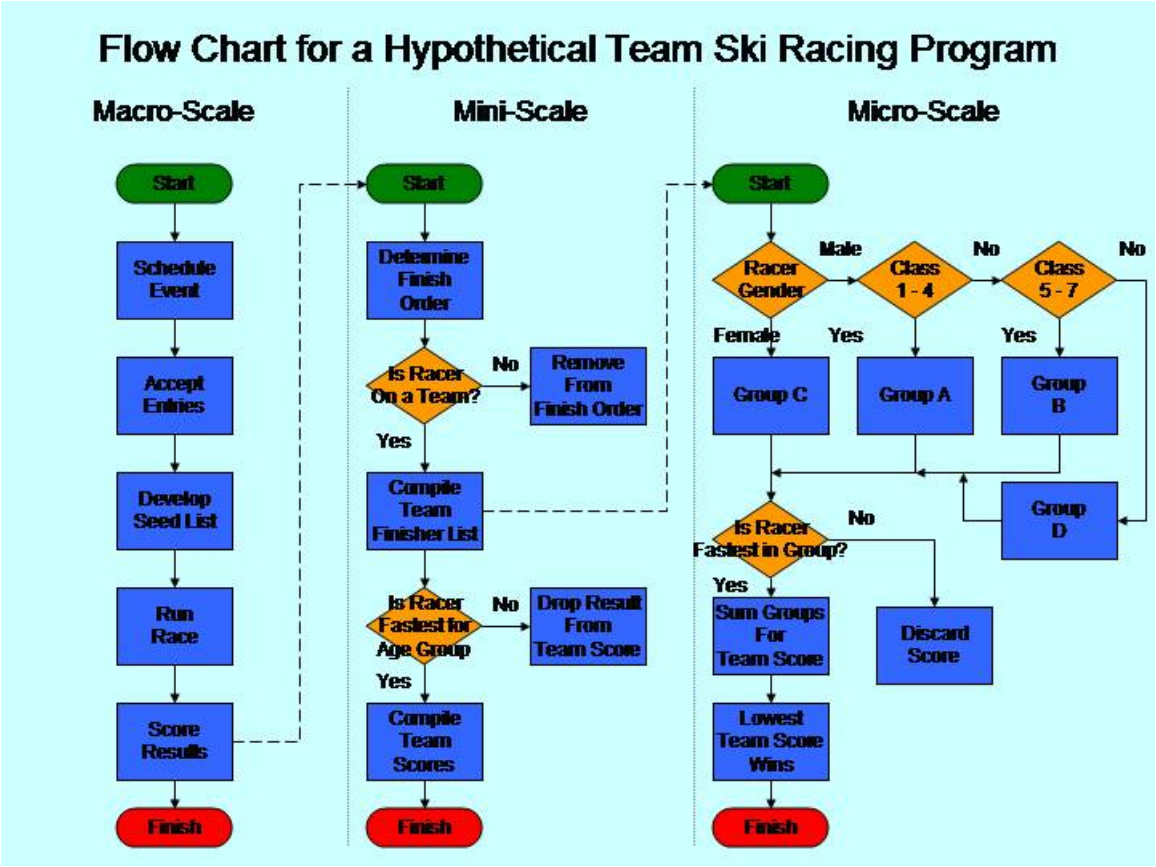
Macro-Scale: A macro-scale flow chart identifies only the major points of a process. A flow-chart at this scale is useful to give a broad overview of a project by eliminating specific detail. A macro-scale flow-chart should contain:

- A general description of what is going to happen.
- Information to help put the project in perspective;
- Sequence of events – what will happen, in what order it will happen, and what is produced.

Mini-Scale: At this level of flow charting, specific project elements identified at the macro-scale are expanded. Decision stages are revealed. A mini-scale flow chart enables a deeper analysis of a project design and specific elements of a design, but may exclude an overall representation of the project. Some portions of the project may be omitted.

Micro-Scale: This is the most detailed level of flow charting. Micro-scale flow charts are very specific, and can outline the finest details incorporated into a project or process. Like a mini-scale flow chart, micro-scale flow charts exclude many of the broader level components that are part of the project.

Figure 7.1 Flow Chart showing Macro, Mini, and Micro-scale detail.



Documentation Process

Project documentation starts at a project's inception, and ties closely to the activities discussed in chapter 4, Project Management and in the Database Design appendix. Documentation begun at the earliest stages of a project can help minimize confusion and miscommunication, and as pointed out in chapter 4, communication is vital. Figure 7.2 (also 5.1) illustrates the data acquisition stages that NETN projects will follow. Documentation is shown to occur throughout all the major stages of the acquisition process, and underscores the importance of documentation throughout the life of a project. In practice, the Northeast Temperate Network plans to front-load the metadata generation process by having cooperators and contractors gather data that will later be used to write the metadata record. For example, the NETN uses the GPS metadata form developed by the Northeast Region to help document the spatial data collection aspect of projects.

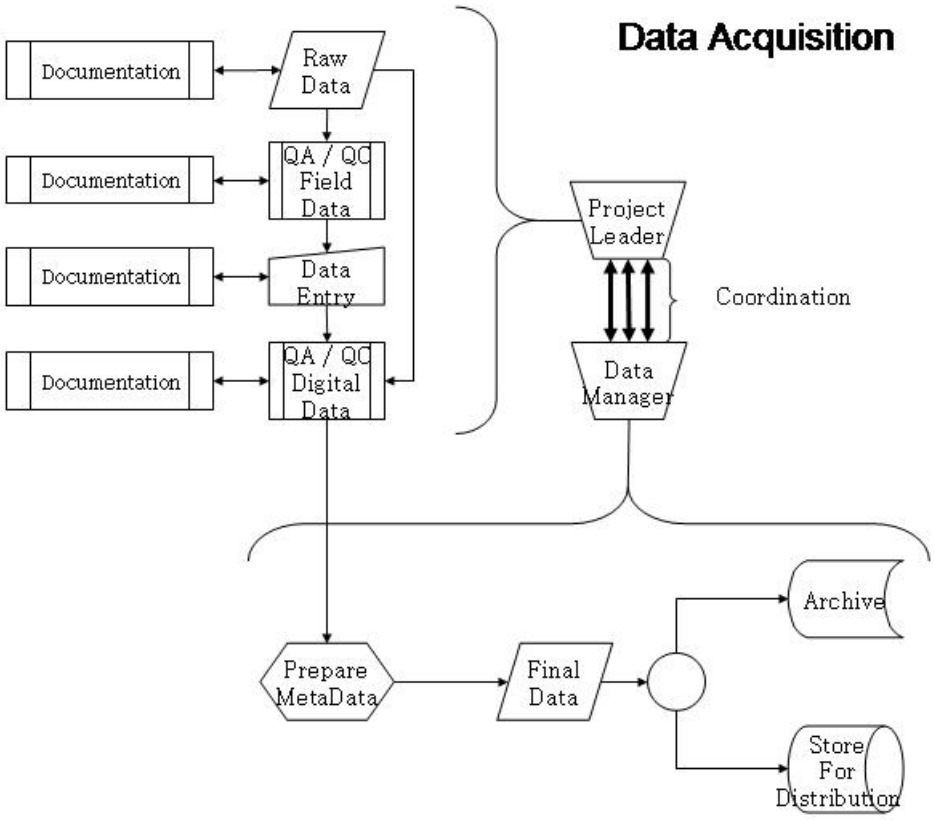
Project Documentation

Project documentation should include everything that will be archived at a project's completion, including proposals, correspondence, flowcharts and models, data, reports, databases, specimens, etc. In general, documentation includes anything that records a project decision, an action, or that is considered to be a deliverable. For a listing of project related elements that should be documented and delivered at the conclusion of a project, refer to the Northeast Temperate Network Guidance on Deliverables Draft.

Biological Metadata

Documentation that describes an entire natural resource or ecological project, not just a spatial subset of the data associated with the project, is termed 'biological metadata.' The Network participates in a regional agreement with North Carolina State University (NCSU) where NCSU develops biological metadata for Network projects. The objective of the agreement is for the Network to send NCSU all project deliverables, including raw data, and NCSU will develop FGDC and NPS compliant metadata records. Refer to the metadata review process appendix for direction on the logical steps the Network follows when submitting data for review and metadata generation.

Figure 7.2. Data Flow Model – Documentation throughout process.



Data Types

Data utilized by the Network are discussed in further detail in chapter 5, but can be broadly grouped into three categories based on origin: legacy; non-programmatic; and programmatic.

Legacy Data

Legacy data, or data that was collected prior to the inception of the Network. Legacy datasets are typically identified as part of data mining efforts. Many legacy data sets will need to be converted to a standard database format. All data and supporting documentation should be compiled, though many legacy datasets will be incomplete.

Non-programmatic Data

Non-programmatic data originates from ongoing efforts by entities outside the direct control of the Network. This could include park managed projects or projects by universities or organizations working independently from the National Park Service. The Network will make every effort to capture and assimilate all relevant data. Metadata will be acquired or developed if necessary. Similar to legacy data, data may need to be converted to a standard database.

Programmatic Data

Programmatic data is data collected by the Network. In most instances, programmatic data will be the most complete form of data that the network will acquire, and will include all appropriate supporting documentation (proposal, SOPs, etc.).

8. Data Dissemination

Background

"*Networks have the information, and communicate with disparate groups. Communicating with these interest groups is an opportunity to demonstrate the value and relevance of the Inventory and Monitoring program*" (Dieffenbach, 2005). This statement was made in the context of data integration, which could be considered a logical next step following data dissemination in the process of data distribution. Why is data integration a step beyond data dissemination? Because data integration demands more than just making data available. Data integration involves bringing data together from various sources and in differing formats to form information, and presenting that information in a form that is useable by the target audience. For example, for many years all cars had odometers, fuel gauges, speedometers, and other instrumentation, but now manufacturers commonly install systems that combine data obtained from each of these sources to give the driver instant information on fuel economy, distance to empty, and time to destination. The data behind this information has always been there, and an inquisitive driver could manually calculate the values, but modern integration technologies have made the information instantly available and constantly current. Similarly, modern data management strategies attempt to do the same thing with data originating from different sources and from different formats. Not coincidentally, one of the hottest Information Technology (IT) issues is termed "dashboarding." An enterprise dashboard brings together data from a wide variety of sources, and integrates it into meaningful chunks of information that managers use to make

decisions. Figure 8.1 Dashboard produced to allow Pioneer Insurance executives to timely monitor insurance key performance indicators (iVizGroup, 2005).

Figure 8.1 shows an example of a corporate dashboard created for an insurance company using a product called iDashboards by the iVizGroup. While the depicted system is for an insurance company application, a keen eye will recognize that a similar strategy could be used to create a dashboard for Natural Resource Managers, Park Superintendents, and anyone else using Inventory and Monitoring program data. For a more complete discussion on performance management, corporate scorecards and dashboards, see BetterManagement.com and SoftwareMagazine.com.

In a white paper by Nimble Technologies (2002), the authors posit that "…*every business is an information business*…", and suggest that business success is dependent on the ability to access pertinent information. This observation logically extends to the business of natural resource management, where the need for timely access to accurate and detailed information is no less important. Accordingly, the Northeast Temperate Network is reviewing the requirements of its audiences and is planning a strategy to use Dashboard technology to provide information that is both targeted and readily available. This approach to integrating information from various sources is termed Enterprise Information Integration, or EII (IPEDO, 2004).

Data Disclosure

All data and associated information from Inventory and Monitoring activities must be assessed to determine their sensitivity. This includes, but is not limited to, reports, maps, metadata, raw and manipulated spatial and non-spatial data. Network staff must carefully identify and manage any information that is considered sensitive. The Network must clearly identify and define those data needing access restrictions and those to make public. Details on classification of sensitive Inventory and Monitoring data shall be outlined in a Standard Operating Procedure (SOP) manual that clearly outlines the individuals responsible for classification and the standards used for doing the classification.

Data Ownership

All data and materials collected or generated using National Park Service personnel and funds are the property of the National Park Service. Thus, NPS bears the responsibility for determining whether it is appropriate to distribute NPS derived data or whether it should be protected from disclosure in accordance with established guidelines (see below). Distribution of data acquired by the National Park Service is governed by any restrictions placed on the data by the data providers or owners of the data. In some instances, it may be possible to distribute a summary of acquired data if distribution of the source data is restricted.

Guidelines and Statutes

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

In some cases, public access to data can be restricted. In accordance with Director's Order #66 (draft), the National Parks Omnibus Management Act (16 U.S.C. 5937), the National Historic Preservation Act (16 U.S.C. 470w-3), the Federal Cave Resources Protection Act (16 U.S.C. 4304) and the Archaeological Resources Protection Act (16 U.S.C. 470hh), the National Park Service is directed to protect information about the nature and location of sensitive park resources. Through these regulations, information that, if released, could result in harm to natural resources can be classified as 'protected' or 'sensitive' and withheld from public release. The Network shall consider each of these acts as well as Department of the Interior FOIA guidance prior to releasing any data and to ensure that its data dissemination procedures are in full compliance.

Data Release Policy

The Network maintains a dataset distribution policy to ensure that:

- Incomplete data, or data that have not been subjected to the full quality control process will not be released;
- All distributed datasets are accompanied by complete metadata;
- Sensitive data are identified and protected from unauthorized access and inappropriate use;
- Any resulting reports or publications acknowledge the National Park Service Inventory and Monitoring Program as the data source.

Internet Accessibility

The Network recognizes that establishing an internet web presence is an important key to making data and information relating to Network parks more readily available, and more widely distributed. The Network established a relationship with the University of Massachusetts to develop a 'home page,' and to create a portal where reports may be downloaded and Network news is distributed. The Network web page is located at the following address:

http://www1.nature.nps.gov/im/units/netn/index.cfm

Data associated with Network projects will be available from this location as data sets are completed and posted to the NRGIS Data store.

To ensure that the NETN web site and all associated pages conform to all federally mandated accessibility requirements, the Network has developed a Web Development Checklist that the cooperator used during development of the page. A copy of the Network Web Development Checklist is contained in the Web Development Checklist Appendix.

Natural Resource and GIS Data Store

The National Park Service has developed a data portal called the Natural Resource and GIS Data Store (NR-GIS) that parks and Networks can use to distribute their data. The system is intended to provide a repository for all park and Network data that is documented with NPS compliant metadata, and the Northeast Temperate Network plans to use the NR-GIS Data Store as the primary destination for all Northeast Temperate Network derived data. The Network's commitment to using the NR-GIS Data Store is best illustrated by the data access portal that is incorporated in the Network web site. The data portal was developed by a cooperator at the University of Massachusetts to reduce the number of places that data are stored and the number of sites that a user must visit to acquire data. The result is a system that effectively streamlines the process of acquiring data from the NR-GIS Data Store.

Reference the NR-GIS web site for background on the NR-GIS system.

Feedback

The Northeast Temperate Network seeks feedback from data users on the utility and format of Network data, ease of acquisition, presence of errors, as well as general comments regarding the Network data distribution system. Feedback may be offered by contacting the Network Office at 802-457-3368, by mail at:
Northeast Temperate Network
Marsh-Billings-Rockefeller NHP
54 Elm Street
Woodstock, VT 050921

or by email (please check the Network "contacts" web page for current email information:

http://www1.nature.nps.gov/im/units/netn/contacts/contacts.cfm).

9. Records Management and Archiving

"Why do I need to save that?"
- Name Withheld

The simple answer to the question is that there are director's orders, agency regulations, and Federal statutes that mandate that certain documents be preserved. Those reasons aside, a Network needs to store and archive program information because these bits of information establish the 'bread-crumb' trail that will lead to a complete understanding of Network activities. Without this record, questions may arise about Network decisions, including expenditures, that cannot otherwise be answered.

The guidance contained in this chapter applies to documents such as final reports prepared by staff or contractors, program administrative documents, contracts and agreements, memoranda of agreements, and other documents related to Network administration, activities, and projects. This chapter also applies to physical items such as natural history specimens, photographs, and audio tapes.

This chapter is not intended to provide a full description of archiving procedures. That information is covered in museum manuals and regulations. It is intended to provide guidance in making the transition from completed products delivered at the conclusion of a project to a secure and long-term storage facility managed by park or regional repositories.

NPS Standards

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

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

Section N of NPS-19 Appendix B, which provides guidelines for managing natural resource-related records, including the results of Inventory and Monitoring Programs, indicates that all natural resource records are considered "permanent" and need to be retained either in an appropriate park museum facility or at the National Archives. It also indicates that non-archival copies of natural resource-related materials are "…*potentially important for the ongoing management of NPS resources*…" and should not, in any instance, be destroyed.

The NPS Museum Handbook provides the overarching guidance for archival procedures. In particular, Part II, Appendix A: Mandates and Standards for NPS Museum Collections, lists the cultural and natural history laws, regulations and conventions for NPS museum collections and should be reviewed prior to any collecting activities.

Network Standards

Direction for managing specimens collected during the implementation of Network projects is described in the Northeast Temperate Network Scope of Collection Statement (SOCS). None of the Network's parks, except Acadia National Park, has expressed an interest in retaining any of the specimens collected during Network activities. The SOCS identifies the limits of the Network's desire to establish a collection. In addition to addressing the disposition of specimens collected during Network projects, the SOCS addresses the disposition of Network operational records.

Curatorial Resources

Several parks in the Northeast Temperate Network have curators and archival specialists on their staff. However, the focus for most of our Network parks is on cultural resources, not natural resources. Acadia National Park is the exception, and in 2005 the curator at Acadia agreed to accept and house the NETN collection.

During the project planning process, project leaders will consult park curators to ensure that all aspects of museum curation are considered, including costs associated with preparation and storage of the materials. As a matter of courtesy, Network parks will have the 'first-right-of-refusal' to curate specimens and project materials associated with work done in their parks. If they decline to retain project specimens and project materials in their collection, the Network will work with staff at Acadia National Park to store the materials. Following project completion, the Network will be responsible for preparing materials in accordance with requirements established by Acadia National Park, or the park where the material will be stored.

10. Glossary

AARWP – Annual Administrative Report and Workplan.

ACAD – Acadia National Park.

AFS – Administrative Finance System.

APPA – Appalachian National Scenic Trail.

Base cartographic data – standard map layers such as boundaries, roads, trails, hydrography; typically not project-specific.

BCIS – Biodiversity Conservation Information System.

BOHA – Boston Harbor Island, a National Park Area.

Certification – a formal process used to officially state that species lists and associated attributes for one or more parks and one or more taxa-categories (birds, fish, vascular plants, etc) have been reviewed for currency, completeness and accuracy

COTS – Commercial-Off-The-Shelf.

Dashboard – a consolicated graphical display of data from disparate sources that is presented to improve the ability of users to interpret the meaning of the data. A dashboard takes its name from an automobile dashboard, which brings together data from a variety of sources and presents it in a way that is easily interpreted by the driver.

EII – Enterprise Information Integration.

ELRO – Eleanor Roosevelt National Historic Park.

EPA – Environmental Protection Agency.

ESRI – Environmental System Research Institute.

FGDC – Federal Geographic Data Committee determines standards for geospatial data.

FOIA – Freedom of Information Act.

GIS – Geographic Information Systems, software used to analyze spatial data and create maps.

GPS – Global Positioning System.

Guidelines – general descriptions of data management practices and procedures.

HOFR – Home of Franklin D. Roosevelt National Historic Site.

I&M – the NPS Inventory and Monitoring Program: http://www1.nature.nps.gov/im/index.html.

IMLS – Institute of Museum and Library Services.

Integration – In the context of data management, integration means the consolidation of data from various sources, and by doing turning these data into information that makes it easier to understand the meaning of the data.

Inventory – "An extensive point-in-time effort to determine location or condition of a resource, including the presence, class, distribution, and status of plants, animals, and abiotic components such as water, soils, landforms, and climate. Inventories contribute to a statement of park resources, which is best described in relation to a standard condition such as the natural or unimpaired state. Inventories may involve both the compilation of existing information and the acquisition of new information. They may be relative to either a particular point in space (synoptic) or time (temporal)." (Source: http://www.nature.nps.gov/im/monitor/index.htm).

IT – Information Technology, composed of hardware, software, digital media, computer and networking services, plus the personnel and knowledge base needed to maintain them.

Local Area Network (LAN) – an interconnected system of computers connected to one or more network servers by cabling and administered as a collective unit.

Long-term monitoring project – a monitoring study designed to assess the status of a natural resource vital sign over time.

MABI – Marsh-Billings-Rockefeller National Historic Park.

MIMA – Minute Man National Historic Park.

MORR – Morristown National Historic Park.

Metadata – information about a data set that tells the who, why, when, where and how the data was collected.

MP – Metaparser.

NBII – National Biological Information Infrastructure, a system of standards for describing biological data.

NCSU – North Carolina State University.

NETN – Northeast Temperate Network.

Network – in the context of the NPS Inventory and Monitoring program, networks are groups of parks linked by geography and shared natural resource characteristics. The network arrangement is intended to provide a minimum infrastructure for initiating natural resource monitoring in the approximately 270 parks with significant natural resources.

NPS – National Park Service.

NPSpecies – The National Park Service database for storing, managing, and disseminating information on all organisms in NPS units.

NRDT – Natural Resource Database Template.

NRGIS Datastore – Natural Resource and Geographic Information System Data Store.

Protocol – a set of sampling procedures and/or experimental processes followed in inventory studies or pilot monitoring projects. In the context of the Inventory and Monitoring Program, this often indicates a monitoring protocol, which is the set of formal documents and sampling processes describing how a vital sign will be monitored. Monitoring protocols are composed of a narrative section, standard operating procedures and supplementary information (databases, reports, tools, hardcopy materials).

Prototype – in the context of the NPS Inventory and Monitoring program, prototype programs play a special role as centers of excellence and as sources of technical guidance and expertise for other monitoring programs across NPS.

ROVA – Roosevelt-Vanderbilt National Historic Sites.

QA/QC – Quality Assurance / Quality Control.

Relational database – a database with data stored in multiple tables where each table contains data on one subject and tables are linked by key fields to link records in different tables.

RIT – Regional Information Technology.

SAGA – Saint Gaudens National Historic Site.

SAIR – Saugus Ironworks National Historic Site.

SARA – Saratoga National Historic Park.

SMMS - Spatial Metadata Management System.

SOCS – Scope of Collection Statement.

Standard Operating Procedures (SOPs) - detailed step-by-step instructions for carrying out sampling procedures in monitoring Protocols; dynamic in nature and frequently updated; can reference overall data management plan, standards and DMOPs.

Standard - a data management practice or policy implemented Service-wide or network-wide (*e.g.,* recommended naming conventions, FGDC metadata format); referenced by overall data management plan, monitoring SOPs and DMOPs.

VAMA – Vanderbilt Mansion National Historic Site.

WASO – Washington Support Office.

WCMC – World Conservation Monitoring Centre.

WEFA – Weir Farm National Historic Site.

XML – Extensible Mark-up Language.

11. References

Busby, John R. 2000. Framework for Information Sharing Series: Volume 5, Metadata. Biodiversity Conservation Information System.

Cox, R.J. 2000. Searching for authority:archivists and electronic records in the New World at the fin-de-siecle. First Monday 5:1. (http://firstmonday.org/issues/issue5_1/cox/index.html)

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

Dieffenbach, Fred W. 2005. Data Integration, a Presentation at the Biennial Inventory and Monitoring Conference, Austin, Tx

Eckerson, Wayne W. 2003. Data Profiling: Minimizing Risk in Data Management Projects. Data Warehousing Institute.

Faccio, S.D. September 2003. A Biological Inventory of Breeding Birds at the Marsh-Billings-Rockefeller National Historic Park and Adjacent Lands, Woodstock, Vermont. Technical Report NPS/NER/NRTR—2005/005. National Park Service. Woodstock, VT.

Institute of Museum and Library Services. 2001. A Framework of Guidance for Building Good Digital Collections. 22pp.

IPEDO Corporation. 2004. Guide to Enterprise Information Integration (EII). 10pp.

iViz Group. 2005. iDashboards Stakes Claim in Insurance: Enterprise Performance Management Dashboards. BusinessWire.Com

Macomber, Hal. 2004. Securing Reliable Promises on Projects: A Guide to Developing a New Practice. Lean Project Consulting, Inc.

Miller, G.A., Galanter, E., & Pribram, K.H. (1960). *Plans and the Structure of Behavior*. New York: Holt, Rinehart & Winston.

National Park Service, 1984. Museum Handbook, Part II, *Museum Records* (revised). Washington, DC:

National Park Service. 2001. NPS Director's Order 19: Records Management. http://www.nps.gov/refdesk/DOrders/DOrder19.html

Nimble Technologies. Next-Generation Data Integration: Harnessing Data for Business Advantage

Oakley, Karen L., Lisa P. Thomas, and Steven G. Fancy. 2003. Guidelines for long-term monitoring protocols. Wildlife Society Bulletin 2003, 31(4):1000–1003

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

Pohl, Ken. 2005. Why Owning the Best Hammer Doesn't Make You a Good Carpenter. DMReview.com

Southwould, Angie. 2002. NPS database specifications for Inventory and Monitoring Studies. 13 pp.

Southwould, Angie. 2002. NPS recommended database strategies, including the Inventory and Monitoring Program database template 53 pp.

Stup, Richard. 2001. Dairy farm business, standard operating procedures: a writing guide. Pennsylvania State University, University Park, PA. 8pp.

Washington Support Office, Inventory and Monitoring Program. 2005. Recommended naming standards for inventory and monitoring databases. National Park Service. Fort Collins, CO. 14 pg.

Wilhelm, H., & C. Brower. 1993. The permanence and care of color photographs: traditional and digital color prints, color negatives, slides, and motion pictures. Preservation Publication Company, Grinnel, IA.

World Conservation Monitoring Centre. 1998. *WCMC Handbooks on Biodiversity Information Management. Volume 2: Information Needs Analysis*. Reynolds, J.H. (Series Editor). Commonwealth Secretariat, London. ix + 23pp.

White, Colin. 2004. Meeting Evolving Business Needs with a Data Quality Service. BI Research