

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
Northeast Region

Date

Brian R. Mitchell, Network Coordinator
Northeast Temperate Network

Date

Fred W. Dieffenbach, Data Manager
Northeast Temperate Network

Date

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 a 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 parks...*” 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 managers, 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.