



Inventory and Monitoring Program

Data Management Plan for the Northern Great Plains Inventory and Monitoring Network



Version: March 2009

Recommended Citation:

Brumm, J. 2009. Data Management Plan: Northern Great Plains Inventory and Monitoring Network. National Park Service Northern Great Plains Inventory and Monitoring Network Report. NGPN/2009.

Acknowledgements

The creation of this plan involved many individuals both within and without the National Park Service. The Data Managers from the first I&M networks to complete their planning process developed a template and drafts of the chapters that have been included in this Plan. Subsequent data management plans have drawn on these initial chapters, and have enhanced and expanded on the original material. Credit is due to all who have preceded NGPN in the completion of network data management plans, and for making these plans available to the rest of us. In addition, the national-level data management staff have provided the vision and created the tools that have been essential to this work. We extend our thanks and appreciation for their support, insight, and guidance.

Change History

This document will initially be recognized as Version 1.0 upon adoption by the NGPN. Version numbers will be incremented by decimals (e.g., Version 1.0 to Version 1.1) when there are minor modifications that do not affect requirements or procedures included in the plan. Version numbers will be incremented by a whole number (e.g., Version 1.3 to Version 2.0) when a change is made that significantly affects requirements or procedures.

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

[illegible]

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION	1
1.1 CONTEXT OF THE DATA MANAGEMENT PLAN	3
1.2 DATA MANAGEMENT CATEGORIES	3
1.3 PLAN SCOPE	4
1.4 PLAN OBJECTIVES	5
1.5 PLAN REVISIONS.....	6
CHAPTER 2. DATA MANAGEMENT ROLES AND RESPONSIBILITIES	7
2.1 PRIMARY ROLES AND RESPONSIBILITIES FOR DATA MANAGEMENT	7
2.2 PROJECT MANAGEMENT	9
2.3 DATA STEWARDSHIP	9
2.4 THE HUB OF DATA STEWARDSHIP	10
2.5 DATA MANAGEMENT COORDINATION.....	13
CHAPTER 3. DATA MANAGEMENT PROCESS AND WORKFLOW.....	15
3.1 PROJECT WORK FLOW.....	15
3.2 CORE DATA MANAGEMENT ELEMENTS OF THE PROJECT WORK FLOW	16
3.3 DATA LIFE CYCLE.....	17
3.4 INTEGRATING AND SHARING DATA PRODUCTS	20
CHAPTER 4. DATA MANAGEMENT INFRASTRUCTURE	22
4.1 ROLE OF INFRASTRUCTURE IN THE DATA MANAGEMENT SYSTEM.....	22
4.2 NPS INFRASTRUCTURE - NGPN COMPONENTS.....	23
4.3 NPS INFRASTRUCTURE - NATIONAL-LEVEL APPLICATION ARCHITECTURE	24
4.4 DATABASE DESIGN	27
4.5 GIS DATA STRUCTURE	28
CHAPTER 5. DATA ACQUISITION AND PROCESSING	29
5.1 PROGRAMMATIC DATA.....	29
5.2 NGPN DATA COLLECTION TOOLS	30
5.3 PROGRAMMATIC DATA ENTRY	32
5.4 CHANGES TO NGPN DATA COLLECTION PROCEDURES	33
5.5 NPS DATA SOURCES.....	33
5.6 EXTERNAL DATA.....	35
5.7 DATA DISCOVERY AND HARVESTING.....	35
5.8 DATA COMPILATION, PROCESSING, AND INTEGRATION.....	37
CHAPTER 6. QUALITY ASSURANCE (QA) AND QUALITY CONTROL (QC).....	38
6.1 DATA QUALITY.....	38
6.2 ROLES AND RESPONSIBILITIES.....	41
6.3 GENERAL OPERATIONS	42
6.4 QUALITY ASSURANCE AND QUALITY CONTROL MECHANISMS	44
6.5 GOALS AND OBJECTIVES	45
6.6 DATA COLLECTION.....	45
6.7 DATA ENTRY/IMPORT	47
6.8 VERIFICATION AND VALIDATION PROCEDURES	48
6.9 DATA QUALITY REVIEW AND COMMUNICATION	52

CHAPTER 7. DATA DOCUMENTATION.....	54
7.1 MANDATE FOR DOCUMENTATION.....	55
7.2 DOCUMENTATION PROCESS.....	56
7.3 NON-PROGRAMMATIC DATA DOCUMENTATION.....	58
CHAPTER 8. DATA ANALYSIS AND REPORTING	59
8.1 IMPORTANCE OF DATA REPORTING.....	59
8.2 TYPES OF REPORTS	59
8.3 OVERVIEW OF PROCESS FOR DATA REPORTING.....	62
CHAPTER 9. DATA DISSEMINATION	63
9.1 NATIONAL PARK SERVICE POLICY ON DATA OWNERSHIP	63
9.2 MECHANISMS FOR DATA DISTRIBUTION	64
9.3 NETWORK WEBSITE.....	66
9.4 FREEDOM OF INFORMATION ACT AND SENSITIVE DATA.....	66
9.5 DATA RELEASE POLICY	68
CHAPTER 10. DATA MAINTENANCE, STORAGE, AND ARCHIVING.....	70
10.1 NPS STANDARDS.....	70
10.2 PARK UNIT AND NETWORK STANDARDS	71
10.3 ROLE OF CURATORS IN CURATION AND MUSEUM STORAGE	71
10.4 NETWORK ARCHIVING PROCESS	72
10.5 ARCHIVE AND STORAGE LOCATIONS	73
10.6 NPS AUTOMATED NATIONAL CATALOG SYSTEM (ANCS +)	74
10.7 PHOTOGRAPHS	74
10.8 SPECIMENS AND OTHER MATERIALS.....	74
CHAPTER 11. IMPLEMENTATION	76
11.1 COMMON DATA MANAGEMENT SINS.....	76
11.2 YEARS 1-5 OF IMPLEMENTATION.....	77
11.3 YEARS 5+ OF IMPLEMENTATION.....	77
LITERATURE CITED.....	79
LIST OF TABLES	
TABLE 1.1. CATEGORIES OF DATA PRODUCTS AND PROJECT DELIVERABLES	4
TABLE 2.1. DATA STEWARDSHIP CATEGORIES FOR ALL NETWORK PERSONNEL AND COOPERATORS.	7
TABLE 2.2. PROGRAMMATIC ROLES AND DATA STEWARDSHIP RESPONSIBILITIES	8
TABLE 7.1. DATA DOCUMENTATION RESPONSIBILITIES.....	55
TABLE 7.2. METADATA FILE FORMATS.....	57
TABLE 9.1. PRIMARY REPOSITORIES FOR NGPN INFORMATION AND ASSOCIATED SPECIMENS.....	65
List of Figures	
FIGURE 1.1. PARK UNITS OF THE NORTHERN GREAT PLAINS NETWORK	2
FIGURE 1.2. CONTEXT OF THE NGPN DATA MANAGEMENT PLAN	3
FIGURE 2.1. CORE DATA MANAGEMENT RESPONSIBILITIES FOR PROJECT AND DATA MANAGERS	11
FIGURE 3.1. GERALIZED PROJECT WORK FLOW MODEL.....	17

FIGURE 3.2.	DIAGRAM OF THE TYPICAL PROJECT DATA LIFE CYCLE	20
FIGURE 4.2.	MODEL OF THE NATIONAL-LEVEL APPLICATION ARCHITECTURE.....	25
FIGURE 6.1.	GENERAL COURSE OF DATA AND ASSOCIATED QA/QC PROCEDURES.....	40
FIGURE 6.2.	AMOUNT OF QA/QC REQUIRED	41
FIGURE 7.1.	PROJECT AND DATA DOCUMENTATION.....	54
FIGURE 7.2.	METADATA REPOSITORIES FOR THE NGPN	57
FIGURE 9.1.	SCHEMATIC OF DATA FLOW FOR WATER QUALITY DATA.....	65
FIGURE 10.1.	SCHEMATIC OF NATURAL HISTORY CURATION FLOW	72

TABLE OF CONTENTS - APPENDICES

APPENDIX A.	BASIC RESOURCE INVENTORIES AND REPOSITORIES FOR THE NGPN.....	80
APPENDIX B.	PROJECT INFORMATION MANAGEMENT PROCESS & WORK FLOW	81
APPENDIX C.	ACRONYMS.....	85
APPENDIX D.	GLOSSARY	86

EXECUTIVE SUMMARY

The central mission of the NPS Inventory and Monitoring Program is to provide timely and usable scientific information to park managers about the status and trends of park resources. To meet this challenge, we need an information management system that can effectively produce, maintain and distribute the products of scientific work done in our parks.

Good data management is the means by which scientific information about our natural resources can become a part of our National Park Service heritage. Data management refers to the framework by which data are acquired, maintained, and made available. Data management is not an end unto itself, but instead is the means of maximizing the quality and utility of our natural resource information. A robust system for data management is particularly important for long-term programs where the lifespan of a data set will be much longer than the careers of scientists who collected the data. Seen in this way, it becomes obvious that data management is vital to the success of any long-term monitoring initiative.

The success of our program hinges upon our ability to produce, manage and deliver timely and usable information to park managers and other audiences. Therefore, we will ensure the quality, interpretability, security, longevity and availability of our natural resource data. In implementing a data management system we will strive for the following:

- Confidence in the security and availability of natural resource data and related information
- Easy access to most information, and appropriate safeguards for sensitive information
- Awareness of the intended use and limitations of each data set
- Infrastructure and documentation that encourages data exploration
- Compatibility of data sets for exploration and analysis at larger scales and across disciplines
- Implementation of standards and procedures that facilitate information management, and that reinforce good habits among staff at all levels of project implementation – project leaders, technicians, and volunteer data collectors
- A proper balance between the standards needed to ensure quality and usability, and the flexibility to meet specific needs and encourage innovation
- A natural resource culture which views data not as a commodity but as the lifeblood of our work

The Northern Great Plains Data Management Plan outlines how we intend to implement and maintain a system that will serve the data and information management needs of our Inventory and Monitoring Program. This plan reflects our commitment to establishing and maintaining a robust system for data management to ensure the availability and usability of high-quality natural resource information.

Data Management Goals

The goal of our data management program is to ensure the quality, interpretability, security, longevity and availability of ecological data and related information resulting from resource inventory and monitoring efforts:

- *Quality* – Awareness of the value of information and its underlying data is fundamental to its proper use. Our objective is to ensure that appropriate quality assurance measures are taken during all phases of project development, data acquisition, data handling, summary and analysis, reporting, and archival. These should reflect current best practices and meet rigorous scientific standards. Since standards and procedures can only accomplish so much, an important part of quality assurance is to continually encourage careful attitudes and good habits among all staff involved in creating, collecting, handling, and interpreting data.
- *Interpretability* – A data set is only useful if it can be readily understood and appropriately interpreted in the context of its original scope and intent. Data taken out of context can lead to misinterpretation, misunderstanding, and bad management decisions. Similarly, data sets that are obscure, complex or poorly documented can be easily misused. Sufficient documentation should accompany each data set and any reports and summaries derived from it, so as to ensure that users will have an informed appreciation of its applicability and limitations.
- *Security* – Our objective is to make certain that both digital and analog forms of source data are maintained and archived in an environment that provides appropriate levels of access to project managers, technicians, decision makers, and others. Our data management program will take advantage of existing systems for network security and systems backup, and augment these with specific measures aimed at ensuring the long-term security and integrity of our data.
- *Longevity* – Countless data sets have been lost over time simply because they were not sufficiently documented and organized when they were created. Too often data are left in a condition that renders them effectively irretrievable – either because the format is outdated, or more often because there is not enough documentation to inform subsequent users of the data collection methods, the scope and intent of the data set, the quality assurance procedures, the data format, or the intended use. Without sufficient information about a data set we lose confidence in its quality and applicability, which leaves it useless and unused. The longevity of a data set can be enhanced by thorough documentation, by maintaining the data in a widely interpretable format, and by appropriate archival measures. Although this requires an initial investment of time and effort upon creation of the data set, this investment almost certainly pays off over time because the data set is much more likely to be used if it is properly documented and formatted.
- *Availability* – Natural resource information can only be useful for informing decisions if it is available to managers at the right time and in a usable form. Our

objective is to expand the availability of natural resource information by ensuring that the products of inventory and monitoring efforts are created, documented and maintained in a manner that is transparent to the potential users of these products.

Although our primary responsibility is for ecological data and related information derived from Inventory and Monitoring Program activities, it is anticipated that the scope of our data management efforts will expand to include all significant natural resource data specific to the Northern Great Plains Network and the parks of which it is composed.

Priorities of Natural Resource Data

The priorities for Network data management efforts are to:

- Produce and curate high-quality, well-documented data originating with the I&M Program
- Assist with data management for current projects, legacy data and data originating outside the I&M Program that complement program objectives
- Help ensure good data management practices for park-based natural resource projects that are just beginning to be developed and implemented

Data Stewardship Roles and Responsibilities

Every individual involved in the I&M Program has data stewardship responsibilities, whether they are involved in the production, analysis, management, or end use of data. It is critical that individuals perform their duties in accordance with the Data Management Plan, the specific Vital Signs monitoring protocols, and any applicable Standard Operating Procedures (SOPs) relating to the data collection methods. Monitoring protocols and SOPs will describe specific roles and responsibilities in detail.

Infrastructure and System Architecture

Infrastructure refers to the network of computers and servers that our information systems are built upon. The NGPN relies heavily on national and regional information technology, personnel, and resources to maintain its computer infrastructure. This includes, but is not limited to: configuration of computers, servers and other related hardware; software installation and support; email administration; security updates; virus-protection; telecommunications; and computer networking.

The infrastructure supports these required functions:

- Provides a central repository for master data sets
- Provides controlled subsets of data for local computing
- Provides a means for uploading and downloading data for both NPS and public
- Supports desktop and internet applications
- Provides security, stability, and backups

The NGPN has developed procedures to maintain, store and archive data to ensure that data and related documents (digital and analog) are both accessible and secure. Content, format, and documentation must be up-to-date so that the data can be easily accessed and properly used. Data must also be physically secure against environmental hazards, catastrophe, and human malice.

Most data maintenance will be performed on the NGPN file server and on Service-wide servers maintained by the I&M Program. The NGPN data management staff are responsible for ensuring that regular data backups are being implemented for all network data. Data and information on NGPN and NPS servers will be kept current and all updates will be described in accompanying documentation. Information files will be properly cataloged and maintained on the NGPN website, and the latest versions of primary data will be available in formats that reflect common usages in the resource management community.

Project data will be electronically archived as stand-alone products and will include:

- Project documentation
- Data in raw, verified, and analyzed conditions
- Respective metadata
- Supporting files (e.g., digital photographs and maps)
- All associated reports

Final deliverables from project data will be added to existing libraries and databases.

Database Design Strategies

The project manager and the data manager will work together to develop conceptual and logical data models to understand the data life cycle and flow of the data collection process. It is necessary to understand where data collection begins (for example, a visit to a site) and what steps are involved in data processing. Understanding conceptual and logical relationships can also help identify how the project information can best be presented.

The NGPN Data Management Plan specifies the standards by which data will be handled. Understanding the relationships between data components is the key to successfully developing and using a database. If the relationships are misunderstood, data entry may be tedious and data output may be cumbersome.

Acquiring and Processing Data

The types of data handled by the I&M Program fall into three general categories:

- *Program data* are produced by projects that are either initiated (funded) by the I&M Program or involve the I&M Program in another manner (e.g., natural resource inventories and Vital Signs monitoring projects).
- *Non-program legacy (existing) data* are produced by NPS entities without the involvement of the I&M Program (e.g., park or regional projects).
- *Non-program external data* are produced by agencies or institutions other than the National Park Service (e.g., weather and air quality data).

Most data acquired by the Network will be collected during field-based inventory and monitoring studies or will be discovered through data mining initiatives. The methods and tools required for the collection of field data (e.g., paper data forms, field computers, automated data loggers, and GPS units) will be specified in individual monitoring protocols and study plans. Field crew members must closely follow the standard

operating procedures (SOPs) in the project protocol. Techniques for handling data acquired from non-program sources, such as data downloaded from other agencies, will also be specified in individual monitoring protocols.

Ensuring Data Quality

High quality data and information are vital to the credibility and success of the I&M Program, and everyone plays a part in ensuring that products conform to data quality standards.

Specific procedures to ensure data quality must be included in the protocols for each Vital Sign. Although many quality assurance/quality control (QA/QC) procedures depend on the individual Vital Signs being monitored, some general concepts apply to all.

Examples of QA/QC practices include:

- Training field crew members uniformly
- Standardizing field data sheets with descriptive data dictionaries
- Use of handheld computers and data loggers
- Maintenance and calibration of equipment
- Procedures for handling data in the field
- Incorporating database features to minimize transcription errors (e.g., validation rules, range limits, pick lists, routines to import data from data loggers)
- Verification and validation, including automated error-checking database routines

Quality assurance methods should be in place at the inception of any project and should continue through all stages of the project. The final step in project quality assurance is the preparation of summary documentation that assesses the overall quality of the data. The project manager will compose a statement of data quality that will be incorporated into the formal metadata. Metadata for each data set will also include information on quality assurance procedures specific to the project and results of the review.

Data Documentation

Documenting data sets, data sources, and the methodology by which the data were acquired establishes the basis for interpreting the data and using it appropriately. At a minimum, all data managed by the Network will require documentation of the project, formal metadata compliant with Federal Geographic Data Committee (FGDC) standards, and data dictionaries and Entity Relationship Diagrams (ERDs) for all tabular databases.

Data documentation will be available via the NGPN website as well as the national I&M Program's NPS Data Store.

Data Analysis and Reporting

The mission of the I&M Program is to provide useful information to managers and scientists; therefore, providing meaningful results from data summary and analysis is a cornerstone of the NGPN data management program. Each monitoring protocol establishes requirements for scheduled and requested data analysis and reporting. Based on such requirements, the associated databases for the protocols will include functions to summarize and report directly from the database and will allow output in formats that can

be easily imported to other analysis software programs. In addition to tabular and charted summaries, the Network provides maps of natural resource data and GIS analysis products to communicate spatial locations, relationships and geospatial model results.

Data Dissemination

The NGPN data dissemination strategy aims to ensure that:

- Data are easily discoverable and obtainable
- Only data subjected to complete quality control are released, unless release is necessary in response to a Freedom of Information Act (FOIA) request
- Distributed data are accompanied by appropriate documentation
- Sensitive data are identified and protected from unauthorized access and inappropriate use

Users have various means at their disposal to browse, search, and acquire Network data and supporting documents. Data products can be accessed via:

- Links to public data products on a NGPN public website.
- The NPS Data Store. Distribution instructions for each data set will be provided in the respective metadata.
- Service-wide databases, such as NPSTORET, NPSpecies, and NatureBib
- Regional, Network, or park data servers protected with read-only access
- External repositories such as the EPA STORET, U.S. Geological Survey, U.S. Forest Service, and the Midwestern Regional Climatic Center
- FTP sites, CDs, DVDs, or hard drives, as appropriate

Ownership, FOIA, and Sensitive Data

Northern Great Plains Network products are considered property of the NPS; however, the Freedom of Information Act (FOIA) establishes that any person may access federal agency records that are not protected from disclosure by exemption or by special law enforcement record exclusions. The NPS is directed to protect information about the nature and location of sensitive park resources under one Executive Order and four resource confidentiality laws:

- Executive Order No. 13007: Indian Sacred Sites
- National Parks Omnibus Management Act (NPOMA; 16 U.S.C. 5937)
- National Historic Preservation Act (16 U.S.C. 470w-3)
- Federal Cave Resources Protection Act (16 U.S.C. 4304)
- Archaeological Resources Protection Act (16 U.S.C. 470hh)

When any of these regulations are applicable, public access to data can be restricted. If disclosure could result in harm to natural resources, the records may be classified as 'protected' or 'sensitive'. The NPS recognizes the following resources as sensitive:

- Endangered, threatened, rare, or commercially valuable National Park System resources
- Mineral or paleontological sites
- Objects of cultural patrimony

- Significant caves

The NGPN will comply with all FOIA restrictions regarding the release of data and information, as instructed in NPS Director's Order #66 and accompanying Reference Manuals 66A and 66B. Managing natural resource information that is sensitive or protected requires the following steps:

- Identification of potentially sensitive resources
- Compilation of all records relating to those resources
- Determination of which data must not be released in a public forum
- Management and archiving of those records to avoid their unintentional release

Classification of sensitive data will be the responsibility of Network staff, park superintendents (or their delegates), and project managers. Network staff will classify sensitive data on a case-by-case, project-by-project basis and will work closely with project managers to ensure that potentially sensitive park resources are identified, that information about these resources is tracked throughout the project, and that potentially sensitive information is removed from documents and products that will be released outside the Network.

Data Archiving and Records Management

In most instances, administrative documents, natural history specimens, photographs, audio tapes and other materials are essential companions to the digital data. 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). Director's Order 19 states that all records of natural and cultural resources and their management are considered mission-critical records (necessary for fulfillment of the NPS mission) and must be permanently archived.

The NGPN data management strategy includes helping project managers comply with archival directives. Whenever possible, physical items that are products of a project (e.g., reports, maps, photographs, or notebooks) will be cataloged and archived by the park(s) involved with the project. When this is not possible, these physical items will be stored in NGPN offices. Physical specimens, such as plants and animals, will be accessioned and housed at appropriate archival institutions.

Water Quality Data

Water quality data are managed according to guidelines from the NPS Water Resources Division. In accordance with these guidelines, the desktop database application NPSTORET will be used to enter, store, document, and transfer water quality data. The NGPN oversees the use of NPSTORET per the Network's water quality monitoring protocols and ensures that data are transferred at least annually to the NPS Water Resource Division for upload to the STORET database.

Implementation

The NGPN Data Management Plan contains practices that may be new to staff and principal investigators. With a few exceptions, however, the DMP does not include any new requirements. Almost every requirement stipulated in the Plan comes from law, Director's Orders, or the I&M Program. The DMP helps put these requirements into context and provides operational guidance to meet these requirements.

1. INTRODUCTION

The central mission of the NPS Inventory and Monitoring Program is to provide timely and usable scientific information to park managers about the status and trends of park resources. To meet this challenge, we need an information management system that can effectively produce, maintain and distribute the products of scientific work done in our parks. Good data management is the means by which a thorough understanding of the value of scientific information about our natural resources can become a part of our National Park Service heritage.

Data management refers to the framework by which data are acquired, maintained, and made available. A robust system for data management is particularly important for long-term programs where the lifespan of a data set will span the careers of several scientists. Good data management is vital to the success of any long-term monitoring initiative. The goals of the I&M Program can only be achieved through the development of a modern information management infrastructure (i.e., staffing, hardware, software) and good data management procedures. Data collected by NPS staff, cooperators, researchers and others must be entered, quality-checked, analyzed, reported, archived, documented, cataloged, and made available to others for management decision-making, research, and education.

This chapter summarizes the goals and objectives for our Northern Great Plains Network (NGPN) data management system. This Data Management Plan (DMP) presents the overarching strategy for ensuring that Program data are documented, secure, accessible, and useful for decades into the future. The plan also refers to other guidance documents, standard operating procedures (SOPs), and detailed protocols that convey more specific standards and steps for achieving our data management goals. The plan acts as a foundation upon which to build as new protocols are developed, advances in technology are adopted, and new concepts in data management philosophy are accepted.

Northern Great Plains Network

The Northern Great Plains Network is comprised of a diverse group of 13 NPS units (Figure 1.1). Several parks, such as Badlands NP, the Missouri NRR, and Theodore Roosevelt NP, consist of discrete and non-contiguous management units. Two units—the Missouri NRR and the Niobrara NSR—are unique in that the Park Service owns no land or only small tracts within park boundaries.



Figure 1.1 Park units of the Northern Great Plains Network

The NGPN I&M Program focuses on basic natural resource inventories and development and implementation of a long-term ecological monitoring program. The natural resource inventory is composed of twelve basic data sets (Appendix A) and forms the basis for planning and development of the Network's monitoring program. The biological inventories focus on vertebrate taxa and vascular plants. These inventories help meet programmatic I&M goals and provide the NGPN with baseline ecological assessments from which long-term monitoring may begin. Data resulting from the natural resource inventories and other short-term studies will be subjected to the same management

standards as monitoring data. Currently, NGPN is working to build a holistic picture of change across the Network ecosystems – specifically, to detect change in major ecological components and in the relationships among those components. Specifics on the status and progress of Network Vital Signs are available on the NGPN website at: <http://science.nature.nps.gov/im/units/ngpn/index.cfm>. Initial NGPN Vital Signs monitoring (for some protocols) will begin in 2009. In addition, monitoring is currently being carried out in Agate Fossil Beds National Monument (AGFO) and Scotts Bluff National Monument (SCBL) by the Heartland Network. These parks were part of the Prairie Cluster prototype monitoring effort. It is planned for NGPN to assume monitoring from Prairie Cluster for these two parks in 2010.

Natural Resource Inventories

The natural resource inventory is composed of twelve basic data sets (Appendix A) and forms the basis for planning and development of the Network’s monitoring program. The biological inventories focus on vertebrate taxa and vascular plants. These inventories help meet programmatic I&M goals and provide the NGPN with baseline ecological assessments from which long-term monitoring may begin. Data resulting from the natural resource inventories and other short-term studies will be subjected to the same management standards as the monitoring data.

1.1 Context of the Data Management Plan

The National Park Service’s I&M Program provides various general data management guidelines for the I&M 32 networks. The NGPN data management strategy draws from these guidelines and formalizes them as Network policy. More detailed data management strategies are documented in standard operating procedures (SOPs) specific to a given data collection effort. These SOPs fall under the aegis of this plan and adhere to the guidance, strategies and policies herein. Figure 1.2 depicts this data management guideline/policy inheritance model for the NGPN.

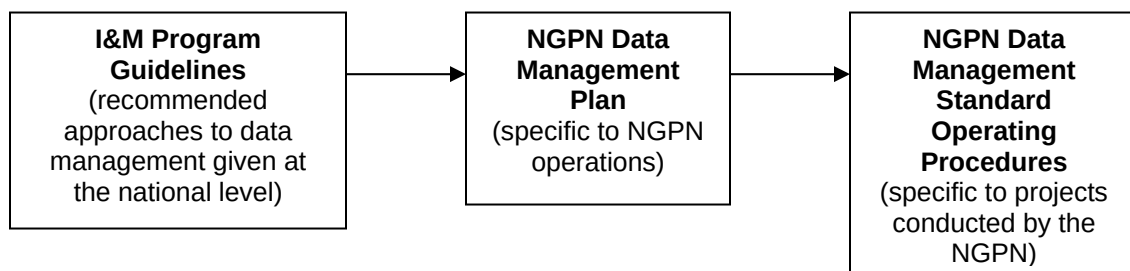


Figure 1.2 Context of national I&M Program guidelines, the NGPN Data Management Plan, and detailed data management procedures for a project

1.2 Data Management Categories

Natural resource data are the vital building blocks for our evolving ecological understanding of park resources. But a set of data – whether collected the previous year or 20 years ago – must be accompanied by sufficient information on how and why it was collected to maintain its value beyond the lifetimes of those who collected it. Therefore, a data management strategy cannot simply attend to the tables, fields, and values that make

up a data set. There must also be a process for documenting the context that makes it interpretable and valuable.

The term “data” frequently encompasses other products that are generated alongside the tabular and spatial data that are the primary targets of our data management efforts. These products fall into five general categories: raw data, derived data, documentation, reports, and administrative records (Table 1.1).

Table 1.1 Categories of data products and project deliverables

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

It is our intent to integrate the products listed in Table 1.1 such that a project’s context within the overall program is maintained. This requires a holistic view of how natural resource data are generated, processed, finalized, and made available. Such an understanding allows us to tailor our data management strategies to meet our objectives in the most efficient manner possible. Chapter 5 explores more thoroughly the types of data and information managed by the NGPN and the general process by which they are generated.

1.3 Plan Scope

This Data Management Plan stipulates the broad principles for handling ecological data from their most elementary state to their final condition in reports or other formats. This plan places heavy emphasis on data generated by, or considered important to, the NGPN monitoring program. However, the management principles presented may be applied to any data-gathering effort, short- or long-term, made by park resource staff. Core data management elements addressed in this plan include data collection, backup, quality control/assurance, validation, verification, documentation, archival, standard manipulation/analysis, organization, availability/access and distribution.

This plan is organized as follows:

- *Chapter 1:* Introduction, background information, plan overview.
- *Chapter 2:* Outline of the data management personnel and their roles and responsibilities. Where possible, shared responsibilities are noted.
- *Chapter 3:* Overview of the I&M data management process. Aspects specific to the NGPN, including Vital Sign monitoring components, are addressed in context of the overall I&M Program.

- *Chapters 4 through 11:* More specific data management infrastructure, standards and practices related to data acquisition, quality control, and documentation.
- *Appendices:* Information that directly supports this document as well as supplementary material supporting data management operations in general. Specifics regarding data management procedures will reside in the NGPN Data Management Standard Operating Procedures (SOPs).

1.4 Plan Objectives

It is important to understand the objectives of this document relative to the national I&M Program and data management goals. The goal of the NPS I&M Program is to provide scientifically and statistically sound data to support management decisions for the protection of park resources. The goal of data management is to ensure the quality, interpretability, security, longevity, and availability of ecological data and related information resulting from resource inventory and monitoring efforts. The goal of the Data Management Plan is to outline the procedures and work practices that support effective data management.

The NGPN seeks to outline sound data management practices that can be easily adopted by any resource management project. When appropriate, the NGPN will promote effective data management by park personnel involved in the collection of natural resource data in the parks. By establishing robust data management and maintaining the transparency of its inner workings, the NGPN hopes to lead by example and positively affect all data management within the parks.

Objectives for this plan fall into two categories. Overall objectives describe how data management fits within the larger I&M Program, and specific objectives detail the procedures and policies necessary for an effective data management program.

Overall Objectives:

- Outline the long-term goals of a comprehensive data management strategy for the Northern Great Plains Network
- Associate data management goals with the long-term goals of the Network and Service-wide I&M Program
- Outline the procedures and work practices that support effective data management
- Guide current and future Network staff to ensure that sound data management practices are followed in any data gathering efforts conducted or administered by the NGPN
- Guide the enhancement of legacy data to match formats and standards put forth in this plan
- Encourage effective data management practices as an integral part of project management so all data are made available and usable for park management decisions now and into the future
- Optimize and promote interagency sharing and development of data, software applications and analyses

Specific Objectives:

- Establish roles and responsibilities of program staff for managing data

- Identify necessary elements for a functional data management program and describe any anticipated changes to those elements
- Establish an organization schema for Network data and information so that they are retrievable by staff, cooperators, and the public
- Establish basic quality control standards
- Establish standards for data, data distribution, and data archiving to ensure the long-term integrity of data, associated metadata, and any supporting information

1.5 Plan Revisions

The plan will be formally reviewed and revised at least every 5 years. Informal review and revision will be ongoing to meet the changing needs of the program. Revisions occurring at intervals between the formal 5 year review will be documented in the change history at the beginning of the document. The latest version of the DMP will always be posted on the NGPN website at <http://www.nature.nps.gov/im/units/NGPN/DataMgt.htm>.

2. DATA MANAGEMENT ROLES AND RESPONSIBILITIES

Fundamentally, data management is as much about people and organizations as it is about information technology and databases. Every person in an organization manages data and information at some level. Their understanding of how their actions affect the quality and utility of the resulting data and information is a critical element of the data management process. Every individual involved with a Network project must share with the data manager responsibilities for the production, analysis, management, and/or end use of the data (Table 2.1). By communicating and working with all the responsible individuals, creating and maintaining data infrastructure and standards, and understanding program and project requirements, the data manager supports and guides people and activities to meet the Network's data management objectives.

This chapter clarifies roles and outlines the principal and ancillary data management responsibilities of Network staff, cooperators, and other participants in any Network project. Individuals who carry out monitoring protocols and inventory study plans are responsible for reading and understanding guidance at two levels — the overarching level of this Data Management Plan and the specific, task-oriented level of the respective protocol or study plan.

Table 2.1 Data stewardship categories for all Network personnel and cooperators

Stewardship Category	Related Activities	Position*
Production	Collecting data or information from any original or derived source such as recording locations, images, measurements, and observations in the field; digitizing source maps; keying in data from a hardcopy source; converting existing data sources; image processing; quality checking / data proofing; and preparing and delivering informative products, such as summary tables, maps, charts, and reports.	Project Crew Leader Project Crew Member Project Data Technician Network/Asst. Data Manager
Analysis	Using data to predict, qualify, and quantify ecosystem elements, structure, and function as part of the effort to understand these components and address monitoring objectives.	Network Ecologist Park Resource Specialist Quantitative Ecologist
Management	Preparing and executing policies, procedures, and activities that keep data and information resources organized, available, useful, compliant, and safe.	Network Data Manager Project Manager Assistant Data Manager National I&M Data Manager
End Use	Obtaining and applying available information to develop knowledge that contributes to understanding and managing park resources.	Network Coordinator Park Manager Superintendent Other (Elected officials, etc)

**NOTE – Each position may be listed in only one category according to primary responsibilities; however, most positions contribute to activities in each category.*

2.1 Primary Roles and Responsibilities for Data Management

The network considers a role as a function or position (*e.g., Data Manager*) while a responsibility is a duty or obligation (*e.g., review data records*). Data stewardship is not treated as a single role because it involves all project staff sharing the responsibilities for managing data and information resources. Establishing a single project data steward role

is avoided because an unmanageable burden (real or perceived) is placed on a single position/person, and others with data management responsibilities are more likely to underperform. An increasing demand for more detailed, higher quality data and information about natural resources and ecosystem functions requires a group of people working together to effectively steward data and information assets. The current and expected capacity, diversity, and rate of change in information technology make managing any large amount of data a greater task than any person can reasonably do alone. Knowledgeable individuals from scientific and technological disciplines must work together to ensure that data are collected using appropriate methods, and that resulting data sets, reports, maps, models, and other derived products are well managed, credible, representative, and available for current and future needs. A typical data set used for inventory and monitoring is handled, viewed, and stewarded by most or all of the people serving in the roles (jobs) listed in Table 2.2. These roles are listed ‘from the ground up’ to help demonstrate the hierarchy and overlap of responsibilities.

According to national I&M guidance, one third of network resources are to be devoted to overall data management, analysis, and reporting. The network intends to avoid cases in which a single position is assigned unattainable duties for multiple roles, resulting in a lack of attention to one or more aspects of the project. Meeting this goal requires awareness and support from park managers to plan for, fund, and provide performance level accountability for the resources required to adequately manage their park’s data and information assets. Learning, understanding, and acting on the responsibilities listed for all categories of data stewardship will help prevent the critical aspects of data management from being overlooked or under-funded. Vital Sign monitoring protocols and SOP documents will describe more detailed data stewardship roles and responsibilities.

Table 2.2 Programmatic roles and associated data stewardship responsibilities

Role	Data Stewardship Responsibilities
Project Crew Member	Record and verify measurements and observations based on project objectives and protocols. Document methods and procedures.
Project Crew Leader	Supervise crew members to ensure adherence to data collection and data processing protocols, including data verification and documentation.
Assistant Data Manager	Provide support for database development and customized programming. Work with Network Data Manager in implementing data management procedures. Duties may include designing and implementing databases, cataloging, data conversion and integration, documentation, and reporting.
Quantitative Ecologist	Develop project sampling design. Analyze data, develop models, and documentation procedures. Review methods for statistical soundness.
Project Manager	Direct project operations. Communicate data management requirements and protocols to project staff, Network Data Manager, and resource specialist(s). Responsible for final submission and review of all products and deliverables.
Park Research Coordinator	Facilitate data acquisition by external researchers. Communicate NPS requirements to permit holders.
Park Resource Specialist	Understand project objectives, data, and management relevance. Make decisions about validity, sensitivity, and availability of data.
Curator (Park or Region)	Manage collection, documentation, and preservation of specimens.
Network Data Manager	Oversee development, implementation, and maintenance of data infrastructure and standards. Perform GIS analysis and mapping. Facilitate and integrate data and metadata. Oversee long-term data storage and maintenance. Design and develop databases and

Role	Data Stewardship Responsibilities
	applications and implement secure file server backup system.
Network Ecologist	Ensure useful data are collected and managed by integrating natural resource science into Network activities and products, including specifying objectives, sample design, data analysis, synthesis, and reporting.
Network Coordinator	Ensure programmatic data and information management requirements are met as part of overall Network business.
Information Technology Specialist (Network or Region)	Maintain LAN, establish and maintain overall system security, and update enterprise software and hardware.
I&M Data Manager (National)	Provide Service-wide database design, support, and services, including processing to convert, store, and archive data in Service-wide databases.
End Users (managers, scientists, interpreters, public)	Provide feedback on scientific information, presentation needs, and interpretation. Use information for management decisions.

2.2 Project Management

Project managers will direct data collection for Vital Signs monitoring and/or inventory efforts. They will provide project oversight, direct on-the-ground data collections, and provide a cohesive link between data collection, synthesis, interpretation, and reporting. Project managers will work with project and NGPN personnel to ensure that project data are handled properly. Project managers and the Network Data Manager share a joint responsibility to collaboratively design the data management procedures for Vital Signs monitoring.

The Importance of Documentation

Among all of the data management responsibilities shared by project and Network personnel, the careful documentation of data sets, data source(s), and data collection methodology is paramount. This careful and thorough documentation establishes the basis for the appropriate use of the data in resulting analysis and reporting, both in the short-term and long-term. Network monitoring protocols contain key elements of data documentation. Network data records collected according to these protocols will include the name, date, and version of the associated protocol. (Chapter 7 presents important guidance and reference for documentation and metadata.)

2.3 Data Stewardship

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.

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 regular 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.

2.4 The Hub of Data Stewardship

Project leaders and the Network and Assistant Data Managers comprise the central data management team for inventory and monitoring projects. Because of the collaborative nature of project data management, communication among this team is essential to meeting program goals. The following section outlines the individual and shared responsibilities of each role.

Project Leaders

The project leader is accountable for data quality during all phases of the project, including collecting, entering, handling, reviewing, summarizing, and reporting data. Developing project documentation and metadata are crucial elements of the project leader's role. Specifically, the role of the project leader includes the following.

- Complete project documentation describing the who, what, where, when, why and how of a project.
- Develop, document and implement standard procedures for field data collection and data handling.
- Enact and supervise quality assurance and quality control measures for the project. Supervise and certify all field operations, including staff training, equipment calibration, species identification, and data collection. Supervise or perform data entry, verification and validation.
- Maintain concise explanatory documentation of all deviations from standard procedures.
- Ensure documentation of important details for each field data collection period.
- Maintain hard copies of data forms and send original data forms to archive on a regular basis.
- Work with program coordinators to identify analysis and reporting mechanisms and to establish a schedule for regular project milestones such as data collection periods, data processing target dates, and reporting deadlines.
- Produce regular summary reports and conduct periodic trend analysis of data, store the resulting reports, and make them available to users.
- Act as the main point of contact concerning data content.

The project leader will also work closely with data management staff to:

- Develop quality assurance and quality control procedures specific to project operations.
- Identify training needs for staff related to data management philosophy, database software use, quality control procedures, etc.
- Coordinate changes to the field data forms and the user interface for the project database.
- Fully document and maintain master data.
- Identify sensitive information that requires special consideration prior to distribution.

- Manage the archival process to ensure regular archival of project documentation, original field data, databases, reports and summaries, and other products from the project.
- Define how project data will be transformed from raw data into meaningful information and create data summary procedures to automate and standardize this process.
- Identify and prioritize legacy data for conversion; convert priority data sets to a modern format.
- Increase the interpretability and accessibility of existing natural resource information.

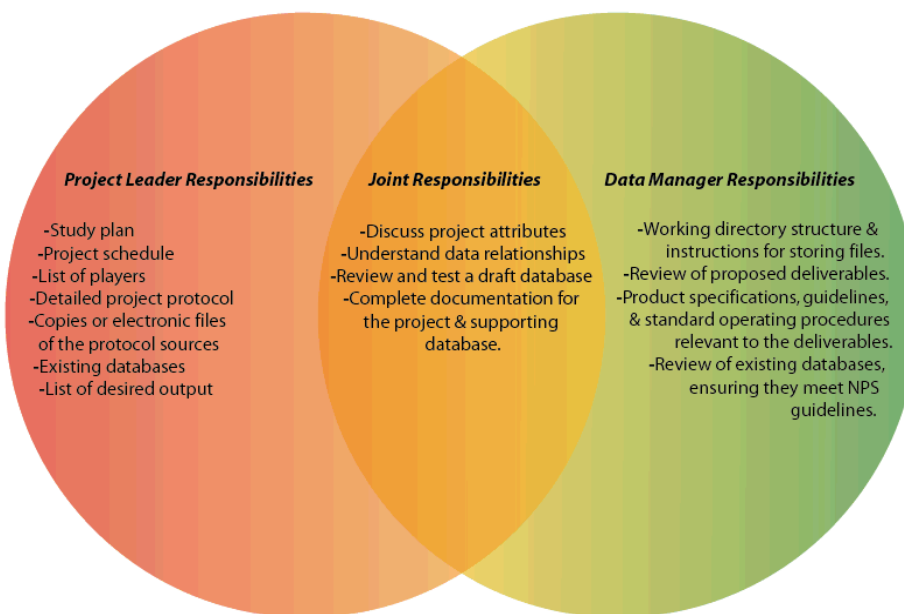


Figure 2.1 Core data management responsibilities for project and data managers

Network Data Manager and Assistant Data Manager

The NGPN has two positions that are assigned data management as their principal task: the Network Data Manager and the Assistant Data Manager. These two positions work together to ensure the compatibility of project data with program standards, design the infrastructure for the project data, and ensure long-term data integrity, security, and availability. The Network data manager coordinates with the project leader throughout the project, as shown in Figure 2.1.

The Network Data Manager will:

- Coordinate and manage the Network data management program, including spatial and tabular data.
- Develop and maintain the infrastructure for metadata creation, project documentation, and project data management.

- Create and maintain project databases in accordance with best practices and current program standards.
- Provide training in the theory and practice of data management tailored to the needs of project personnel.
- Develop ways to improve the accessibility and transparency of digital data.
- Establish and implement procedures to protect sensitive data according to project needs.
- Integrate tabular data with geospatial data in a GIS system in a manner that meets project objectives.

The Assistant Data Manager will:

- Design and enhance project databases in accordance with best practices and current program standards.
- Create and enhance customized interfaces for project databases, including data entry forms for a variety of platforms.
- Program database query and reporting functions to enable program staff to more easily access Network data.
- Create user interfaces to streamline the data entry process and simplify the review and validation of project databases.
- Work with technicians and project team members to ensure correct data entry procedures are being utilized in transcribing field data forms into databases.

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

- Define the scope of the project data and create a data structure that meets project needs.
- Become familiar with how the data are collected, handled and used.
- Review quality control and quality assurance aspects of project protocols and standard procedure documentation.
- Identify elements that can be built into the database structure to facilitate quality control, such as required fields, range limits, pick lists and conditional validation rules.
- Create a user interface that streamlines the process of data entry, review, validation, and summarization that is consistent with the capabilities of the project staff.
- Develop automated database procedures to improve the efficiency of the data summarization and reporting process.
- Make sure that project documentation is complete, complies with metadata requirements, and enhances the interpretability and longevity of the project data.
- Ensure regular archival of project materials.
- Inform project staff of changes and advances in data management practices.

The NGPN data management staff will also manage geospatial data associated with network inventory and monitoring projects, as well as other spatial data related to the full range of park resources. They also maintain standards for geographic data and are responsible for sharing and disseminating GIS data throughout the network.

The Network Data Manager will work in collaboration with project leaders to:

- Determine the GIS data, imagery, and analysis needs for the project.
- Acquire and convert GIS and imagery data necessary for the project.
- Develop procedures for field collection of spatial data including the use of GPS and other spatial data collection techniques.
- Display, analyze, and create maps from spatial data to meet project objectives.
- Properly document data in compliance with spatial metadata standards.
- Design geospatial databases and other applications for the network.
- Create relationships between GIS and non-spatial data and create database and GIS applications to facilitate the integration and analysis of both spatial and non-spatial data.
- Establish and implement procedures to protect sensitive spatial data according to project needs.
- Develop and maintain an infrastructure for metadata creation and maintenance.
- Ensure that project metadata are created and comply with national and agency standards.

2.5 Data Management Coordination

The Natural Resource Challenge states that collaboration among the National Park Service, other public agencies, universities, and non-governmental organizations is necessary to effectively acquire, apply, and promulgate the scientific knowledge gained in parks. The I&M Program encourages coordination to help ensure that data collected by NPS staff, cooperators, researchers and others are entered, quality-checked, documented, analyzed, reported, archived, cataloged, and made available for management decision-making, research, and education. The Network Data Manager and Assistant Data Manager both work with the national I&M data management staff and regional information management personnel to maintain awareness and involvement in service-wide and regional databases and data management policy and guidance. Staff at the Washington office in Fort Collins provides critical assistance necessary for NGPN to accomplish data management tasks.

Network data management staff work locally with other network personnel, park staff, and cooperators to ensure integration and availability of data sets. Regular and productive communication among these personnel leads to common understanding and better synchronization of network and park data management activities. Such communication uses a variety of methods, including personal visits, phone calls, email, joint meetings and training sessions, as well as the meetings and work of the network's Technical Committee and Board of Directors. It is also anticipated that the Network will utilize SharePoint for future collaboration, allowing for cooperation without regard to geographic location. The development of network planning materials, including inventory study plans, monitoring protocols, and SOPs includes involvement and input from park scientists and information management staff.

Data managers throughout the NPS regularly collaborate with each other and national program staff via annual meetings, conference calls, workgroups, a data management

listserver, web sites, and informal communication. This communication and cooperation promotes compatibility of protocols and data sets among networks and organizations.

3. DATA MANAGEMENT PROCESS AND WORK FLOW

The network is involved with short-term and long-term projects that generate data sets, reports, and other products. Short-term projects usually last from one to three years and include individual research projects, inventories, or pilot work done in preparation for long-term monitoring. Long-term projects include Vital Sign monitoring and other multi-year research and monitoring projects performed by other agencies and cooperators. Products produced by all of these projects must be managed and made available to current and future users. Data stewardship begins as soon as the project starts. Successful project management requires clearly defined data stewardship responsibilities communicated to project staff at all stages of a project.

3.1 Project Work Flow

At an operational level, the process used to implement any data-gathering project includes clearly defined steps for effective data management. A generalized project work flow model provides the framework for data management. While the work flow presented in this plan may not apply to all situations, it does address both the long- and short-term data collection efforts anticipated by the NGPN monitoring program. This section outlines five stages of a project shown (Figure 3.1) and describes the roles and responsibilities for project and data management.

Planning and approval

The project leader and program administrator are responsible for defining the scope and objectives of the project, establishing funding sources, and obtaining permits and compliance clearances. The data management staff contribute as data processing and reporting schedules are set and budgets for data management activities are discussed and incorporated.

Design and testing

The project leader is responsible for adopting or developing and testing project methodology to meet clear objectives. The project leader and data management staff work together to establish appropriate data management procedures, develop the data design, and prepare a data dictionary which comprehensively lists and describes the database fields and data value domains. The data managers develop a new database application or adapt an existing database to meet project requirements. Geospatial data needs are identified and data is acquired or flagged for development by the Network Data Manager to support the project objectives.

Implementation

During implementation data are acquired, processed, error-checked and documented. Reports, maps, geospatial data, and other products are developed and distributed. The project leader oversees all aspects of implementation, including logistics, contracting, training, and equipment procurement, data acquisition, report preparation and final delivery of products to resource managers. Data management staff provide training and support for database applications, GIS, GPS, and other data processing applications; facilitation of data summarization, validation and analysis; and assistance with the technical aspects of documentation and product development.

Product integration

The network maintains master copies and archives of all network-generated data on a local server. These are used to update NPS databases at appropriate intervals, e.g. annual updates to

NPSSTORET for water quality data. Chapter 4 lists and describes the Service-wide databases used by the network for archival and distribution of data and information resources. The network also works with other partners to share and integrate results from network projects.

Evaluation and seasonal or final project wrap-up

For long-term monitoring and other cyclic projects, this phase occurs at the end of each field season as part of an annual review of the project. For non-cyclic projects, this phase represents the completion of the project. Upon season or project closure, project records are updated to reflect the status of the project and its associated deliverables. After products are catalogued and made available, program administrators, project leaders, and the data management staff work together to assess how well the project met its objectives. They determine what might be done to improve various aspects of the methodology, implementation, and formats of the resulting information. For Vital Signs monitoring protocols, careful documentation of all changes is required. Changes to methods and procedures are maintained in a tracking table associated with each document. Significant revisions, as defined in each protocol, could potentially require additional peer review.

3.2 Core Data Management Elements of the Project Work Flow

Essential data and information management includes planning for and implementing data collection, data entry, validation (quality control, etc.), and submittal to designated archives and audiences. Project activities involving these elements are highlighted in Figure 3.1. As indicated in Figure 3.1, metadata should develop as the project progresses rather than after data are collected, entered, validated, verified, and submitted (Chapter 7). Similarly, quality control of data should begin with the data design. Field sheets and databases should be constructed to minimize and restrict the variety of values that may be recorded for a given parameter. Pick lists, whether built into a data logger or included on a field sheet, should be employed wherever possible to help narrow the range of inputs to acceptable values.

Certain key information is common both to data sets and to the organization as a whole. For example, lists of contacts, projects, parks, and species are relevant to many projects, but are often complex and dynamic. It is a good strategy to centralize such information so that users have easy access to the most updated versions in a single place. Centralization also avoids redundancy and versioning issues among copies. Centralized information is maintained in database tables that can be linked or referred to from several distinct project databases. Network applications – for project management, administrative reporting, or budget management – can also link to the same tables so that all users have instantaneous access to edits made by other users.

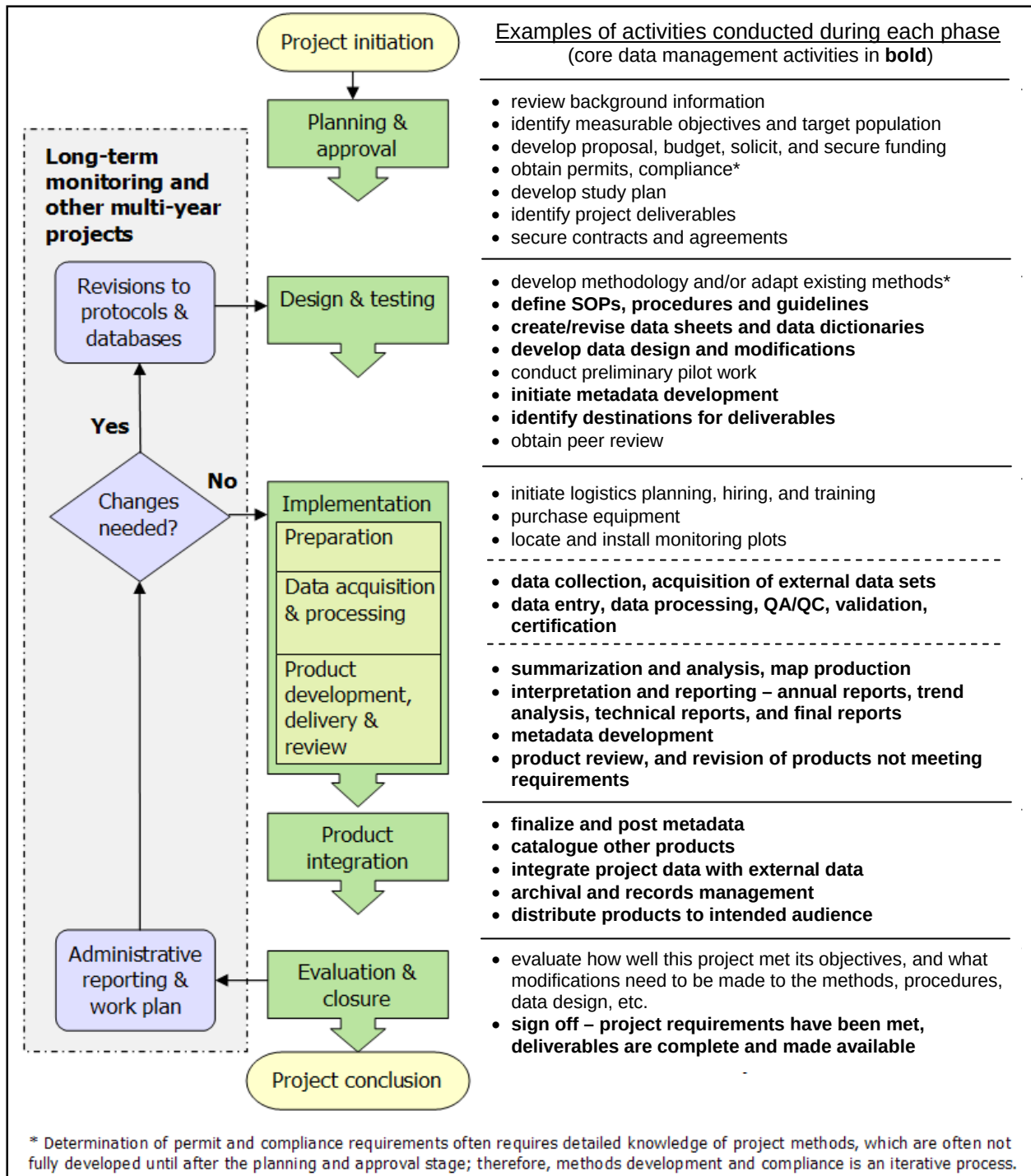


Figure 3.1 Generalized project work flow model for short- and long-term data collection efforts (Bold text indicate core data management elements)

3.3 Data Life Cycle

Project data take on multiple forms and are maintained in different places as they are acquired, processed, documented, and archived (Figure 3.2). Types of databases and forms of data include:

- Raw data* – Analog data recorded by hand on hard-copy forms and digital files from handheld computers, GPS receivers, telemetry data loggers, etc.

- *Working database* – A project-specific database for entering and processing data for the current season (or other logical period of time).
- *Certified data and metadata* – Completed data and documentation for short-term projects, or one season of completed data for long-term monitoring projects. Certification is a confirmation by the project leader that the data have passed all quality assurance requirements and are complete and ready for integration and distribution. Metadata records include the detailed information about project data needed for its proper use and interpretation (see Chapter 7).
- *Master database* – Project-specific database for storing the full project data set, used for viewing, summarizing, and analysis. Only used to store data that have passed all quality assurance steps.
- *Reports and data products* – Information that is derived from certified project data.
- *Edit log* – A means of tracking changes to certified data.
- *National databases and repositories* – Applications and repositories maintained at the national level, for the purpose of centralization, archival, and distribution among NPS units and with cooperators and the public.
- *Local archives and digital library* – Locally stored data, metadata and other products generated by projects.

Steps in the typical data process flow (Figure 3.2):

1. *Acquire data* – For data recorded by hand in the field, data forms should be reviewed regularly (e.g., daily or weekly) for completeness and clarity in order to identify and resolve errors as close to their origin as possible.
2. *Archive raw data* – Copies of all raw data files are initially stored intact. Digital files are copied to the digital library section for the project; hard copy forms are either scanned and placed in the digital library, or are copied and placed in the archives. Archival or scanning of hard copy data forms may occur at the end of a season as a means of retaining all marks and edits made during the verification and validation steps. Once a project is completed or annual data results are validated and reported, original and intermediate data sets may be disposed in order to avoid unnecessary clutter in the information system. Disposing of data assets must be approved by at least the project leader and the data manager based on established criteria for disposition that is documented in the data management procedures for the project.
3. *Data entry/import* – Analog data are entered manually, and digital data files are uploaded to the working database.
4. *Verification, processing and validation* – Verify accurate transcription of raw data and validate data using database queries and exploratory data analysis (Tukey 1977) to identify and resolve missing data values, out-of-range values, and logical errors.
5. *Documentation and certification* – Develop or update project metadata and certify the data set. Certification is a confirmation by the project leader that the data have passed all

quality assurance requirements and are complete and documented. It also means that data and metadata are ready to be posted and delivered externally or used locally.

6. *Archive versioned data set* – Copies of the certified data and metadata are placed in the digital library. This can be accomplished by storing a compressed copy of the working database, or by exporting data to a more software-independent format (e.g., ASCII text).
7. *Upload data* – Certified data are uploaded from the development database to the production database.
8. *Post data and update national databases* – To make data available to others, certified data and metadata are posted to national repositories such as the NPS Data Store. In addition, national databases such as NPSpecies and NPSTORET are updated. Data and data products containing protected information are posted according to the guidelines and access levels outlines in chapter 9.
9. *Reporting and analysis* – Certified data are used to generate data products, analysis, and reports, including semi-automated annual summary reports for monitoring projects. Depending on project needs, data can be exported for analysis using other software applications and/or summarized within the database.
10. *Store products* – Reports and other data products are stored according to format and expected demand – either in the digital library, on off-line media, or in the document archives.
11. *Distribute data and information* – Deliverables can be shared and distributed in a variety of ways - via the web-based national databases and repositories, by FTP or mailing in response to specific requests, or by providing direct access to project records. Distribution is subject to legal requirements to protect information about sensitive resources (see Chapter 9).
12. *Track changes* – All subsequent changes to certified data are documented in an edit log, which accompanies project data and metadata upon distribution. Significant edits will trigger reposting of updated data and products to national databases and repositories.

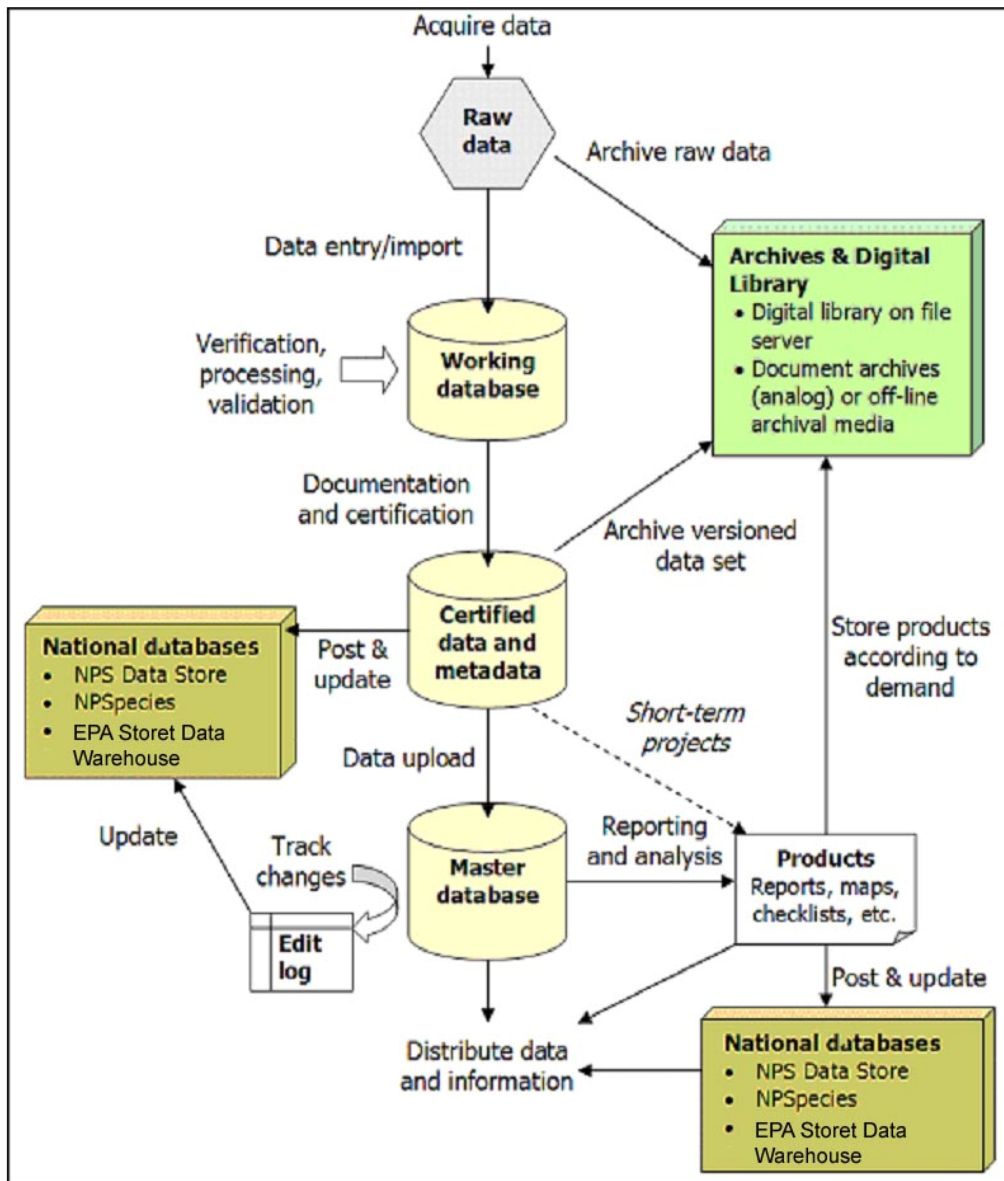


Figure 3.2 Diagram of the typical project data life cycle

This data flow occurs iteratively (usually annually) for long-term monitoring projects and once for short-term projects. For projects spanning multiple years, decision points include whether or not a separate working database is desirable, and the extent to which product development and delivery is repeated year after year.

3.4 Integrating and Sharing Data Products

Certified project data and products need to be secured in long-term storage and made available. This requires using a variety of information systems such as product repositories, clearinghouses, and web applications. Network data will be archived at both the Network level, on the NGPN file server, and at the national level, in the NPS Data Store, in NPStoret, NPSpecies, and other national databases.

The process of product distribution occurs in several ways. In most cases the data manager reviews products for conformance with format standards then transmits the products to the appropriate repository. All products are reviewed for completeness and accuracy by the appropriate resource specialist prior to delivery to the data manager. Stored products are documented by posting to the NPS Data Store. The final and certified data will also be available on the NGPN website. Chapter 9 will discuss data distribution in more detail.

4. DATA MANAGEMENT INFRASTRUCTURE

The ultimate measure of success for the NGPN I&M information management program is to supply quality information in a timely manner to a wide-range of stakeholders (e.g. NPS personnel, scientists, and the public) with variable end-uses. The most common method for delivering data and information is via data/information clearinghouses. Clearinghouses facilitate one-stop shopping through the internet and provide a mechanism for searching, browsing and downloading or requesting information at the convenience of the user. Typically the data or information products provided directly through clearinghouses are “snapshots” that do not change. These static information products and data sets will be stored in national level systems already developed for such purposes or on the NGPN website. Through the use of metadata, these clearinghouses also facilitate data discovery by external users who may then request access to the data from its originator.

Our computer resource infrastructure is composed of computers and servers that are functionally or directly linked through computer networking services. This infrastructure represents the foundation upon which our network information system is built. Systems architecture signifies the applications, database systems, repositories, and software tools that make up the framework of our data management enterprise.

The NGPN monitoring program relies heavily on park, regional, and national information technology (IT) personnel and resources to maintain the computer resource infrastructure. Ongoing maintenance of computer resources includes, but is not limited to, hardware replacement, software installation and support, security updates, virus-protection, and telecommunications networking. Communication with park, regional and Network IT specialists is essential to ensure service continuity for the Network’s systems architecture. Rather than focusing on a detailed description of the current computer resources, this chapter will focus on the systems architecture that is central to data management.

At this time, the NGPN data management system is in a developmental stage. This chapter also specifies current plans and standards for NGPN data management operations. In general, the NGPN will conform to National Park Service standards and policies in all aspects of Program data management operations. Further, the NGPN will conform to national I&M Program standards and mandates in the interest of program integration and information sharing.

4.1 Role of Infrastructure in the Data Management System

Chapter 3 presented an overall project-flow model as a framework for data and information management. Core data management practices were presented in the context of the project work flow in Figure 3.1. The major data management elements in the process include:

- Collect or acquire data (field work, satellite download, etc.)
- Archive raw data (physical and possibly digital material)
- Enter or import data
- Verify and validate data (quality assurance and quality control)
- Produce documentation (metadata, etc.)
- Archive validated data & documentation (digital)
- Integrate project data with external data (NPS Data Store, NatureBib, etc.)

- Summarize and analyze data, and produce reports
- Archive data products
- Catalogue data products

The elements listed above comprise the data life cycle for most NGPN projects. Figure 3.1 presents the relative timing of each element during a project. Procedural details will be documented in the standard operating procedures for each Vital Sign. Common data management procedures are included in this plan.

4.2 NPS Infrastructure – NGPN Components

Because NGPN offices are separate from a park or other NPS office with conventional NPS computer network service, NGPN maintains a small client-server computer architecture that relies on disconnected data transfer to exchange information with the parks. The Northern Great Plains Network digital infrastructure consists of an in-house server connected to workstations through the Regional Office using Microsoft's Active Directory architecture. Databases, documents, and geospatial data are stored on the server, which has a redundant drive array. Network staff also maintains copies of data on their personal workstations.

The Network maintains a consistent LAN/WAN connection in order to access data and information stored on remote computers. Distributed data are accessed through NPS servers, which host applications such as NatureBib, NPSpecies, and the NPS Data Store, and through sites maintained by partner agencies (e.g., the National Weather Service, USGS or EPA).

There are distinct advantages to this distributed data architecture. Data are maintained and archived by the relevant data stewards and the data provided is consistent in quality and format. The NGPN staff utilizes the Network LAN for data exchange and common storage. The configuration and components in the LAN will change as equipment is replaced and as new technology becomes available. It is also probable that Network staff will access LAN resources over the Service-wide WAN in the future. At the time this document was developed, the NGPN server is a Dell PowerEdge 2900 III Server with a RAID 5 disk array totaling 8 Tb of disk space. Server access is restricted to the Network Data Manager, the Assistant Data Manager, the Network Coordinator, and the regional IT specialist.

The following list outlines the data components that encompass the distributed data within the NGPN I&M Program.

Network data server

The network server components are:

- Master project databases – compiled data sets for monitoring projects and other multi-year efforts that have been certified for data quality
- Common lookup tables – e.g., parks, projects, personnel, species
- Network digital library – Network repository for finished versions of project deliverables for Network projects (e.g., reports, methods documentation, data files, metadata, etc.)
- Working files – working databases, draft geospatial themes, drafts of reports, administrative records
- GIS files – base spatial data, imagery, and project-specific themes

Network computers

The network computer components are:

- Draft project databases – databases in development
- Working files – Draft reports, administrative records
- Draft GIS files – Projects in development
- Local applications – desktop versions of national applications such as NPSpecies

Network software

The NGPN maintains current versions of all enterprise NPS applications, including Microsoft Office and ESRI's ArcGIS products. The Network Data Manager, Assistant Data Manager, and regional IT staff work collaboratively to ensure that applications and operating system software are patched and maintained at the current level.

The components of the information system architecture will change as the I&M Program matures. Additions to existing regional and Network-level infrastructure will be made in order to meet data management objectives.

Midwest Regional Office

Two regional programs contribute to the Network's data management infrastructure. The Midwest Regional Office (MWR) provides software updates, some hardware upgrades, and technical support related to security and networking issues. The Midwest Regional GIS Office distributes software updates for ESRI's ArcGIS products and helps develop and complete GIS projects for the Network and MWR parks.

Chapters 9 and 10 of this plan present more details on the distribution, maintenance, and storage of NGPN data. These chapters may be summarized as:

- The NGPN will manage a primary repository located in the Network office for data and information generated by the Network.
- A portion of the NGPN databases required for the efficient operation of the Network's IMS gateway may be mirrored on WASO servers, and access restricted to authorized users.
- The primary NGPN repository is backed up with duplicate media (stored on separate physical machines and external hard drives). The Dell server also has a 1.6 Tb compressed tape drive for backups, which are run daily, with selected tapes being stored offsite to ensure redundancy in the event of a fire or natural disaster.
- Required portions of the NGPN data will be stored in national I&M Program databases such as NPSpecies, NatureBib, and NPS Data Store.
- Certain NGPN data sets will be primarily maintained by outside organizations; however, metadata for these data sets will be maintained in the primary NGPN repository. An example of this category is the climate data which will be handled by National Oceanic and Atmospheric Administration (NOAA) or a designee, such as the Midwest Regional Climate Center, under formal agreement with the NPS and/or the NGPN.

4.3 NPS Infrastructure – National-level Application Architecture

To achieve an integrated information management system, two of the national-level data management applications (NatureBib and NPSpecies) utilize a distributed application

architecture with both desktop and internet-accessible (master) components (Figure 4.2). Under development at the national level is a single portal or interface that will combine access to NPSpecies, NatureBib, and the NPS Data Store (initially). This project is the Integration of Resource Management Applications, or IRMA.

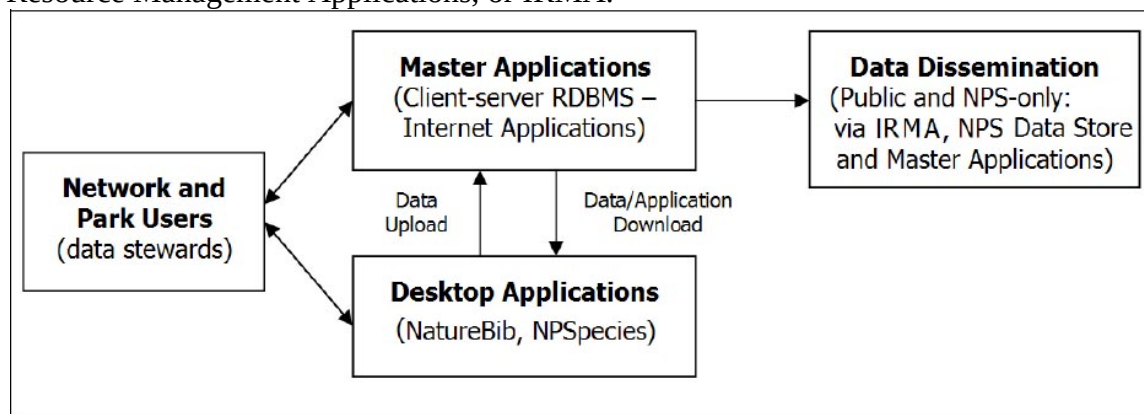


Figure 4.2. Model of the national-level application architecture

NatureBib

NatureBib is the master database for bibliographic references that merges a number of previously separate databases such as Whitetail Deer Management Bibliography (DeerBib), Geologic Resource Bibliography (GRBib), and others. It also contains citation data from independent databases like NPSpecies and the Dataset Catalog and NR-GIS Metadata Database. It currently focuses on natural resource references, but may eventually be linked to references on cultural resources and other park operations. See the following web link for more information on NatureBib: <http://www.nature.nps.gov/nrbib>

The online version of NatureBib can be accessed at <https://science1.nature.nps.gov/naturebib/nb/simple/clean>

NPSpecies

NPSpecies is the master species database for the NPS. The database lists the species that occur in or near each park, and the physical or written evidence for the occurrence of the species (e.g., references, vouchers, and observations). Taxonomy and nomenclature are based on ITIS, the interagency Integrated Taxonomic Information System. The master version of NPSpecies for each Park or Network can be downloaded from the master website into an MS Access version of NPSpecies. The internet-based version is the master database, which can be accessed via password-protected logins administered by park, Network and regional data stewards. The master database requires that species lists are certified by Networks before any data will be available to the public. NPSpecies is linked to NatureBib for bibliographic references that provide written evidence of a species' occurrence in a park. NPSpecies data can be accessed without the need for a login through the Natural Resource Information Portal (developed as part of the IRMA project) at: <http://nrinfo/Portal/Home.mvc>. The MS Access application and additional details can be found at the NPSpecies website: <http://science.nature.nps.gov/im/apps/npspp/index.htm>

NPS Data Store

The NPS Data Store is a key component of the data dissemination strategy employed by the I&M Program. The NPS Data Store utilizes a graphical search interface that links dataset metadata to a searchable data server on which datasets are organized by NPS units, offices and programs. The interface allows customized public or protected searches of natural, cultural, facilities and fire resource geospatial data, tabular data, map products, data standards and other products produced by the I&M and other NPS programs. Each park or Network is able to post and curate its own data on the server. See the NPS Data Store website for further information:

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

NPSTORET

STORET is an interagency water quality database developed and supported by the Environmental Protection Agency (EPA) to house local, state, and federal water quality data collected in support of managing the nation's water resources under the Clean Water Act. STORET is used by NPS as a repository of physical, chemical, biological, and other monitoring data collected in and around national park units by park staff, contractors, and cooperators. The NPS operates its own service-wide copy of STORET and makes periodic uploads to the EPA STORET National Data Warehouse so that data collected by and for parks will be accessible to the public. NPS Director's Order 77 indicates that the NPS should archive water quality data in STORET, and the NPS Water Resources Division (WRD) requires that any data collected as part of a WRD funded project get archived in STORET. NPSTORET is the NPS master database designed to facilitate park-level standardized reporting for STORET.

More details on NPSTORET can be found at:

<http://www.nature.nps.gov/water/infoanddata/index.cfm>

Additional information on STORET can be found at:

<http://www.epa.gov/storet>.

Natural Resource Database Template

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

Monitoring protocols, including databases that are based on the Natural Resource Database Template, are made available through a web-based protocol clearinghouse (see below). A description of the NRDT, a data dictionary, and example implementations are located on the NR Database Template website:

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

Natural Resource Monitoring Protocols Clearinghouse

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

4.4 Database Design

Rather than developing a single, integrated database system, our design relies upon modular, standalone project databases that share design standards and links to centralized data tables. Individual project databases are developed, maintained, and archived separately. There are numerous advantages to this strategy:

- Data sets are modular, allowing greater flexibility in accommodating the needs of each project area. Individual project databases and protocols can be developed at different rates without a significant cost to data integration. In addition, one project database can be modified without affecting the functionality of other project databases.
- By working up from modular data sets, we avoid a large initial investment in a centralized database and the concomitant difficulties of integrating among project areas with very different – and often unforeseen – structural requirements. Furthermore, it is uncertain whether such an initial investment would ever payoff in terms of increased efficiency of interdisciplinary use.

The I&M Program recognized early on that it needed database standards and detailed documentation. Database standards promote compatibility among data sets and allow data to be aggregated and summarized. Well thought-out standards help to encourage sound database design and facilitate interpretability of data sets.

The I&M Program has developed a series of recommendations in database design which include:

- Database specifications for I&M Studies
- Recommended database strategies, including the Natural Resource Database Template (NRDT)
- Recommended naming standards
- Natural Resource Database Template data standards

Each database must ultimately meet the needs of the Network scientists. These needs may include compatibility with databases produced by other agencies and ease of use, maintenance, integration, and customization. The NGPN will model and document databases as described in this Data Management Plan, will use the above recommendations in database design as a guideline where feasible, and will standardize attributes across databases where feasible. Each Vital Sign will have an associated database to store monitoring data.

Fundamental Database Structure

The Natural Resource Database Template (NRDT) provides an example of the fundamentals for database structure to be used by the I&M Program. These fundamentals are as follows:

1. Mandatory (core) tables and fields – These tables and fields are used to manage the information describing the “who, where, and when” of project data. These tables contain critical data fields that are standardized with regard to data types, field names, and domain ranges. For example: tbl_Sites, tbl_Locations, tbl_Events.
2. Mandatory if Applicable (master look-up elements) tables and fields – Described in more detail below (Integrated Data Dictionary), these tables contain commonly used attributes, such as park codes, and are centrally located to minimize redundant information in each database.
3. Optional (project specific) tables and fields – These are tables and fields that are needed for a specific project or protocol, but will not likely be used in other databases or in integration.

NGPN has adopted the Natural Resources Database Template (NRDT) as the standard for database design and recommends this method be used as much as possible. The NRDT has been extensively reviewed by I&M Program data managers and has been applied to some of biological inventory projects in the Program. It has not, however, been applied extensively to monitoring. Although the fundamentals of this database structure will be applied to NGPN databases, its applicability towards Vital Signs monitoring may require specific mandatory tables to be reconsidered depending on the sampling methods, information collected, and potential for integration with the NGPN geodatabase GIS structures.

Collaborative Database Development

The development of the Vegetation Community protocol has been accomplished jointly with the USGS Northern Prairie Wildlife Research Center and the Northern Great Plains Fire Effects group. The Network intends to develop future databases utilizing the NRDT template to the greatest extent possible. However, other considerations such as working jointly with the Fire Effects program offer significant benefits in terms of collaborative effort and shared workload. The NGPN has adopted the joint USFS/NPS FFI (FEAT and Firemon Integrated) database for use with the Plant Community protocol. Standardized methods within FFI will be utilized to the extent possible, and additional methods will be constructed with the FFI Protocol Manager application.

4.5 GIS Data Structure

The coordinate system standard the National Park Service is *Universal Transverse Mercator with North American Datum 1983* (UTM NAD83). Currently, Network staff and collaborators use many different raster and vector data formats for storing and managing spatial data. Vector data formats commonly include ESRI ArcInfo coverages, shapefiles, and personal geodatabases. Raster data formats commonly include MRSID, TIFF, GEOTIFF, and Grid structures. Since ESRI software products have historically demonstrated excellent backward compatibility with older data structures, all the above listed formats will be accepted for existing GIS datasets. The new ESRI file geodatabase data structure does carry numerous advantages such as scalability, ability to handle large datasets and compatibility across operating systems. Future monitoring GIS datasets will be constructed as file geodatabases to the greatest extent possible.

5. DATA ACQUISITION AND PROCESSING

The NGPN collects information from multiple sources and processes it to meet national and Network standards. This chapter describes the steps involved with acquiring data as well as the initial stages of data processing. These steps are integral to strengthening the scientific foundation of the I&M program and providing the park managers with high quality data.

There are three general classifications for the types of data handled by the I&M Program:

- **Programmatic Data** – any data produced from projects that are initiated (funded) by the I&M Program or projects that in some way involve the I&M Program.
- **Non-programmatic NPS Data** – any data produced by the NPS that did not involve the inventory and monitoring program.
- **Non-programmatic External Data** – any data produced by agencies or institutions other than the National Park Service.

This distinction addresses only on data sources, not the importance of the different data types to the I&M Program. The importance or value placed on a data set will be based on the quality or usefulness of data itself and the impact it has on the I&M Program and parks.

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

5.1 Programmatic Data

Data collection initiated by the Inventory and Monitoring Program usually takes place as part of either natural resources inventories or Vital Signs monitoring projects but can also include collaborative efforts between the NGPN I&M Program and other regional programs or institutions. Data for these projects are typically collected by I&M personnel, park staff or by cooperators/contractors.

Natural Resources Inventories

Resource inventories were designed to identify the primary resources of each park and as such they represent an important phase in the management of park resources.

Vital Signs Monitoring Projects

Vital Signs are measurable, early warning signals of changes that could impair the long-term health of natural systems. Early detection of potential problems allows park managers to take steps to restore ecological health of park resources before serious damage can happen. The NGPN I&M Program, park resource managers, and scientific collaborators selected natural resources Vital Signs that were most important to its parks and to the network as a whole (while still taking limited monitoring resources into account).

5.2 NGPN Data Collection Tools

Biological inventories and monitoring projects are the most common examples of field studies conducted by the NGPN I&M Program. The project manager is responsible for ensuring that data collection, data entry, verification, storage, and archiving for all field projects are consistent with NGPN standards. Project managers will make use of NGPN general standard operating procedures (SOPs) and guidance documents as applicable. The data manager will work closely with the principal investigator and Network staff to develop protocol-specific SOPs for the collection, storage, and maintenance of project data. This may range from detailing the proper usage of data entry forms or databases to outlining calibration procedures for automated data loggers.

Some of the tools that can be used to collect field data are listed below. This list serves as a guide for consideration as the Vital Signs protocols are developed. Details on how these tools may apply to protocol-specific SOPs will be contained in the individual protocol documentation.

- **Field Forms** – These are the most common method of recording field data. Formatted, project-specific data sheets are recommended as opposed to a field notebook. Field notebooks are important for entering additional notes and observations. Acid-free paper should be used to prevent fading and subsequent data loss. Some circumstances may warrant the use of paper and writing implements that can withstand moisture, dust, and other extreme environmental conditions (e.g., Rite in the Rain[®] paper). Standardized data sheets that identify the pieces of information to be recorded and forms that reflect the design of the computer data entry interface help ensure that all relevant information is recorded and that subsequent data entry errors are minimized. Data sheets should contain as much basic, preprinted project information as possible and should provide sufficient space to record relevant metadata (e.g., date, collectors, weather conditions). They should clearly specify all required information, using examples where needed, to ensure that the proper data are recorded. Data recorders should adhere to the following guidelines:
 - All information added to the data sheet must be printed and clearly legible.
 - If alterations to the information are necessary, the original information should be crossed out with a single line and the new information written next to the original entry. Information should never be erased and old information should not be overwritten.

Upon return from the field or when field crews have access to photocopiers near field sites, all original data sheets should be copied and checked for legibility and completeness (i.e., no data cut off at the edges). Because lost data sheets may contain data that is irreplaceable or may cost hundreds of dollars to collect again, each protocol will specify aggressive guidelines for backing up field forms as soon as possible during field seasons. The original data sheets should be used for data entry and then they should be stored in a fireproof cabinet. Copies of original data sheets will be archived offsite.

- **Cameras** –Cameras are useful for capturing photo point records of long-term study sites and visual records of field visits. They also serve well for automated data collection by remotely recording visual information using continuously recording webcams or event-triggered cameras.
- **Field Computers** – Data collection and data entry efficiency can often be increased by eliminating paper field forms. Data can be downloaded directly from field computers to office desktop computers, thereby eliminating the data entry step. This results in less chance for error because QA/QC checks can be built into the field database; however, potential limitations and problems to be avoided must be considered.

The use of field computers can be inefficient if copious amounts of notes or comments need to be recorded in the field. Field computers may be vulnerable to heat, dust, moisture, rage, and loss. When handheld computers are used for data entry in the field, the data should be downloaded daily to avoid potential loss of information. No field computer should be used unless it is equipped with a removable flash memory card to store field data in case of a catastrophic failure of the field unit. Batteries should be checked prior to a data collection trip, and they should be charged at the end of every field day. Also, in case the unit becomes inoperable in the field, printed data sheets should always accompany field teams.

- o **Personal Digital Assistants (PDAs)** – The small size and relatively low cost of these devices make them attractive options for entering field data. They are good for small projects but not powerful enough for large, data-intensive field projects. These devices can be ruggedized fairly easily and at a relatively low cost. Most run the Windows Mobile operating system and may require additional processing or programming to transfer the database structure to the field units.
- o **Tablet PCs** have the same properties as most laptops and provide the user with the convenience of a touch screen interface. They are bulkier, more expensive, and harder to ruggedize than PDAs, but are more powerful as well. Tablet PCs are good for field projects that are very data intensive. Because these units run Windows XP (Tablet Edition), the project database can be directly transferred from desktop units to field units without additional programming steps.
- **Automated Data Loggers** are mainly used to collect ambient information such as air or water quality information. In addition to environmental sensors, this category includes auto-triggered event recorders or cameras, automated RFID readers, and telemetry receivers. These devices can be calibrated and programmed to automatically record data and store them for later download to a computer, thereby eliminating the possibility of manual data entry errors. Automated units must be properly calibrated and maintained and field crews must be properly trained. Calibration and maintenance procedures of the units must be documented in SOPs attached to each protocol and followed by field crews using these devices.

- **Permanently deployed devices** – The high initial cost of permanent devices may be offset by reduced labor costs if fewer field visits are required. Data from these devices must be retrieved and batteries must be changed on a regular basis. The intervals for each activity should be defined in the protocol.
- **Portable hand-held devices** – These units are deployed for sampling only during site visits. They are generally less expensive than units that are permanently deployed in the field. Prior to and following field visits, device components should be inventoried to ensure that all necessary equipment is accounted for.
- **GPS Units** – Three types of GPS units are often used during field work in NGPN parks to collect location information.
 - **Recreational Grade GPS Receivers (e.g., Garmin)** are good for collecting general position information and are inexpensive (usually \$150-\$400). They are not suitable for obtaining high- accuracy location information, but are suited for navigating to preexisting marked locations or for recording incidental location information. The display on a Garmin GPS cannot display GIS data or reflect multiple data layers, and the data must be imported into ArcGIS on returning to the office.
 - **Recreational Grade GPS Receivers (e.g., Trimble JUNO) with ArcPad**
 The accuracy of a Trimble JUNO or Magellan MobileMapper GPS is similar to that of a Garmin recreational GPS unit. Pricing is in between the Garmin units and the mapping grade GPS. The principal benefit of utilizing a unit like the Trimble JUNO is that the Windows Mobile operating system can run ArcPad, and GIS base data and aerial photography can be displayed during GPS data collection. This can greatly assist in locating features and in placing the data collection in a larger context. These units also capture GPS data directly to ESRI shapefile format, simplifying data collection and streamlining the workflow as no conversion is needed upon return to the NGPN offices. NGPN intends to utilize this grade of GPS units for most data collection work.
 - **Mapping Grade GPS Receivers (e.g., Trimble)** are good for collecting highly accurate (submeter) location information. They are usually employed with well developed forms or a database application that ties in with a desktop processing program and GIS standards. The increased cost of these units (typically several thousand dollars) may be outweighed by their additional accuracy and simplified data flow capabilities.

5.3 Programmatic Data Entry

All field studies funded by the NGPN I&M Program will have a Microsoft Access database associated with them. The NGPN I&M Program has adopted the Natural Resources Database Template (NRDT) as the foundation for its databases and will provide assistance in database development for any programmatic project. The database template is highly flexible and can be modified and customized for each project to meet the needs and requirements of the researcher. The database incorporates mechanisms

such pick lists and validation rules for quality assurance purposes (*guidelines in development*).

- Field crews/project staff enter all data into the database provided to them. As protocols are developed, each protocol will identify a target schedule for completing data entry following each data collection effort.
- Field crews are required to periodically forward the data manager the project data files. Refer to individual protocols for the requirements regarding forwarding data.
- The data manager maintains the master copy of the database and updates it with data files received from the field crews.
- All data must undergo QA/QC procedures (see Chapter 6 of this document for more specifics relating to data verification and validation).

5.4 Changes to NGPN Data Collection Procedures

Changes to data collection procedures must be based on valid reasons for altering the methodologies. Most issues will be identified while planning, designing, and testing the protocols, and changes will be implemented prior to collection of field data. Unforeseen issues that arise after data collection has begun may require revision of procedures or protocols.

Improvements in technology may also require changes to procedures. Significant changes to protocols must be approved by the project manager, the Network coordinator, and the data manager. The Network Coordinator will determine if proposed changes to protocol procedures require additional peer review before the changes are accepted and implemented.

Changes to protocols and associated data collection procedures may also occur as a result of scheduled Program reviews. During the review, data may be analyzed to determine if the current protocol is meeting stated objectives. If a protocol has not achieved the desired results then changes should be recommended. All changes must be carefully documented within the SOPs and any associated databases. All data gathered by NGPN must note the version and date of the SOP and/or protocol under which it was gathered; field data forms will have fields to capture this information.

5.5 NPS Data Sources

Non-Programmatic NPS Data

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

Park data

Parks in the NGPN also use base funding or receive funding through NRPP (Natural Resources Protection and Preservation) programs to support park-level projects.

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

One impetus for the development of the National I&M program was the historically short shelf life of large numbers of park-specific data sets. Often, useful data has effectively been lost or degraded due to personnel changes or due to inadequate resources and support for robust data management. The NGPN I&M Program will provide expertise to the parks and lead by example to reduce this problem.

Regional and National NPS Programs

The NPS regional and national programs summarized below support all of the parks within the Midwest Region and are good resources for natural resources information.

- **Air** – National-scale programs collect data, maintain databases, assure data quality, and perform the trend analyses relevant to NGPN air quality issues. The NGPN I&M program will rely on the data analyses from these national scale monitoring networks to obtain trends for most attributes of the air quality Vital Sign.
- **EPMT** - Exotic Plant Management Teams (EPMT) collect and maintain data regarding the presence of exotic species in regional parks, and develop and document the methods used to treat these species. This information is stored in the Alien Plant Control and Monitoring Database (APCAM) which is maintained by the EPMT data manager.
(<http://www1.nrintra.nps.gov/BRMD/invasivespecies/exoticplants/index.htm>). The Network will collaborate with the EPMT to integrate data on exotic plant occurrences and treatment actions under the Exotic Plant Early Warning Vital Sign.
- **Fire Program** – Data concerning the occurrence of fires within the NGPN are maintained both at the individual park and regional level. National databases such as Fire-Pro, SACS and the Fire Program Analysis (FPA) package (<http://fpa.nifc.gov>) have been and will be used to maintain information regarding fire incidence and the resources dedicated to fire management. The NPS is also involved in efforts such as the Joint Fire Science Program (<http://www.firescience.gov>) that provide scientific information and support for fuel and fire management programs. Additionally, the NGPN I&M program has been working closely with the Northern Great Plains Fire Effects team to collaborate on vegetation monitoring.
- **GIS** – The NGPN I&M Program is supported by regional GIS specialists to help ensure that regional GIS data are available and accurate. Much of this data is also available through the NPS Data Store.

- **Inventory and Monitoring** – The regional and network coordinators provide support and guidance for I&M related tasks and activities and communicate national program directives.
- **Water** - Almost all of the field data collection by the regional water resources program is done in support of the water resources Vital Signs monitoring projects. The program also synthesizes, analyzes and interprets water resources data collected by parks.
- **Wildlife Management** - The Regional Wildlife Biologist provides management support to the parks in the Midwest Region.

5.6 External Data

External data sources outside of the NPS often provide information relevant to the mission of the I&M program. It should be noted that such sources need not be directly connected to NGPN parks but may instead pertain to methodologies or protocols that could assist NGPN staff with the development of a more productive program. There are a myriad of possible data sources available to provide information pertinent to the NGPN I&M Program, including academia, private organizations and non-profit groups, and state, local, and federal government agencies. Some NPS units in the NGPN collect these data and contribute them to the larger regional or national data sets maintained by the other agencies. The agencies or organizations that compile these data have the expertise to conduct the proper quality control procedures and the capability to function as a repository and clearinghouse for the validated data.

Data sources to be accessed and used by the NGPN in interpreting causes and effects of trends in Vital Signs include:

- Weather/climate data collected by the National Weather Service and other programs in and adjacent to NGPN parks.
- Data on river flow, water quality, and aquatic biota collected by USGS, states, and other entities.
- North American Breeding Bird Survey and Christmas Bird Counts.

Additional data sources will be identified as protocols are developed.

When the data are not kept in-house, there can be two modes of acquisition: downloads from online databases or requests for data on CD or other media. In some cases, portions of external databases may be incorporated into NGPN databases and thereby made more accessible to NGPN staff.

5.7 Data Discovery and Harvesting

Data discovery or data mining is the process of searching for existing data or information that may be useful to the I&M Program mission and that is related to the natural resources of the Network parks. This is a continual process that begins with the collection of background information and data to assist in the development of project methodologies and designs. The process involves reviewing many different sources for varying types of information. Although data mining is an important part of any I&M project, some Vital Signs monitoring projects may depend largely on data collected by other agencies or institutions and harvested through Network data mining endeavors. Many of the data

sources listed below are accessible via the internet, but some require visiting local archives, research or academic institutions, museums, or local parks in order to search reference material.

Bibliographic/Literature

- National NPS databases (e.g., NatureBib)
- Online literature databases
- Park archives through ANCS+

Geographic Data

- National centralized GIS data from NPS Data Store
- Federal Geographic Data Clearinghouse(s)
- Local, state, and federal government offices
- Regional GIS specialists

Biological/Natural Resources Data

- NPSpecies
- Voucher collections (museums, parks, universities)
- Network parks' records and documents
- Local, state, and federal government offices

Relevant information collected during a data discovery process is maintained at the Network either electronically or in hard copy format, depending on how it was collected. Any data collected during data discovery should be accompanied by as much documentation (metadata) as possible. Geographic data sets collected during this process are documented with FGDC-compliant metadata. In some cases, data sets will be documented using the metadata tools described in Chapter 7.

The process of harvesting data from other sources should be standardized as much as possible. Prior to harvesting data, the following questions should be considered:

- Does the data source organization know of the I&M Program and its data needs?
- Does a Memorandum of Understanding (MOU) need to be in place?
- What is the contingency if this data source is no longer available?
- Can downloads or requests for data be consistently exported/imported and scheduled?
- How will the downloaded data be stored and integrated into the I&M Program?
- How should errors be addressed?
- Is the data source organization interested in integrating I&M Program data where appropriate?
- Is the documentation adequate?

For Vital Signs monitored through data harvesting, these questions should be addressed in their data management protocols.

Data harvested from another agency may not exactly match the needs of the NGPN. Boundaries used for summary data often do not coincide with park boundaries, or specific parameters may not be currently collected. In addition, data may not meet

minimum NGPN accuracy standards. Project managers are encouraged to work with these agencies to incorporate NGPN needs and standards as much as possible. Improving existing systems, rather than initiating new ones, can save money and lead to more efficient monitoring.

5.8 Data Compilation, Processing, and Integration

All data, whether collected in the field by the Network or downloaded from another organization, will need to be compiled and processed. This compilation should follow data management guidelines developed for the specific data type. For example, tabular data will need to be entered or imported into a pre-approved Microsoft Access database or file geodatabase. All GIS data obtained from other entities are stored in the proper format, have the correct spatial reference information and have FGDC compliant metadata. Quality assurance and quality control (QA/QC) procedures are more fully described in Chapter 7.

Certain data sets will require more than the basic processing steps illustrated in the examples above. The level of data procession required for external data sets such as those used in the Vital Signs monitoring program depends on the desired output. If more intensive analyses are intended than are provided through these means – such as conducting in depth analyses at a specific location rather than monitoring regional trends – the data may require additional processing steps. In such cases, the specific protocols should provide the necessary data processing requirements.

6. QUALITY ASSURANCE (QA) AND QUALITY CONTROL (QC)

We must have confidence in the data we use. Data of inconsistent or poor quality can result in loss of sensitivity and incorrect interpretations and conclusions, potentially to the detriment of park resources. However, avoiding such problems is a major challenge for a large-scale, multi-disciplinary monitoring program. The potential for problems with data quality increases dramatically with the size and complexity of the data set (Chapal & Edwards 1994). Therefore, the NGPN has developed procedures for QA (planning to obtain the highest possible data quality) and QC (monitoring the system or appraising the product after the product is developed).

The NGPN will establish and document protocols in order to identify and reduce error at all stages in the data lifecycle. These stages include project planning and database design, data collection, data entry, verification and validation (certification), documentation (including data quality and sensitivity review), archiving (Figure 6.1), and dissemination and integration. Detailed QA/QC procedures for these stages will be included in each protocol initiated by the NGPN. This chapter of the data management plan presents more broadly based procedures and policies that govern specific operations within a NGPN project. Figure 6.2 illustrates selected QA/QC procedures relative to the amount of planning and quality control necessary to have confidence in the data. Quality assurance and quality control cannot be considered without specifying certain roles and responsibilities native to such procedures. The duties outlined in this chapter are inherited from those listed in Chapter 2 for all members of the data-stewardship team for the Network.

6.1 Data Quality

Data Quality Expectations

Although a data set containing no errors would be ideal, the cost of attaining 95%-100% accuracy may outweigh the benefit. Therefore, we must consider at least two factors when setting data quality expectations:

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

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

National Park Service Mandate for Quality

High quality data and information are mandated by directives and orders, and they are vital to the credibility and success of the I&M program. Data need to meet national-level quality standards and need to be accessible to be used for wise and defensible decision-making at all levels. Data need to be able to be shared and aggregated with data from

other parks and from adjacent lands to support landscape-level and national planning and decision-making.

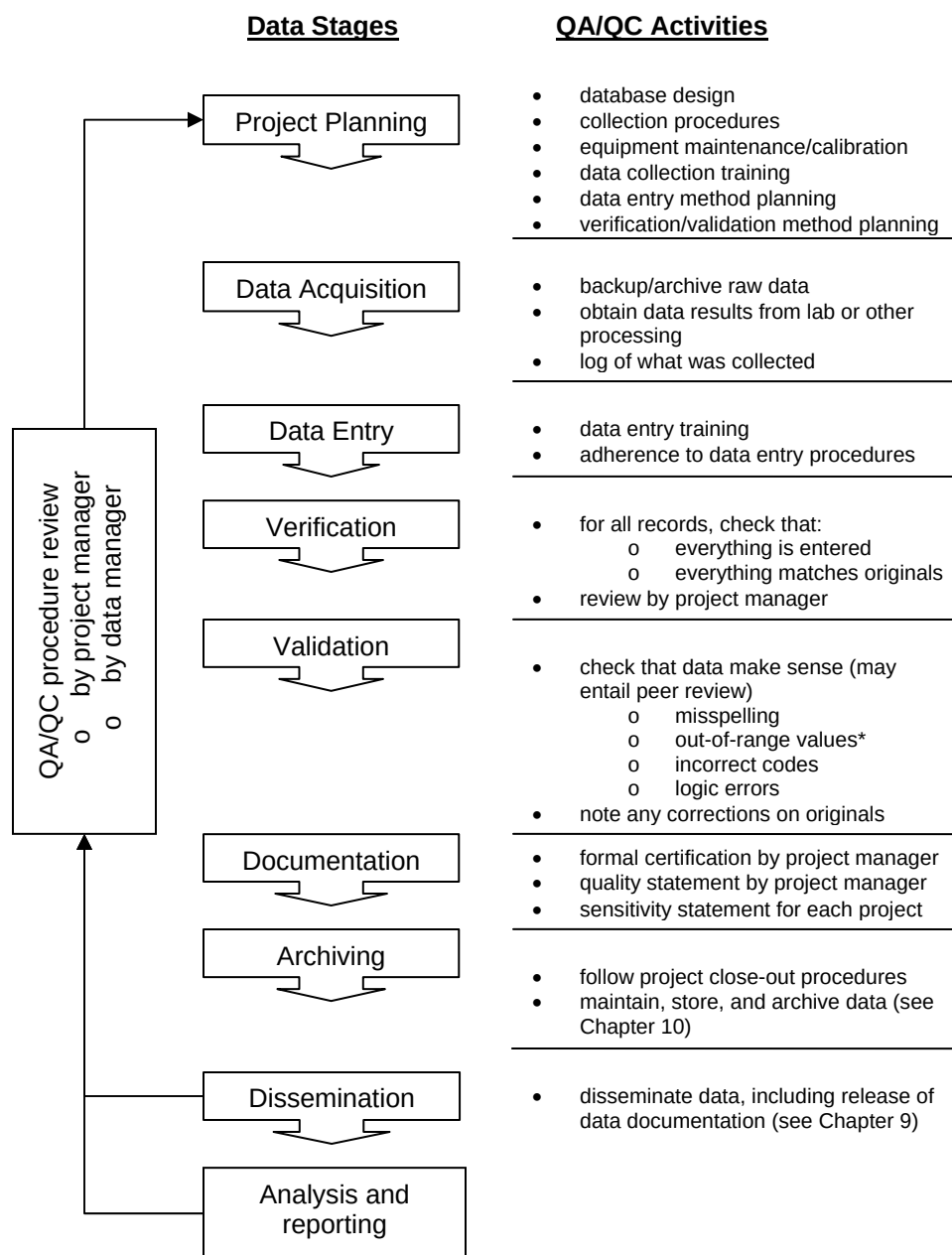
Director's Order #11B, which was issued by the NPS in 2002, "Ensuring Quality of Information Disseminated by the National Park Service," defines '*quality*' as incorporating three key components—*objectivity*, *utility*, and *integrity*.

Objectivity consists of: 1) *presentation*, which focuses on whether disseminated information is being presented in a proper context, in an accurate, clear, complete, and unbiased manner; and 2) *substance*, which focuses on the accuracy, usability, and reliability of the information.

Utility refers to the usefulness of the information to its intended users, from the perspectives of both the Network and the general public.

Integrity refers to the *soundness* of the data or the confidence one has in the data. Integrity is closely related to objectivity; however, it is possible to have subjective data of high integrity. The integrity of data is also related to data security. Data must, for example, be protected from unauthorized access or revision to ensure that the information is not compromised through corruption or falsification.

Director's Order #11B also specifies that information must be based on reliable data sources, which are accurate, timely, and representative of the most current information available. These standards apply to NPS-generated information as well as information provided by other parties if the NPS disseminates or relies upon this information.



* CAUTION! Care must be exercised when culling 'out-of-bounds' data. See text for discussion.

Figure 6.1 General course of data and associated QA/QC procedures. Quality control with regards to data analysis is specific to each project and addressed in appropriate standard operating procedures

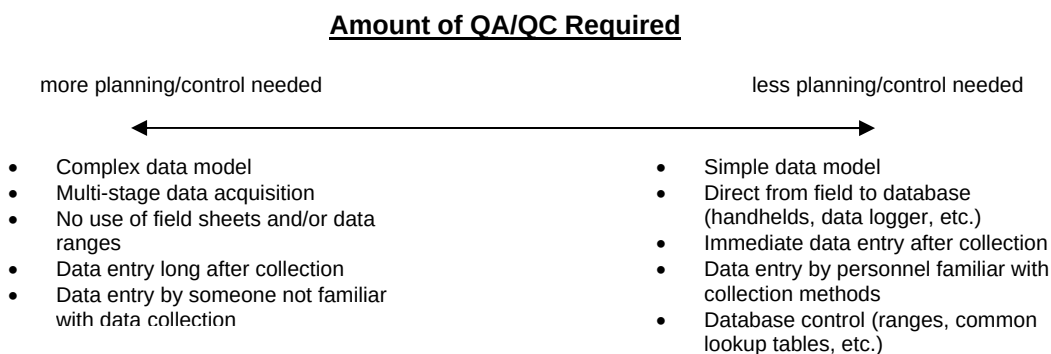


Figure 6.2 Some common data management elements affecting the amount of QA/QC needed. Planning and training for data entry and collection is always a critical component

6.2 Roles and Responsibilities

Quality assurance methods should be in place at the inception of any project and continue through all project stages to final archiving of the data set. It is critical that each member of the data management group work to ensure data quality. Everyone plays a part in producing and maintaining high quality data. Anyone assigned to a project is responsible for the quality of the results generated from his or her task(s). While Chapter 2 specifies data management roles and responsibilities, selected QA/QC duties are emphasized here.

Project managers must:

- be aware of quality protocols and convey their importance to technicians and field crews
- ensure compliance with the protocols
- plan for and ensure proper execution of data verification and validation
- review all final reports and information products

Technicians must

- follow established protocols for data collection, data entry, and verification established in the project SOPs
- help make project manager aware of opportunities to improve data quality

The Network Data Manager and Assistant Data Manager are responsible for:

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

6.3 General Operations

Version Control and File Naming Standards

Version control is the process of managing copies of changing files over the course of a project; file naming standards are critical for effective version control. Because of the scope and quantity of files being consolidated into one place, it is critical that computer files be given names that will uniquely identify them.

The following conventions apply to all NGPN file names:

- No spaces or special characters within the name
- Include date for version control, in yyyyymmdd format (e.g., 20070807 for Aug. 7, 2007)
- Use underscore as delimiters

File names for final products will begin with “NGPN” and contain a brief but clear explanation of the file content followed by an eight digit date and the extension. Most file names will require an indication of the QA/QC status of the information as a name element. Examples of QA/QC status include “draft”, “raw”, “verify”, and “valid”. Each of these file name elements (excepting the extension) is preceded by an underscore (“_”). Additional underscores may be added for clarity. As an example, “NGPN_Data_Mngmt_Plan_draft_20070824.doc” indicates a draft version of the data management plan dated August 24, 2007.

Before making major changes to a file, a copy of the file with the appropriate version control (in this case the file name) should be made. As indicated above, files are stored with the appropriate eight digit date which serves as version control. This allows changes to be tracked over time and facilitates collaboration between multiple personnel working on common files. With proper controls and communication, versioning ensures that only the most current version is used in any analysis. Including the date, formatted as yyyyymmdd, in the file name provides logical version control.

File Structure

The Network has developed a directory structure to allow for an organized and efficient method for storing digital files. The directory structure provides a framework for the file structure to evolve through time. Any changes to the directory structure should conform to the following rules:

- Provides a consistent and logically organized structure that complements the Network’s day-to-day operations
- Minimizes confusion about where information is located
- Allow for efficient, comprehensive, regular and secure means of backing up information
- Accommodate various security settings for different types of users
- Be hierarchical and minimize the ambiguity of where information should go. Ideally, the organization should be logical enough that one could navigate through the entire structure without ever requiring a map
- Separate in-progress/draft/unofficial/in-progress files from final/official information

- Be broad enough to clearly accommodate all types of information handled by the Network staff
- Be simple without sacrificing functionality
- Minimize the number of sub-folder levels. Ideally there should never be more than 8-10 subfolders within any primary directory.
- Newly created sub-folders should be organized by themes (e.g., hydrology, soil, etc.), rather than format (e.g., documents, GIS)

Information should never be arbitrarily divided in a way that increases the possibility of corruption or loss (e.g., metadata should never be separated from the data). When possible and appropriate, the structure should conform to other NPS and national directory structure standards. In keeping with these procedures, the following rules will apply to any folder expansions:

- Fiscal Years are represented by FY0x. Fiscal years begin October 1. Example: FY04.
- Subdirectories named “old” indicate older drafts of documents that are being temporarily archived. These folders are to be removed following final publication and archiving.
- Folder names should never have spaces or special characters (e.g., \$%\&*).
- Primary folders, or those that are close to the root directory and unlikely to change in name or location are in UPPERCASE.
- Two words are separated by an underscore (e.g., FIRST_SECOND).
- Secondary folders, or those that are many levels below the primary or are not accepted as a stable element of the NGPN directory structure are to follow Proper Case naming convention (e.g., FirstSecondThird).
- Sensitive information that must be managed by the Network will be placed in a /SENSITIVE/ sub-folder.
- In cases where a folder or file is inappropriately located, staff should notify the Network Data Manager.

The primary Network drive contains all of the information related to administration, monitoring, projects, and tools; this drive is located on the NGPN network server.

Roles and Responsibilities

The Network Data Manager is responsible for:

- Coordinating changes in the Network data structure
- Modifying any root level changes to the directory structure
- Informing Network staff of changes to the directory structure

All NGPN staff are responsible for:

- Regularly being informed of the directory structure and following the guidelines
- Adhering to the Network’s sensitivity guidelines when storing and accessing sensitive information.

External Data

Some of the NGPN Vital Signs involve laboratory analysis conducted via contract with certified laboratories. All data for a given Vital Sign, including lab results, will be housed in a distinct relational Microsoft Access database. Laboratories that will be providing data for a given Vital Sign will be supplied with a copy of the application so that data may be entered in a format consistent with the rest of the data for that Vital Sign. While most professional laboratories exercise their own QA/QC procedures, results received by a project manager are subject to the same QA/QC measures exacted on other project data.

6.4 Quality Assurance and Quality Control Mechanisms

Palmer and Landis (2002) define *Quality Assurance* (QA) as “an integrated system of management activities involving planning, implementation, documentation, assessment, reporting, and quality improvement to ensure that a process, item, or service is of the type and quality needed and expected by the consumer.” They define *Quality Control* (QC) as “a system of technical activities to measure the attributes and performance of a process, item, or service relative to defined standards.” Quality Assurance procedures maintain quality throughout all stages of data development. Quality Control procedures monitor or evaluate the resulting data products.

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

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

Overview of NGPN QA/QC

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

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

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

The NGPN holds as policy that techniques and procedures chosen for a project will maximize data quality. Quality assurance will be achieved by stipulating that:

- Common lookup tables are created for values of parameters recorded in an identical field for more than one project (such as common weather metrics)
- Each project has SOPs for core data management (collection, entry, verification, etc.)
- Each project use, at a minimum, standard field sheets for data collection (see Chapter 5)
- Database data entry computer applications resemble field sheets
- Automated error checking features be included in database applications
- Database application design will maximize the use of auto-fill, auto-correct, value range limits, pick lists, and other constraints specific to projects
- Database applications will include a means to track errors reported on the data after dissemination
- Database maintenance logs will be maintained for each NGPN database and housed in association with database files

6.5 Goals and Objectives

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

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

The most effective mechanism for ensuring that a project produces data of the right type, quality, and quantity is to provide procedures and guidelines to assist the researcher in accurate data collection, entry, and validation. We will initiate a comprehensive set of SOPs and data-collecting protocols for quality control, field methodologies, field forms, and data entry applications with built-in validation mechanisms.

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

6.6 Data Collection

Chapter 5 addressed the general acquisition and initial handling of data. Attention to detail during data collection is crucial to overall data quality. With sufficient effort, errors made in data entry or subsequent processing can always be corrected. However, faulty

data recorded during data collection may be uncorrectable. Such errors, if not captured during data collection, either are detected subsequently and treated as suspect (potentially making the record useless) or are not detected, potentially causing bias or imprecision in the results.

Discussion of Methods for Reducing Collection Errors

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

The project leaders and data manager need to ensure that field crews receive proper training. Although protocols and SOPs are in place, they cannot guarantee that high quality data will be collected. Prior to routine data collection for a project, conduct training sessions to ensure that field personnel have a clear understanding of data collection procedures described in the SOPs. A training program may also include a process to certify that field staff understand and can perform the specified data collection procedures. Moreover, data quality, accuracy, and especially consistency among observers can degrade even after extensive training. Such “data drift” frequently occurs even with experienced, highly skilled crew members. In many field projects, such errors may be impossible to correct once the field survey ends. Project leaders should assume that humans collecting data require calibration and validation at least as frequently as automated data-collection equipment. It is often invaluable to have crew members periodically measure the same attribute independently, compare values, and discuss results with their supervisor. It is important that everyone involved see these steps simply as part of good scientific practice, not as managerial tools for assessing shortcomings or ranking performance. The SOPs should outline how project leaders or crew leaders will implement periodic checks for consistency among crew members and accuracy of each person’s data. The development of a training manual may be helpful for long-term monitoring data collection efforts and those that will involve a large number of field staff. Palmer and Landis (2002) provide an outline for a training manual and suggestions for planning training sessions.

Data entry and validation are often seen as an “end-of-season” activity disconnected from data collection. However, many data collection problems may not be obvious or suspected until one can sort, summarize, and plot data. Rapid data entry and screening help reveal inconsistencies and systematic biases among crew members. Such screening may reveal, for example, that an individual is recording a particular species far more frequently than other crew members despite visiting the same sites, or that his/her measurements for a parameter are consistently higher. With any protocols involving multiple crew members and relying on crew expertise (e.g., in visual identification of species), daily data entry and exploratory summaries are critical during the first part of each field season.

6.7 Data Entry/Import

‘Data entry’ is the initial set of operations where raw data are transferred to database tables using a computerized form. When data are gathered or stored digitally in the field (e.g., on a data logger), data entry consists of transferring the data (downloading) to a file in an office computer where they can be further manipulated. The goal of data entry is to transcribe field observations into a computer database with 100% accuracy. In other words, that which is recorded in the field should be entered exactly in the database. Subsequent data verification is conducted to ensure that raw data matches entered data. Following verification, data validation may result in changes *to the entered data*. Data entry is a separate operation from data validation and care must be taken to not impose validation (beyond that automatically imposed by programming rules in a database) during data entry.

The NGPN adopts the following precepts affecting data quality with regards to data entry:

- Data entry will be done by someone familiar with data collection. The project manager (with assistance from the data manager if needed) must ensure that data entry staff is familiar with the database software, database structure, and any standard codes for data entry used by the Network. At minimum, data entry technicians should know how to open a data entry form, create a new record, and exit the database properly. They must learn how to commit both a ‘field’ entry and a ‘complete record’ entry and to correct mistakes made while typing.
- If feasible, data will be entered by two qualified persons; one person will read the observations and the other will enter the data.

The Network Data Manager, Assistant Data Manager, and the project leader should provide training in the use of the database to all potential users – especially those responsible for data entry and quality assurance. The project leader makes certain that data entry technicians understand how to enter data and follow the protocols.

Methods for Reducing Data Entry Errors

- 1) *Enter or download data in a timely manner.* All data should be entered or downloaded into the project database as soon as possible, preferably at least once a week if feasible. Try to avoid delaying data entry until all the project data have been collected. Timetables for data entry will be specified in the SOPs.
- 2) *Design efficient data entry forms and methods.* A full-screen data entry form that mimics the field data forms can effectively reduce manual data entry errors due to the 1:1 correspondence of the attributes. A strategy to distinguish between validated data and newly entered data should be adopted. Data can be entered into an empty, fresh database table to avoid contaminating existing data and the new data appended to the master data only after formal verification, validation, and documentation. Alternatively, we can include validation attributes that indicate which data have been checked and validated by the project leader in the database. Regardless of strategy, we must clearly document the process for validation in the protocol data management SOP.

- 3) *Build automated error checking features into the database.* The most robust QA/QC measures for data entry should be built into the database design to perform automatic validation checks of data. Data entry forms reduce transcription errors through auto-filled fields, range limits, pick lists, and spelling checks. They provide controlled access to the database (i.e., forms are set for data entry only, which prevents accidental deletion or alteration of existing data). They control the sequence of data entry (i.e., certain fields require an entry before more information can be entered). They warn the operator when errors are made and provide an opportunity for correction before the data are committed to a file.
- *Auto-filled fields.* Whenever possible, the data in a field should be auto-filled by the computer. For example, if a location ID is composed of a park code, project code, and a unique number, those elements are automatically inserted into the location ID field, ensuring that the record always contains a unique identifier.
 - *Range limits.* Where the appropriate values for a particular field span a finite range, the data entry program can check the entered value against the specified minimum and maximum values for that parameter. When a value is outside the accepted range, a warning message appears and asks the user to reenter a valid value. For some fields, values outside a specified 'normal' range may be acceptable. In this case, the warning message asks the user to verify the entry before continuing.
 - *Pick lists.* The data entry application may also use pop-up pick lists for standardized text items where spelling errors can occur. For example, rather than typing in a species code or name (where a misspelling generates a new species in the database), the code or name is selected from a list of valid species codes or scientific names and automatically entered into the species field. A pick list may also be used when only certain entries are acceptable. Lists are not appropriate for all written fields but should be used when appropriate.
 - *Unique constraints.* Duplicate and incorrect data entry can often be caught with the application of unique constraints on data entry fields. These constraints are particularly useful when importing data from other applications.
- 4) *Provide a clean, organized work environment.* Desktop space near the computer should be free of clutter and distractions that could cause the technician to lose her place. There should be enough space for two stacks of paper documents, one from which data are being entered and one from which data have been entered. A pad or notebook and some fine colored markers should also be available for making notes. (The need for a clean workspace also applies to the verification and validation phases.)

6.8 Verification and Validation Procedures

We appraise data quality by applying verification and validation procedures as part of the quality control process. These procedures are more successful when preceded by effective quality assurance practices. *Data verification* checks that the digitized data

match the source data, while *data validation* checks that the data make sense. It is essential that we validate all data as truthful and do not misrepresent the circumstances and limitations of their collection. Failure to follow SOPs for data entry, validation, and verification will render a data set suspect. Although data entry and data verification can be handled by personnel who are less familiar with the data, validation requires in-depth knowledge about the data and the resource being measured.

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

Methods for Data Verification

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

- 1) *Visual review at data entry.* The data entry technician verifies each record after input. Recorded values in the database are compared with the original values from the hard copy. Identified errors are immediately corrected. This method is the least complicated since it requires no additional personnel or software. Its reliability depends entirely upon the person keying data and thus, is probably the least reliable data verification method.
- 2) *Visual review after data entry.* All records are printed upon completion of data entry. The values on the printout are compared with the original values from the hard copy. Errors are marked and corrected in a timely manner. If only one technician is available to perform the review, the method's reliability increases if someone other than the person keying the data performs the review. Alternatively, two technicians can perform this review, a reader and a checker. One technician reads the original data sheets, and the second reads the same data on the printout.
- 3) *Duplicate data entry.* The data entry technician completes all data entry, as normal. Random records are selected (every *n*th record) and entered into an empty replica of the permanent database, preferably by someone other than the person keying the permanent data. A database query is then used to automatically compare the duplicate records from the two data sets and report any mismatches. Disparities are manually reviewed and correction applied if necessary. This method adds the overhead of retyping the selected records, as well as the creation of a comparison query. However, it becomes increasingly successful as the number of duplicate records increases. Professional data entry services frequently use this method.

We can also use the entered data to calculate simple summary statistics with statistical software. These summary statistics can help us catch a duplicate or omitted entry. For

example, we can view the number of known constant elements, such as the number of sampling sites, plots per site, or dates per sample. We can pose the same question in different ways; differences in the answer provide clues to errors. The more checks we devise to test the completeness of the data, the greater our confidence that we have completely verified the data.

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

Validation

Although we may have correctly transcribed the data from original field notes or forms, they still might be inaccurate or illogical. For example, entries of stream pH of 25.0 or a temperature of 95°C in data files raise doubt about their accuracy; and such entries almost certainly are incorrect, whether or not they were properly transcribed from field forms. This process of reviewing computerized data for range and logic errors we call *validation*, and it can accompany data verification *only* if the operator has comprehensive knowledge about the data. More often, validation is a separate operation carried out *after* verification by a project specialist who can identify generic and specific errors in particular data types. Corrections or deletions of logical or range errors in a data set require notations in the original paper field records about how and why the data were changed. Modifications of the field data should be clear and concise while preserving the original data entries or notes. Validation efforts should also include a check for the completeness of a data set since field sheets or other sources of data could easily be overlooked. Often, it is valuable to have field crews keep a log of the total number of data sheets (specimens, etc.), collected at each site visit, so it is simple to do a tally later to determine all data sheets are present.

General step-by-step instructions are not possible for data validation because each data set has unique measurement ranges, sampling precision, and accuracy. Nevertheless, validation is a critically important step in the certification of the data. Invalid data commonly consist of slightly misspelled species names or site codes, wrong dates, or out-of-range errors in parameters with well defined limits (e.g., elevation). But more interesting and often puzzling errors can be detected as unreasonable metrics (e.g., stream temperature of 70°C) or impossible associations (e.g., a tree 2 inches in diameter and 100 feet high). We call these types of erroneous data *logic errors* because using them produces illogical (and incorrect) results. The discovery and correction of logic errors has direct, positive consequences for data quality and provides important feedback to the methods and data forms used in the field. Histograms, line plots, and basic statistics can reveal possible logic and range errors.

The NGPN adopts the following procedures regarding data validation:

- Each NGPN project protocol will address a process for data validation that includes at least one of the methods outlined in the Network data management plan. That process will adhere to the precepts given in the plan.

- Corrections or deletions as a result of data validation require notations in the original paper field records about how and why the data were changed, with the editor's initials.
- Modifications of the field data should be clear and concise while preserving the original data entries or notes (i.e., no erasing!).
- Validation efforts should also include a check for the completeness of a data set since field sheets or other sources of data could easily be overlooked.
- The NGPN will maximize the use of automated routines and/or data summary and visualization such as histograms, line plots, and basic statistics to reveal possible logic and range errors.
- The NGPN will maximize the use of database programming to control data entry. In general, this will be achieved via the use of lookup tables, but may be accomplished by field-type design in a data base (such as "yes/no" field-types).

Methods for Data Validation

Specific procedures for data validation depend upon the Vital Sign being monitored (project subject) and will be included in the project protocols. The following general methods can be used to validate data:

- 1) *Data entry application programming.* Certain components of data validation are built into data entry forms. This method is essentially part of database design and is discussed earlier in this chapter. Not all fields, however, have appropriate ranges that are known in advance, so knowledge of what are reasonable data and a separate, interactive validation stage are important.

Caution must be exercised when using lookup tables to constrain variable values. Values occurring outside the range set by a lookup table (established during database design) may not always be invalid. As part of data validation procedures, the project manager is responsible for correct use of lookup tables or other automated value range control.

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

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

Database, graphic, and statistical tools can be used for ad-hoc queries and displays of the data to detect outliers. Some of these outlying values may appear unusual but prove to be quite valid after confirmation. Noting correct but unusual values

in documentation of the data set saves other users from checking the same unusual values. Moreover, outliers are relevant both to data validation and to planned data analyses (e.g., presence of outliers can be addressed through robust statistical approaches).

- 3) *Other exploratory data analyses.* Palmer and Landis (2002) suggest that in some cases, calculations for assessments of precision, bias, representativeness, completeness, and comparability may be applicable and that for certain types of measurements, evaluation of a detection limit may also be warranted (these authors provide examples of procedures that may be applicable). Normal probability plots, Grubb's test, and simple and multiple linear regression techniques may also be used (Edwards 2000).

6.9 Data Quality Review and Communication

The National Park Service requires QA/QC review prior to communicating or disseminating data and information. Only data and information that adhere to NPS quality standards may be released (see earlier treatment in this Chapter). Data and information disseminated to the public must be approved by the appropriate reviewing officials and programs. Documentation of the QA/QC standards used in producing the information and that substantiate the quality of the information must be formally certified and distributed with related data and information. Also, mechanisms must be in place for receiving and addressing comments or complaints pertaining to the quality of data. Chapter 9 addresses the dissemination of NGPN data and information.

Monitoring Conformance to Plans and Standards

As part of the close-out and evaluation stage of each NGPN project, QA/QC procedures will be reviewed by the project manager and recommendations for change will be included in the annual report. Additionally, the NGPN data manager will review and revise the QA/QC procedures included in the data management plan as needed. While the data management plan will be reviewed on a regularly scheduled plan, QA/QC procedures are subject to revision as needed by the Network Data Manager.

The data manager will conduct periodic 'spot checks' of random NGPN monitoring projects to ensure compliance with the data management plan and project protocol QA/QC procedures. The data manager will track and facilitate the correction of any deficiencies found during this process. The project manager and Network Data Manager are responsible for ensuring that non-conformities in data management practices are corrected.

Periodic checks by the data manager to see if Network staff are adhering to the data quality procedures established in the Data Management Plan and protocols SOPs may include verification that:

- Data collection and reporting requirements are being met
- Data collection and reporting procedures are being followed
- Verification and validation procedures are being followed
- Data file structures and maintenance are clear, accurate, and according to plan
- Revision control of program documents and field sheets are adequate

- Calibration and maintenance procedures are being followed
- Seasonal and temporary staff have been trained in data management practice
- Metadata collection and construction for the program proceeds in a timely manner
- Data are being archived and catalogued appropriately for long term storage

Documenting and Communicating Quality

The final step in the QA/QC for a given data set is the preparation of summary documentation that assesses the overall data quality. A statement of data quality will be composed by the project manager and incorporated into formal metadata as well as the NGPN primary data repository. Metadata for each data set or database will also provide information on the specific QA/QC procedures applied and the results of the review. Typically, data quality information will be conveyed as part of FGDC-compliant metadata (see Chapter 7 for data documentation). Metadata and data will be available via both the NGPN website and the NPS Data Store (see Chapter 9 for dissemination information).

7. DATA DOCUMENTATION

Collecting and maintaining high quality data is not enough to ensure its utility for future users. Even the highest quality data can become nearly useless if users cannot determine how data were collected and processed. 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. In all instances documentation provides 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. Other documentary elements might include Standard Operating Procedures (SOPs), flow charts, project justification statements and study plans, and a record of changes to a sampling protocol. For these reasons, the focus of this chapter goes beyond a discussion and justification for metadata by emphasizing the importance of more complete documentation. While metadata are addressed in context throughout this plan, this chapter outlines the NGPN strategy for generating and handling metadata (Figure 7.1).

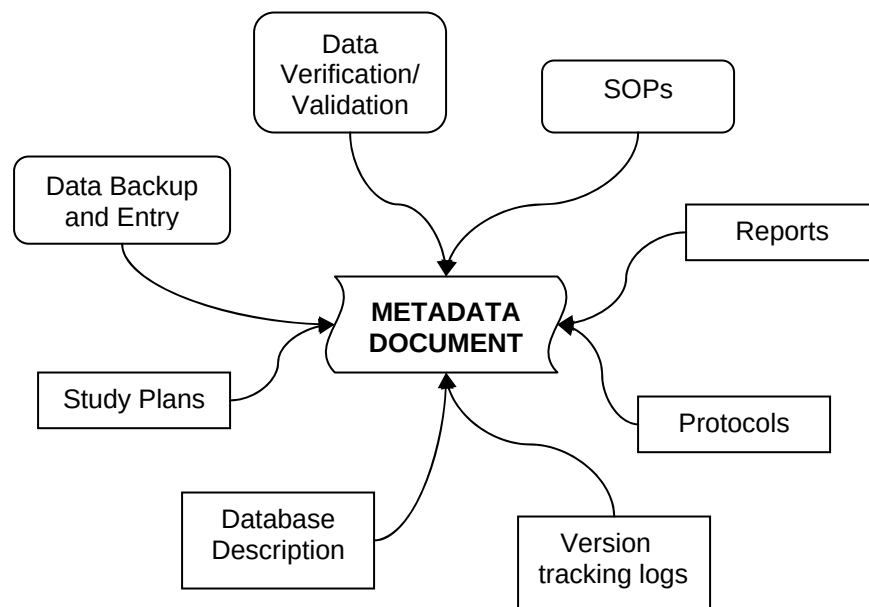


Figure 7.1 Project and data documentation. The metadata document contains all information pertaining to project data

7.1 Mandate for Documentation

Executive Order 12906 mandates that federal agencies “...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, www.fgdc.gov/metadata/contstan.html). In addition, EO 12906 directs agencies to plan for legacy data documentation and to provide metadata and data to the public.

The FGDC Biological Data Profile (www.fgdc.gov/standards/status/sub5_2.html) 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, www.nbii.gov/datainfo/metadata) Clearinghouse. Although not a requirement, completion of the Biological Data Profile for appropriate data sets is recommended.

All GIS data layers must be documented according to applicable FGDC and NPS metadata standards. The NPS GIS Committee requires that all GIS data layers be described using FGDC standards and the NPS Metadata Profile (nrdata.nps.gov/profiles/NPS_Profile.xml).

Documentation Roles and Responsibilities

The project manager and Network Data Manager are responsible for documenting data, as shown in Table 7.1. See Chapter 2 for additional discussion of project roles and responsibilities concerning documentation.

Table 7.1. Data documentation responsibilities

Project Manager	Data Manager
• Document project planning, execution, and progress in reports, correspondence, project-tracking database, etc. • Generate formal data set metadata	• Facilitate metadata generation by providing tools and assistance in their use • Establish metadata dissemination protocols (cataloguing of metadata on accessible servers) • Format metadata for dissemination

The duties for the project manager, including direction of project personnel and general project management, are aimed at producing a comprehensive metadata document.

The Network data manager is cast in the unwritten role of motivator of those with metadata generation responsibilities. The data manager ensures that metadata are available in formats suitable for distribution (see [Metadata Formats](#) below).

7.2 Documentation Process

Metadata Process

Figure 3.1 indicates the points within the standard project workflow for metadata development. Metadata development begins with project design and planning. The

NGPN will utilize the NPS Metadata Tools & Editor for the creation of project metadata. As the project progresses, this document will be augmented by the project manager to include relevant project details. Full development of metadata for project data will be completed after the data set is certified by the project manager. Once metadata are complete, the Network data manager saves the document in one or more of four acceptable formats (Table 7.2). Metadata are saved on the primary NGPN server along with data and copied to the I&M Program NPS Data Store (Figure 7.2). Hypertext links within the metadata document will point to accompanying data sets served on the primary NGPN server. In general, a single metadata document will apply to both raw and certified versions of the data. Metadata records will be stored with both hard copy and digital archive data.

Metadata Tools

As noted above, the NGPN will principally utilize the NPS Metadata Tools & Editor (standalone or ArcCatalog plugin) for creating metadata. The end goal is a formal, comprehensive metadata document that meets federal (FGDC) and NPS standards. Additional applications for metadata creation are available, and the NGPN may utilize these programs for specific components of metadata creation, recognizing that the tools and means of achieving this end product are numerous and changing. Project managers are encouraged to utilize the NPS Metadata Tools & Editor as well, but may use any tool deemed suitable for the project that meets program requirements.

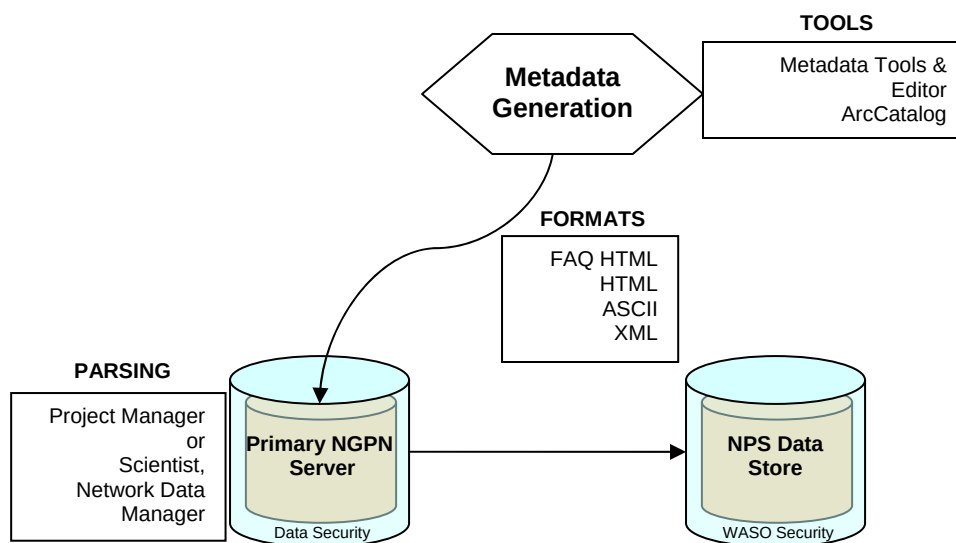


Figure 7.2 Metadata repositories for the NGPN. The NGPN server acts as primary storage of metadata. Metadata are available publicly via the NPS Data Store

Metadata Formats

Table 7.2 lists the metadata file formats to be saved on the primary NGPN server.

Table 7.2. Metadata file formats.

<u>Format</u>	<u>Purpose</u>
HTML	Standard format for web browsers
FAQ HTML	Frequently Asked Questions version of metadata
ASCII	Standard text file
XML	Extensible Markup Language for application programming

These file formats will be viewable on the NGPN website. The XML format will be the preferred format for metadata exchange and archiving. The HTML file version will serve as the primary file for display on the website, and internal hyperlinks will allow for easy viewing, saving, and printing of the other formats.

Metadata Maintenance

The NGPN procedures for metadata generation and maintenance will provide details on the mechanics of handling metadata. Generally, metadata will be created by the project manager, with assistance from the data management staff, via an application interface tied to project data on the primary NGPN server. The Network Data Manager/Assistant Data Manager ensures that updated metadata are copied to the NPS Data Store.

Protocol Versions

Chapter 6 noted that database design for NGPN data will include a means for recording the protocol version under which each piece of data was collected. The metadata document will contain information about protocol versions used to collect the data.

7.3 Non-Programmatic Data Documentation

Data which are generated and/or managed outside the I&M Program, but used in analysis with NGPN data (see Chapter 5) or distributed in any manner by the NPS, require the same level of documentation as NGPN-generated data. This includes data produced under contract with the NPS. For non-Program data, the project manager will ask the originating entity for metadata. Any contract entered into by the NGPN with data producers will stipulate that FGDC-compliant metadata in HTML, ASCII, or XML format, or in a format specified by the project manager and approved the Network Data Manager. The Network Data Manager will assist the project manager in metadata acquisition by providing tools, format protocols, and file transfer services.

Derived Data Documentation

Several of the NGPN Vital Sign monitoring projects will produce data sets derived from data collected and certified by other entities. When a derived data set becomes a data product, metadata are required for that data set. Like regular project data, derived data products may not be released without proper documentation. Metadata for any such derived data will include a data lineage pointing to parent data sets.

Legacy Data Documentation

Data acquired via data mining or other means does not enjoy the same level of stewardship afforded project data. When older data that have no current steward (typically a project manager) are found to be useful, metadata for those data must be

established. Also, as noted in Chapter 9, no data may be released without accompanying metadata. If ancillary data are used in the creation of derived data products (for example, ancillary data are used in an analytical comparison to project data), metadata for those data are necessary to fully document the derived data product.

8. DATA ANALYSIS AND REPORTING

Well-designed monitoring and a sound data management plan are necessary but not sufficient for ensuring a successful monitoring program. Monitoring in the NGPN will be credible and useful only if data are analyzed, interpreted, and provided to managers and others at regular intervals and in a format appropriate for each audience.

8.1 Importance of Data Reporting

Each Vital Sign protocol establishes requirements for data summary, analysis, and reporting. Integrated queries and reports are integrated into the reporting function of most Vital Signs databases. Monitoring protocols also include spatial data analysis as an essential part of data reporting. The graphical presentation of locations, spatial relationships, and patterns of change over time can be powerful tools for both analyzing data and communicating results to a broad audience. The primary role of data management staff is to prepare data as needed for analysis and reporting in coordination with the project manager.

For each Vital Sign and for the NGPN I&M Program overall, we will specify the types of reports that will be generated, who will be responsible for generating each report, and the schedule for reoccurring reports. Such planning will clarify expectations for these reports and ensure that there is sufficient program accountability, documentation, and evaluation. The NGPN Monitoring Plan presents an overview for monitoring projects whose implementation will be funded in the near-to mid-term. More detailed description of data analysis will also be contained in the protocols associated with each monitoring project.

8.2 Types of Reports

Annual Reports

The major purposes of annual reports are to:

- summarize annual data and document monitoring activities for the year
- describe current condition of the resource being monitored
- document changes in monitoring protocols
- increase communication within the park and network.

Many of our monitoring programs will be active each year, and those programs will generate annual reports. However, some sampling regimes do not require annual activity. For example, some protocols may collect data less frequently. Those programs will produce “annual” reports only when there are significant monitoring activities to document.

Annual reports will be generated from automated data analyses developed for each monitoring project.

- *Audience:* Network staff, park staff including administration, scientists working in parks
- *Review:* internal network review

Analysis and Synthesis Reports

The role of analysis and synthesis reports is to:

- determine status and trends in condition of resources being monitored
- document previously unknown characteristics of resources and correlations among resources being monitored
- analyze data to determine amount of change that can be detected by this type and level of sampling
- provide context, i.e. interpret data for the park within a multi-park, regional or national context
- recommend how results can be used for resource management (feedback for adaptive management).

These reports can provide critical insights into resource status and trends, which can then be used to inform resource management efforts and regional resource analyses. This type of analysis, more in depth than that of the annual report, usually will require several years of sampling data. Therefore, these reports are not written more frequently than every three to five years for resources sampled annually. For resources sampled less frequently, or which have a particularly low rate of change, intervals between reports may be longer.

It is important that results from all monitoring projects within and across all NGPN parks be integrated across disciplines in order to interpret changes to park resources. This will be accomplished with a network synthesis report produced at no more than 10-year intervals.

- *Audience:* superintendents, park resource managers, network staff, and external scientists
- *Review:* external, blind peer-review with at least 3 subject-matter experts including one statistician

NPS Natural Resource Publications Series

The Natural Resource Publications series consists of five series: Natural Resource Report, Natural Resource Technical Report, Scientific Monograph, Natural Resource Year in Review, and Park Science. Manuscripts must adhere to a set of minimum standards and are peer-reviewed to ensure that information is scientifically credible, technically accurate, appropriately written for the intended audience, and designed and published in a professional manner. Publication in the Natural Resource Report (NRR) and Natural Resource Technical Report (NRTR) is overseen by NPS Key Officials who are designated by the Associate Director for Natural Resource Stewardship and Science for the national series, and the Regional Director for regional series. Regional series address natural resource topics that are of interest and applicability to a specialized or limited geographical readership in the National Park Service and to others charged with managing natural resources, and may feature preliminary data and results of study or summaries of ongoing study for rapid dissemination to park managers. The NPS Key Official oversees the process, but most of the responsibility for using the correct format/style and incorporating comments from peer reviewers is assumed by the contributors (authors) since the NPS does not have the funding to hire editors for the series. The contributor bears any costs for printing and distribution of the reports. The majority of reports are distributed via the Internet.

Step-by-step instructions, examples, and templates for preparing, submitting, and publishing reports in these two series are available on the Natural Resource Publications Management (NRPM) website at

<http://www.nature.nps.gov/publications/NRPM/index.cfm>. To facilitate the preparation and peer review of draft manuscripts, and the timely publishing of routine scientific reports over the internet, the minimum style/formatting standards and templates use readily available fonts and minimal formatting that do not require desktop-publishing skills. However, reports that will be printed and distributed by hardcopy, or for which a large readership is expected, contributors should produce a final report using Frutiger and NPS Rawlinson fonts that is consistent with the “Communicating the National Park Service Mission” design standards and Director’s Order 52. Instructions, examples and templates are available on the NRPM website to assist contributors with the final design and layout steps.

- *Audience:* superintendents, park resource managers, network staff, and external scientists
- *Review:* optional, at discretion of NPS Key Official

Scientific Journal Articles and Book Chapters

This aspect of the program will be directed by the program managers, and is more at their discretion than previous reports. Publishing scientific journal articles and book chapters is primarily conducted to communicate advances in knowledge and is a very important, widely acknowledged means of quality assurance and quality control, via the academic peer-review process. Putting a program’s methods, analyses, and conclusions under the scrutiny of a scientific journal’s peer-review process is one of the best ways to ensure scientific rigor.

Scientific journal articles and book chapters produced by NGPN efforts are tracked by the NGPN monitoring program; new publications are part of the Annual Administrative Report and Work Plan, which is sent to the regional and national offices each year. Additionally, all scientific journal articles and book chapters will be entered into the NatureBib database. Principle Investigators of recently published work in the NGPN frequently make presentations at professional workshops and conferences and may be invited to present their findings at the annual NGPN meeting.

- *Audience:* scientific community
- *Review:* peer-review conducted by journal or book editor

Interpretation and Outreach

Scientific information gained from monitoring programs usually requires a concerted effort to be translated for the general public. Network staff share discoveries with the public by contributing articles to natural history newsletters and other media in the parks which produce them. In the future, the Network plans to produce brochures and fact sheets regarding monitoring. The NGPN program will also provide information to the community in a digestible form, easily available on the NGPN website.

- *Audience:* the interested public
- *Review:* reviewed by project leads for accuracy

8.3 Overview of Process for Data Reporting

Providing meaningful results from data summary and analysis is a cornerstone of the I&M Program and characterizes the program's data management mission to provide useful information for managers and scientists. The associated data management objective is to provide valid data in formats that support scheduled and ad hoc display, query, analysis, summary, and reporting of data. Routine and scheduled data summary and analysis requirements and procedures are identified in each Vital Sign monitoring protocol.

The data managers will work with the project managers, network ecologists, and others to specify and design or adapt database objects, fields, and values to support the formats and functions necessary for analysis using statistical software applications such as R. Summary and reporting functions required by Vital Sign monitoring protocols can be developed within Microsoft Access database applications where data are stored. The statisticians, ecologists, and others will use existing and custom data conversion and export functions in MS Access to prepare datasets for import into other software applications.

Spatial analysis and maps will be produced by the Network Data Manager. Ad hoc queries and reports will be handled on a case by case basis due to their dynamic nature. Data analysts and the Network Data Manager will identify and develop the data conversion routines necessary to generate these formats for analysis. GIS functions can also contribute to understanding long term status and trends of Vital Signs and ecosystems. The network will provide geographic data display, summary, and mapping services for scheduled reports and other requests. Methods may be developed for using GIS to visualize time-series data, perform geostatistical, functions, and do spatial network analysis with hydrography, transportation, and other linear features. Tabular and spatial results will be shared in reports and made available on the NGPN web site.

9. DATA DISSEMINATION

Data Management at the NGPN aims to ensure that:

- Data are easily discoverable and obtainable by all stakeholders
- Only data subjected to complete quality control are released, unless necessary in response to a Freedom of Information Act (FOIA) request
- Distributed data are accompanied by appropriate documentation
- Sensitive data are identified and protected from unauthorized access and inappropriate use

The NGPN will continue to provide links to public data products via the NGPN public website, located at <http://science.nature.nps.gov/im/units/ngpn/>. This will be the Network's portal for distribution of the most current data products available as the monitoring program develops. Distribution instructions for each data set will be provided in the respective metadata.

9.1 National Park Service Policy on Data Ownership

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

- All data and materials collected or generated using NPS personnel and funds become the property of NPS.
- Any important findings from research and educational activities should be promptly submitted for publication. Authorship must accurately reflect the contributions of those involved.
- Investigators must share collections, data, results, and supporting materials with other researchers whenever possible. In exceptional cases, where collections or data are sensitive or fragile, access may be limited.

Establishing Data Ownership Guidelines

The network has established guidelines for the ownership of data and other research information. To ensure that proper ownership, format, and development of network products is maintained, all cooperative or interagency work must be conducted as part of a signed collaborative agreement. Every cooperative or interagency agreement or contract involving the network must include [OMB Circular A-110](#) cited under the Reports and Deliverables Section of all agreements and contracts. The following shows appropriate language to use when citing Circular A-110:

“As the performing organization of this agreement, *<institution or organization name>* shall follow the procedures and policies set forth in Office of Management and Budget (OMB) Circular A-110. (<http://www.whitehouse.gov/omb/circulars/a110/a110.html>)”

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

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

“All reports and deliverables must follow the most recent versions of the Northern Great Plains I&M Network product specifications.”

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

9.2 Mechanisms for Data Distribution

The Network’s data products (final deliverables or periodic milestones) will be bundled and made available for distribution as a complete set that will enable managers, scientists, and the general public access to all pertinent data in a single download. The network will use the following applications and repositories to distribute data developed by the program:

- **NatureBib** – a web-based database housing natural resource bibliographic data for I&M Program parks.
(<https://science1.nature.nps.gov/naturebib/nb/simple/clean>)
- **NPSpecies** – a web-based database to store, manage and disseminate information on the status of organisms known or suspected to occur in NPS units. Information includes data on species status, abundance, residency, nativity, vouchers, observations, and supporting references.
(<https://science1.nature.nps.gov/npspecies/web/main/start>)
- **NPS Data Store** – an online repository for spatial and non-spatial metadata and associated data products. (<http://science.nature.nps.gov/nrdata>)
- **NGPN Website** – a source of detailed information about the network and its program. Data and products will either be available through the site, or users will be directed to where the data are stored. Metadata on all inventory and monitoring products developed as part of the Network’s I&M plan will also be posted to this site.
- **Water Quality Data**
Water quality data are managed according to guidelines from the NPS Water Resources Division. This includes using NPSTORET desktop database application to help manage data entry, documentation and transfer. The NGPN oversees the use of NPSTORET according to the network’s water quality

monitoring protocols and ensures the content is transferred at least annually to NPS Water Resource Division for upload to the EPA STORET database.

Information will be made available to two primary audiences: public and NPS employees, as determined by data sensitivity and development status. Only fully documented, certified, non-sensitive data and data products may be uploaded to public distribution repositories or otherwise released to the public. Access to NGPN data products will be facilitated via a variety of means that allow users to browse and download Network data and supporting documents (Table 9.1).

Table 9.1 Primary repositories for NGPN information and associated specimens

Item	Repository
Reports (public) • digital • hardcopy • bibliography	NPS Data Store, NGPN website and Park and Network libraries, park archives, NatureBib
• Network-generated digital data sets and data products (public, non-sensitive) • certified data and data products (including photos) • metadata	NGPN website, NPS Data Store, NPSpecies, NPSTORET
• Network-generated digital data sets and data products (NPS staff, sensitive) • raw, validated, and analyzed data • metadata • submitted reports • digital photos • digital presentations	NGPN intranet website; selected NGPN data may be housed externally with an established MOU.
• project product materials • specimen vouchers • photograph film	Park archives or other curatorial facility (according to project protocol)
Project administrative records or miscellaneous items (hard copy)	NGPN office, NPS archive facility

The national water quality database (NPSTORET) maintained by the NPS Water Resources Division provides a good example of data transfer from a Network project to an on-line master database, from which data and summary products are made widely available via a web-based clearinghouse service (Figure 9.1).

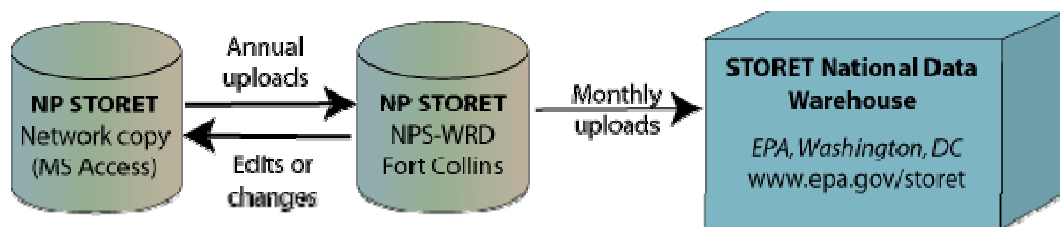


Figure 9.1 Schematic of data flow for transferring water quality data from a Network to the national master database, and for Network and public access to that master database via an online clearinghouse.

9.3 Network Website

The Network's web page is primary tool for Network communication. It will be a site for distributing reports, analyses, datasets, and other materials to its staff, member parks, the National Park Service, other government agencies, and the public. The web page is partitioned in two sites that are nearly identical structure, but differ in content. The first is the public internet-based web site with non-sensitive and non-proprietary materials. The second web site contains proprietary and sensitive materials considered of use and interest to the network staff, member parks, and other NPS personnel. The general format of the pages is guided by the standards set by the NPS and the I&M administration.

The NGPN website will provide an opportunity for NPS staff, cooperators and the public to provide feedback on data and information gathered as part of the Network's I&M Program. The metadata attached to each dataset available to the public will include information on where to send comments or questions regarding the specific data.

9.4 Freedom of Information Act and Sensitive Data

The Freedom of Information Act, 5 U.S.C. § 552, referred to as FOIA, stipulates that the United States government, including the National Park Service, must provide access to data and information of interest to the public, 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 exemption or by special law enforcement record exclusions. Under the terms of FOIA, agencies must make non-protected records available for inspection and copying in public reading rooms and/or from the internet.

The NPS is directed to protect information about the nature and location of sensitive park resources under one Executive Order and four resource confidentiality laws:

- Executive Order No. 13007: Indian Sacred Sites
- National Parks Omnibus Management Act (NPOMA; 16 U.S.C. 5937)
- National Historic Preservation Act (16 U.S.C. 470w-3)
- Federal Cave Resources Protection Act (16 U.S.C. 4304)
- Archaeological Resources Protection Act (16 U.S.C. 470hh)

When any of these regulations are applicable, public access to data can be restricted. If disclosure could result in harm to natural resources, the records may be classified as 'protected' or 'sensitive' and information withheld from the public. The following resources are recognized as sensitive by the NPS:

- endangered, threatened, rare, or commercially valuable NPS resources
- mineral or paleontological sites
- objects of cultural patrimony
- significant caves

The Network will comply with all FOIA restrictions regarding the release of data and information as instructed in NPS Director's Order #66 and accompanying Reference Manuals 66A and 66B (currently in development). Managing natural resource information that is sensitive or protected requires the following steps:

- Identification of potentially sensitive resources
- Compilation of all records relating to those resources
- Determination of which data must not be released in a public forum
- Management and archival of those records to avoid their unintentional release

Classification of sensitive data will be the responsibility of Network staff, park superintendents, and project managers. The Network Data Manager will work with project managers and the Network Coordinator to classify sensitive and/or protected information. Network staff will classify sensitive data on a case-by-case, project-by-project basis and will work closely with project managers to ensure that potentially sensitive park resources are identified, that information about these resources is tracked throughout the project, and that potentially sensitive information is removed from documents and products that will be released outside the Network.

Consider the following questions to determine whether information should be protected:

- Has harm, theft, or destruction occurred to a similar resource on federal, state, or private lands?
- Has harm, theft, or destruction occurred to other types of resources of similar commercial value, cultural importance, rarity, or threatened or endangered status on federal, state, or private lands?
- Is information about locations of the resource in the park specific enough so that the park resource is likely to be found at these locations at predictable times now or in the future?
- Would information about the nature of the park resource, which is otherwise not of concern, permit determining locations of the resource if the information were available in conjunction with other specific types or classes of information?
- Even if relatively out-dated, is there information that would reveal locations or characteristics of the park resource such that the information could be used to find the park resource as it exists now or is likely to exist in the future?
- Does NPS have the capacity to protect the park resource if the public knows its specific location?

Access Restrictions on Sensitive Data

Sensitive park resources will be identified through a cooperative effort. The Network and park staffs should identify all potentially sensitive park resources to the manager for each project. Reciprocally, the project manager must identify any known references to potentially sensitive park resources.

Network staff is responsible for managing access to sensitive data handled by the Program. All potentially sensitive park resources will be identified and investigators working on Network projects will be informed that:

- All data and associated information must be made available for review by Network staff prior to release in any format; and
- Any information classified as protected should not be released in any format except as approved in advance by the National Park Service.

When preparing information for any repository, the Network staff ensures that all protected information is properly identified and marked. All references to protected information are removed or obscured in any reports, publications, maps, or other public forum.

Network staff will remove any sensitive information from public versions of documents or other media. They will isolate sensitive from non-sensitive data and determine the appropriate measures for withholding sensitive data. The main distribution applications and repositories developed by the I&M Program are maintained on both secure and public servers and all records marked 'sensitive' during uploading will only become available on the secure servers. Procedures for assigning a sensitivity level to specific records when uploading to both the NPSpecies and NatureBib databases are given on the following websites:

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

Thus, access to data on sensitive park resources can be limited to Network staff and/or research partners. However, limits to how these data are subsequently released must also be clearly defined. It is crucial that the person(s) uploading records to the online applications (repositories) is familiar with the procedures for identifying and entering protected information.

NPS Only versus Public

Only data subjected to complete quality control are released, unless release is necessary in response to a FOIA request. Products of incomplete, poor, or questionable quality (typically legacy data) may not be appropriate for the public and should be managed separately from those of acceptable quality. These lower quality products may be the only source of information on the natural history of the park, may have been the basis for early management decisions, and may still be valuable in-house.

Provided low quality products do not contain sensitive data (as described earlier in this chapter), these data may be released to the public upon specific FOIA request. They must be accompanied with qualifying documentation.

National I&M Program applications such as NatureBib provide a means to flag data as "NPS Only" or "Public". Additional notes describing the quality should be added to the records in these applications, as resources permit.

9.5 Data Release Policy

The NGPN will maintain the following policy regarding the release of data and related products:

- 1) Only fully documented, certified, non-sensitive data and data products may be uploaded to public distribution repositories or otherwise released to the public or other non-NPS recipient.
- 2) Any released data or data products must be accompanied by full metadata and any additional supporting documentation as determined by the project manager.

- 3) Data from the NGPN monitoring program will be made available according to NGPN reporting and data integration schedules specified in the most recent versions of this document (see Chapter 8) and the NGPN Monitoring Plan.
- 4) Non-monitoring data will be released by NGPN upon completion of data certification (verification and validation) and on condition of project manager approval. However, data for which analysis and reporting have not been completed but which are otherwise certified will be released no later than one year after certification.
- 5) Distribution instructions for each data set will be included in respective metadata.
- 6) Distribution logs specifying recipient name and contact information, intended use of data, export file format, delivery date and method, data content description, and range (by date and geography) of data delivered and description of distributed items will be kept for each distinct data set or product managed by the NGPN.
- 7) At the discretion of the project manager, any reports utilizing the data may also accompany distributed data.

Datasets or other information will never be released before passing the classification procedure to ensure that sensitive or protected data are not made public.

10. DATA MAINTENANCE, STORAGE, AND ARCHIVING

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

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

Effective long-term data maintenance is inseparable from proper data documentation, and an essential part of any archive is accompanying explanatory materials (Olson and McCord 1998). This chapter will refer to, and in some cases elaborate on, metadata standards and dataset documentation procedures that are more fully explained in Chapter 7 (Data Documentation) of this plan. This chapter applies to documents such as final reports prepared by NPS staff or contractors, program administrative documents, contracts and agreements, memoranda of understanding, and other documents related to NGPN activities. This chapter also applies to physical items such as natural history specimens, photographs, and audio tapes. In most instances these documents and objects are essential companions to the digital data described throughout this plan.

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 for transferring completed project deliverables to a secure and long-term storage facility managed by park and/or regional repositories.

10.1 NPS Standards

Direction for managing physical and 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) (<http://www.nps.gov/policy/DOrders/DOrder19.html>). Director's Order 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 on records related to natural resources (including, specifically, products 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 of 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 object collections.

10.2 Park Unit and Network Standards

Direction for managing materials for each park unit is described in each collection facilities “Scope of Collection Statement”. In general, only materials directly “... related to one or more of the park’s themes or site-related materials that the NPS is legally mandated to preserve” will be accepted. All of these materials are required to be cataloged in the NPS cataloging system (ANCS+, see 36 CFR, Section 2.5).

The NGPN will adhere to the standards described above. All specimens collected as part of NGPN projects are considered property of the respective park. Procedures for the collection and processing of specimens or samples will be dictated by curatorial staff at the park or regional level, as appropriate. In some instances, the parks may delegate the NGPN as curator for specimens that are central to the Network’s research.

10.3 Role of Curators in Curation and Museum Collection Storage

Curators and archival specialists for the region and NGPN parks are an excellent source of expertise, advice, and guidance on curatorial issues, and they have a role in almost every project undertaken by the Network. Project leaders should consult with curatorial staff early in the project planning stage to ensure that all aspects of museum curation of documents, specimens, and other objects are considered, and that any associated expenses are included in project budgets.

When there be field collections associated with a park research permit, an accession number is assigned by the park curator when the permit is issued. After collection, a range of catalog numbers can be assigned by the curator for the principal investigator to use. The accession and catalog number assignments must be requested by the project leader.

The project leaders will package all project materials and give the package to the park or its designated curator, as established in their collection permit. The curator will ensure that the materials are properly stored, cataloged, and archived. The project leader is ultimately responsible for verifying that cataloging and archiving has been completed.

Project managers are responsible for preparing materials for the curator. The workflow illustrated here and described in this chapter refers to the general procedure for making the transition between the project leaders and the museum staff (Figure 10.1). The museum staff will follow a much more detailed procedure for actual curation. The workflow schematic begins with the delivery of the products, walks through the electronic storage of these products, and then highlights the storage and archiving of materials (Figure 10.1).

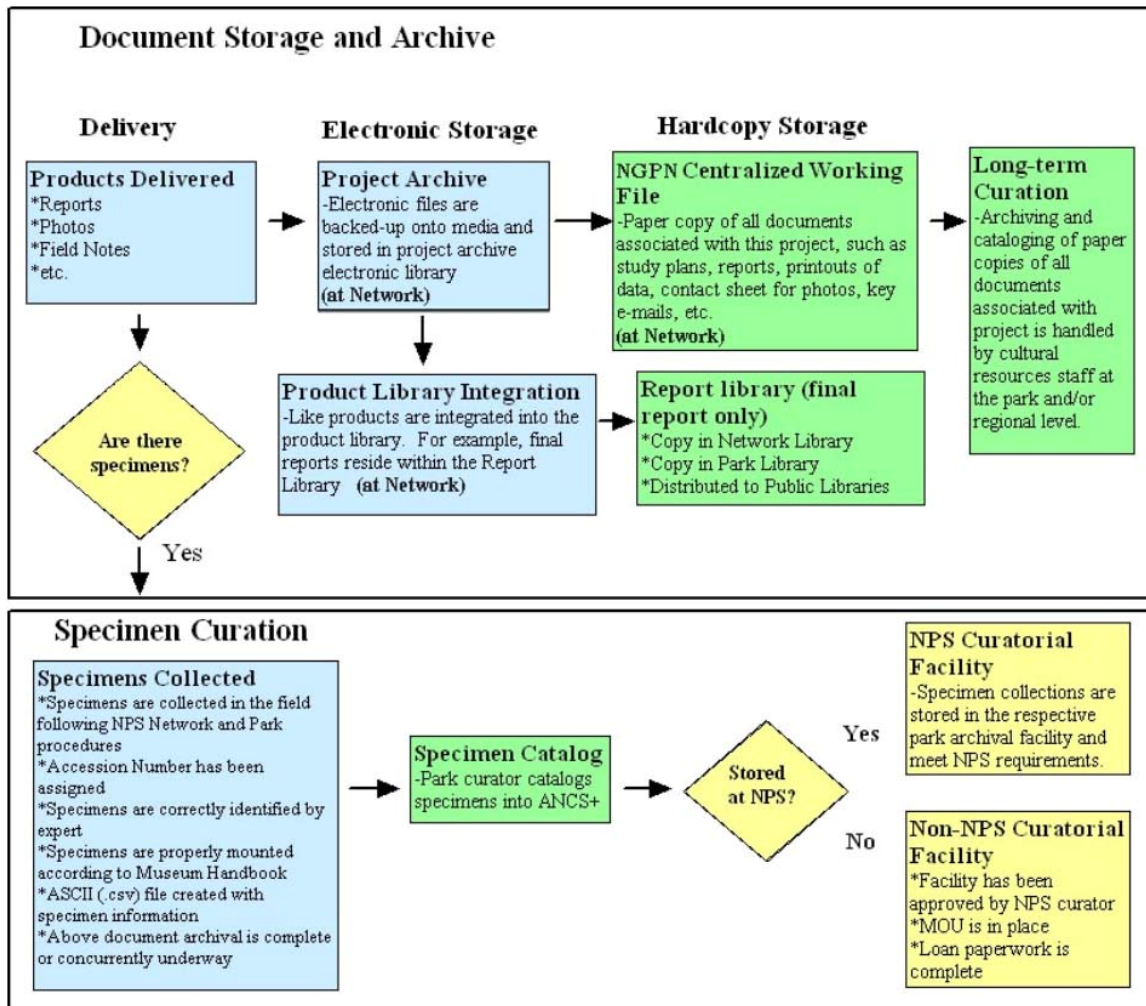


Figure 10.1 Schematic of natural history curation workflow for both hard copy documents and specimens

10.4 Network Archiving Process

Project materials should be consolidated and packaged for curation when a project is complete or, for long-term projects, reaches a milestone. The project leader should acquire an accession number from the respective curator.

The project leader should package the project information as follows:

- Include a cover sheet, memo, or Collection Catalog Worksheet listing contact information, project abstract and purpose, sensitivity and use of materials, and all materials included in the package.
- All materials should be clearly labeled with the:
 - Park acronym
 - Date or range of dates
 - Accession number
 - Project number
- Field notes should be on acid free paper (preferred) or copy paper and stored in a 3-ring binder or book box.

- Other paper materials such as reports and data printouts should be stored on acid free paper (preferred) or copy paper and placed individually in an acid free, labeled folder.
- Photographs should be printed and documented.
- Voucher specimens should be labeled according to the Museum Handbook
- For specimens not residing in park collection facilities, loan paperwork should be completed and a copy of the form should be stored with the project package.
- Specimens not residing in NPS repositories should be stored in a selected facility that meets NPS museum collection standards.
- Curation of specimens must be addressed in NGPN Vital Signs protocols and should specify:
 - The name and address of the storage facility
 - Memorandum of Understanding, if applicable
 - What is being curated or loaned (list of items)
- All electronic materials should be included on CD or DVD media. Jewel case should be labeled with the:
 - Park acronym
 - Date created
 - Range of dates for information
 - Accession Number
 - Project Number
 - List of contents

The museum staff will catalog this information into ANCS+ and will properly store the materials according to the Museum Handbook. If specified in arrangements with the Network and each affected park, similar packages should be prepared as “working copy” for the Network and the park(s).

10.5 Archive and Storage Locations

Many of the NGPN parks do not have adequate facilities to properly store and archive documents, photographs, artifacts, and specimens from all natural and cultural resource projects. Most NGPN parks actively store documents, but many look to other institutions for storage and archiving of artifacts and specimens. Storage of artifacts and specimens in non-NPS facilities presents challenges for park staff wishing to access materials. In contrast, an institution that has ample resources committed to an organized and well-maintained collections department serves scientists and researchers well because it provides a centralized location for materials from many collection sites.

The NGPN parks will continue to cooperate with various regional museums or curatorial facilities to archive artifacts and specimens collected in the parks.

NGPN Digital Data Backup and Storage

All digital data will be stored on the NGPN Server and backed up daily to tape, with offsite storage of a subset of the tapes to ensure redundancy in the event of a disaster. Offsite storage is currently arranged with Mount Rushmore NMEM, with which the NGPN enjoys a reciprocal arrangement. Mount Rushmore NMEM stores recent tape

backups for the NGPN, and the NGPN stores recent tape backups for Mount Rushmore NMEM, to ensure backup redundancy for both sites. Tape exchanges occur on a routine basis, at least several times monthly. The tapes are stored in a secure, access-restricted lockbox at both Mount Rushmore NMEM and NGPN.

10.6 NPS Automated National Catalog System (ANCS+)

The NPS Service-wide cataloging system is ANCS+ (using ReDiscovery software customized for NPS). Park staff with cataloging duties are typically the only individuals with access to this software. Park collections are entered into ANCS+ at the local level, then “rolled up” to the Service-wide database.

The cataloging system uses an accession number to identify a collection and catalog numbers to identify the items within the collection. If specimens were collected, the project will have a single accession number, but will contain multiple catalog numbers (one for each specimen collected). Park staff may access the catalog system by one of three methods:

- Contact a park curator and review together
- Request a copy of the program and data from the park curator
- Access the information online using ReDiscovery Web Catalog (<http://www.museum.nps.gov/>)

Uploading of a park’s ANCS+ records to the online ReDiscovery Web Catalog is provided at additional cost to the park. It should be noted that the online catalog may not represent all collection holdings at a given park.

10.7 Photographs

Museum curators have been reluctant to fully embrace digital photography and some have expressed concern that, with the accelerating rate of technological change, documentary heritage is in danger of being lost in the information age (Cox 2000). Photographic methods for any given project will be developed by the project manager and NGPN staff. The NGPN accepts both digital and analog photographs. Every image, regardless of format, will have an entry into a photographs database (ThumbsPlus or other NGPN-approved database) where attributes such as the electronic file name, keywords, project, photo description, photographer, date, and location are cataloged.

10.8 Specimens and Other Materials

Planning for monitoring projects includes budgeting for specimen preparation, records processing, cataloging, and storage. Specimens collected under the auspices of the NGPN will be provided to the respective park in which they were collected for curation, or to a repository approved by a park. Project managers are responsible for providing park curators with necessary data for cataloging each specimen. Network data management staff will assist project leaders and curators to ensure that specimens records are in comma-delimited format (.csv) for automated uploading into ANCS+. Data provided to non-NPS curators will be in an appropriate format specified by the institution.

All materials related to a project should be included in the collection. For materials (e.g., audio and visual) not described in this chapter, project managers should discuss archival requirements with the appropriate curatorial staff.

11. IMPLEMENTATION

As a data provider, our data management processes may be satisfactory and may meet our needs for a number of years. The I&M Program, however, focuses attention on our data consumers – managers who will make decisions based on the data we provide or researchers who will use these data decades from now. The central goal of Vital Signs monitoring is to understand the long-term patterns and processes of park ecosystems at multiple spatial scales. Good data management practices are essential to achieving this goal. It costs less to save data than to collect it again, and ecological data often cannot be collected again at any cost.

Some of the practices in this Data Management Plan may be new to staff and principal investigators. However, with few exceptions, the DMP does not include any new requirements. Almost every requirement comes from law, Director's Orders, or the I&M Program. The DMP helps to put these requirements into context and provides operational guidance to fulfill these requirements. The DMPs for each of the 32 I&M networks are the first comprehensive documents of their kind and will establish the means by which managers, educators, and the greater scientific community can reach data and information about network and park resources.

Good data management practices take time. Some Vital Sign collection procedures and data management practices are already in use and may require minimal revisions. Other data management practices may involve several iterations before the procedures and databases are acceptable and functional.

11.1 Common Data Management Sins

Data mining provides a means of discovering existing data and is an important component of all projects (Chapter 5). The Network has put considerable effort into mining data and information useful to the I&M Program. This consisted of looking through archives, office filing cabinets, natural resource folders, libraries, and electronic files. Not surprisingly, some data and information was well-maintained and documented, and some was not. Examples of common data management problems are:

- *Unfinished projects.* Failure to finish a project can happen for a number of reasons beyond the staff's control. Unfinished projects, however, are the most difficult to manage. It takes longer to document an incomplete project than a finished one. The best data management practice is to finish the project or reach a logical milestone.
- *Lack of header information.* Often, sheets of information do not have a date, name or subject. Reading the material and determining if the information is relevant is extremely time consuming. Relevant information is often discarded because it does not have adequate header information. A good data management practice is to use a memorandum format or a standard header for all files for a project, no matter how insignificant.
- *Lack of documentation.* Most projects and data are inadequately documented. Without documentation, the data are often considered unusable. Documentation

- *Lack of file naming standard and version control.* In many cases it was impossible to know which electronic file contained the final product.
- *No archive or final storage.* Data tucked away in an individual's desk or personal computer isn't accessible to anyone else. It often requires in-house knowledge to know that the information even exists. When there is a turnover in staff, this information is often lost.

11.2 Years 1-5 of Implementation

The NGPN is just beginning to implement this data management plan. The first few years of implementation will involve learning, testing, and refining. During the coming years, the data management plan will be implemented for the first Vital Signs monitoring protocols and data management training will be provided to Network and park staff. Templates and SOPs will be developed and tested, and will then be applied to other Vital Signs monitoring protocols. Some bottlenecks are anticipated; identifying and eliminating problems will be an important part of implementation.

Implementation Goals for Years 1-5:

The following outcomes should be achieved within the first five years of this Plan's implementation:

- All Network staff understand the fundamentals of data management, which include:
 - File management
 - Documentation
 - Quality assurance and quality control
 - Electronic storage
 - Archive storage
- Data management practices will be improved by implementing:
 - Conceptual Data Models for Vital Signs monitoring protocols
 - Testing of data entry prior to field work
- Common SOPs that can be used for multiple protocols will be developed
- Data management staff will inform and direct the data management aspects of the first Vital Signs monitoring protocols to be implemented

The Network plans to perform a data management plan review upon completion and implementation of the final Vital Signs monitoring protocol.

11.3 Years 5+ of Implementation

The NGPN data management plan will be revisited and revised every five years, beginning with the first review upon implementation of all Vital Signs monitoring protocols. Subsequent versions may be more streamlined and concise. With prior experience and commonplace practices, generalizations may be eliminated.

Implementation Goals for Years 5+:

- The data management plan will be revised and streamlined for Vital Signs monitoring.
- All implemented Vital Signs monitoring protocols will follow the data management plan guidance.
- Databases and custom reporting will continue to be enhanced.
- Framework and gateway for integration of data with other agencies or networks will be established.

LITERATURE CITED

Cox, R.J. 2000. Searching for authority: archivists and electronic records in the New World at the fin-de-siècle. *First Monday* 5:1.

(http://firstmonday.org/issues/issue5_1/cox/index.html).

Edwards, D. 2000. Data quality assurance. Pages 70-91 *in* Michener, W.K., and J.W. Brunt, editors, 2000, *Ecological data: design, management and processing*. Blackwell Science Ltd. Methods in Ecology Series, Malden, MA.

Palmer, C.J., and E.B. Landis. 2002 Draft. Lake Mead National Recreational Area Resource Management Division: Quality System Management Plan for Environmental Data Collection Projects.

Southwold, A. 2002. Recommended Database Strategies, including I&M Database Template. Version 1.0. Draft. Unpublished manuscript.

Tukey, John Wilder (1977). *Exploratory Data Analysis*. Addison-Wesley.

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

APPENDIX A: BASIC RESOURCE INVENTORIES AND REPOSITORIES FOR THE NGPN

Inventory	Information Access
Natural resource bibliography	NatureBib (www.nature.nps.gov/nrbib)
Base cartographic data	http://science.nature.nps.gov/nrdata/
Geology map	NPS Geologic Resources (www2.nature.nps.gov/geology/inventory/gre_pub2.htm)
Soils map	http://websoilsurvey.nrcs.usda.gov/app/WebSoilSurvey.aspx
Meteorological data	Midwestern Regional Climate Center (http://mrcc.sws.uiuc.edu)
Air quality	NPS Air Resources (www2.nature.nps.gov/air/)
Air quality related values	NPS Air Resources (www2.nature.nps.gov/air/)
Water body location and classification	NPS Water Resources (www.nature.nps.gov/water)
Water quality data	NPS Water Resources (www.nature.nps.gov/water)
Documented species occurrence of vertebrates and vascular plants	NPSpecies (science1.nature.nps.gov/npspecies)
Species distribution and status of vertebrates and vascular plants	NPSpecies (science1.nature.nps.gov/npspecies)

APPENDIX B: PROJECT INFORMATION MANAGEMENT PROCESS & WORK FLOW: DELIVERABLES, REPOSITORIES, AND GUIDANCE DOCUMENT REFERENCES

Project phase	Deliverables	Repository	Data management guidance
Planning and approval	- proposal (optional) - study plan, work plan - permit application and permit - contracts and agreements - new record in project management database	- park project funding requests in PMIS/RAMS - permit applications in RPRS - network project management database - working project files, then document archives - final copies of study plan sent to document archives and/or placed in the digital library; records created in NatureBib	- guidelines for data distribution and data discovery - guidelines for organizing project info (including a project file organizer and a template directory structure) - project information management SOP - template language for contracts and agreements (protected information, deliverables) - specifications for deliverables (includes reference to GIS specifications)
Design and testing	protocol / methodology and SOPs	documents and supporting materials created and	naming conventions for

Project phase	Deliverables	Repository	Data management guidance
	• blank field forms and data dictionary • data design documentation • metadata questionnaire – partially complete • Project file documentation – partially complete	• maintained in working project folders • final copies of protocols/SOPs sent to document archives and/or placed in the network digital library; records created in NatureBib	• files and data objects • data design standards • SOP - metadata procedures and specifications • guidelines for quality assurance and quality control • SOP – GPS data collection and processing
Implementation	• contracts and agreements • raw data forms, field notebooks, trip reports • databases, GIS layers, GPS rover files, list of coordinates • biological specimens, species checklists, catalog of specimen data • photographs, illustrations, sound and video recordings • data certification report	• documents, databases, GPS rover files and GIS layers created and maintained in working project folders • specimens to the park collection or other designated repository	• guidelines for managing third-party data • guidelines for labeling and storage of analytical samples and specimens • standards for voucher collection and cataloguing • photo management strategy • template language for contracts and agreements • naming

Project phase	Deliverables	Repository	Data management guidance
			conventions for files and data objects • SOP - metadata procedures and specifications • specifications for deliverables (including GIS specs) • guidelines for quality assurance and quality control • SOP – GPS data collection and processing
Product delivery and review	• annual report • trend analysis report • final reports - technical or general audiences • publications • completed metadata questionnaire • Project file documentation	• documents and supporting materials created and maintained in working project folders • final copies of reports sent to document archives and/or placed in the network digital library • permit investigator's annual reports to RPRS	• reporting guidelines and templates • specifications for deliverables (including GIS specs) • SOP - metadata procedures and specifications
Product integration	• finalized metadata posted on the internet • records in searchable corporate databases (e.g., NPSpecies, NatureBib) • products are secure and available	• NPS metadata clearinghouse • NatureBib • NPSpecies • network databases • all remaining materials moved to document archives and stored by calendar year	• integration of project data with corporate data • archival procedures for project records • sensitive info

Project phase	Deliverables	Repository	Data management guidance
			management strategy • NPSpecies and NatureBib guidelines • project information management SOP • guidelines for data distribution and discovery
Project evaluation and close out	• update record in project management database • completed checklists (data manager, permitting officer, project manager) • documentation of needed modifications	• network project management database • place completed checklist in document archives • maintain modifications documentation with related project documents	• project information management SOP

APPENDIX C: ACRONYMS

ANCS+	Automated National Catalog System
CDM	Conceptual Data Model
CSDGM	Concept Standards for Digital Geospatial Metadata
DMP	Data Management Plan
EPA	U.S. Environmental Protection Agency
ERD	Entity Relationship Diagram (for databases)
ESRI	Environmental Systems Research institute
FGDC	Federal Geographic Data Committee
FOIA	Freedom of Information Act
FTP	File Transfer Protocol
GIS	Geographic Information System
NGPN	Northern Great Plains Inventory & Monitoring Network
GPS	Geographic Positioning System
HTML	Hypertext Markup Language
I&M	Inventory & Monitoring (Program)
IT	Information Technology
ITIS	Integrated Taxonomic Information System
JPEG	Joint Photographic Experts Group
LAN	Local Area Network
LDM	Logical Data Model
MOU	Memorandum of Understanding
mp	Metadata Parser
MS	Microsoft
MWR (MWRO)	Midwest Region(al Office)
NBII	National Biological Inventory Infrastructure
NOAA	National Oceanic and Atmospheric Administration
NPS	National Park Service
NRDT	Natural Resource Database Template
NRTS	Natural Resource Table Standards
OMB	Office of Management and Budget
PC	Personal Computer
PDA	Portable Digital Assistant
PDF	Adobe Portable Document Format
PDM	Physical Data Model
QA/QC	Quality Assurance/Quality Control
RDMS/RDBMS	Relational Database Management System
SOP	Standard Operating Procedure
SQL	Structured Query Language
TIFF	Tagged Image File Format
USGS	U.S. Geological Survey
WAN	Wide Area Network
WASO	Washington Support Office
XML	Extensible Markup Language

APPENDIX D: GLOSSARY

Conceptual data model is a detailed model that shows the overall structure of organizational data while being independent of any database management system or other implementation considerations. It is typically produced as the first modeling step in system design, and it is frequently a precursor to the logical data model. Conceptual data models are often presented as entity relationship diagrams.

Data refers to observational information gathered directly (in the field) from a natural resource via specific protocols and organized for analysis, summary or reporting. Data may be in either analog or digital form (generally stored either on paper or a variety of computer-compatible media), though the latter is encouraged where ever feasible. Data may exist in several states (conditions) including ‘raw’, ‘validated’ and ‘analyzed’. ‘Analyzed’ data includes ‘reported’ or ‘summarized’ data and may represent ‘information’ as a final form of the data from which decisions or conclusions may be made. Ultimately, data are intended to contribute to the knowledge and decisions regarding the conditions, processes, and changes within the ecosystem.

Data set can best be considered a convenient grouping of data, or individual observations, such that the summary of the information will be meaningful to prospective users.

Entity relationship diagrams (ERDs) are high-level data models that are useful in developing conceptual designs for databases. Creation of an ER diagram, which is one of the first steps in designing a database, helps the designer(s) to understand and to specify the desired components of the database and the relationships among those components. An ER model is a diagram containing entities or "items", relationships among them, and attributes of the entities and the relationships.

Indicators are a subset of monitoring attributes that are particularly information-rich in the sense that their values are somehow indicative of the quality, health, or integrity of the larger ecological system to which they belong. Indicators are a selected subset of the physical, chemical, and biological elements and processes of natural systems that are selected to represent the overall health or condition of the system.

Inventories are “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>)

Legacy Data are data collected before the NGPN Data Management Plan was implemented. Generally, this refers to data collected in the park units that the NGPN deems important to the monitoring mission.

Logical data model (LDM) is an abstract representation of a set of data entities and their relationship, usually including their key attributes. It may leave out non-key attributes or use abstractions of actual types, and there may be physical tables that are represented in the logical model by relationships or as aggregates within a larger entity, depending on the notation used and the level of abstraction. The logical data model is intended to facilitate analysis of the function of the data design, and is not intended to be a full representation of the physical database. It is typically produced early in system design, and it is frequently a precursor to the physical data model that documents the actual implementation of the database.

Metadata is data about data. It is a document that contains specific and detailed information about a data set, including who, what where, when, why and how the data was collected, analyzed, or manipulated.

NatureBib is the National Park Service bibliographic database, and serves to help staff learn of relevant reports, locate them, and (where possible) obtain digitally.

Natural Resource Database Template (NRDT) is a relational database template that provides a basic underlying relational structure for developing databases. It includes standards for describing location and event information, to which protocol-specific tables and relationships can be added.

Ortho-rectification is the process of correcting an image for distortions due to camera lens, sensor non-verticality, and terrain, a process collectively included in photogrammetry.

Orthophoto is a digital image in which the pixels have been corrected to an orthogonal projection, produced by removing errors due to tilt and relief displacement.

Physical data model is a representation of a data design which takes into account the facilities and constraints of a given database management system. In the lifecycle of a project it is typically derived from a logical data model, though it may be reverse-engineered from a given database implementation. A complete physical data model will include all the database components required to create relationships between tables or achieve performance goals, such as indexes, constraint definitions, linking tables, partitioned tables or clusters.

Protocol, (Monitoring Protocol), refers to the formal documents and sampling processes describing how a Vital Sign will be monitored; composed of a narrative section, standard operating procedures and supplementary information (databases, reports, tools, hardcopy materials).

Raw Data are data obtained from the environment and that has not been subjected to any quality assurance or control beyond those applied during field work. Typically, raw data

constitutes field data sheets but may also include remotely sensed data, data gathered electronically on field computers and photographic imagery.

Relational Database System (RDMS/RDBMS) is a type of database management system (DBMS) that stores data in the form of related tables. Relational databases are powerful because they require few assumptions about how data is related or how it will be extracted from the database. As a result, the same database can be viewed in many different ways. An important feature of relational systems is that a single database can be spread across several tables. This differs from flat-file databases, in which each database is self-contained in a single table.

Sensitive Data are data that is not intended for release to the public, and are not subject to FOIA requests. These data are commonly include information about threatened and endangered species, or archaeological resources, whereby disclosure of such information could result in harm or vandalism.

Standard Operating Procedures (SOPs) are detailed step-by-step instructions for carrying out sampling procedures in monitoring Protocols (the ‘How?’ of data management planning); dynamic in nature and frequently updated; can reference overall data management plan and standards.

STORET/NPSTORET is a database application maintained by the U.S. Environmental Protection Agency that contains raw biological, chemical and physical data on surface and ground water collected by federal, state and local agencies, Indian Tribes, volunteer groups, academics, and others. All 50 states, and jurisdictions of the U.S. are represented in these systems. NPSTORET is the NPS version of this database used for data entry by NPS staff. Data from NPSTORET are subsequently transferred to STORET.

Structured Query Language (SQL) is a language used to interrogate and process data in a relational database. Originally developed for mainframes, all database systems designed for client/server environments support SQL. SQL commands can be used to interactively work with a database or can be embedded within a programming language to interface to a database. Programming extensions to SQL have turned it into a full-blown database programming language, and all major database management systems (RDBMS s) support the language. There is an ANSI standardized SQL, but most database management systems (RDBMS s) have some proprietary enhancement, which if used, makes SQL non-standard.

Validated data are data that have been verified according to the standard operating procedure under which the data were gathered (typically the protocol for a given project) and are deemed ready for reporting and/or analysis.

Vital Signs, as used by the National Park Service, are a subset of physical, chemical, and biological elements and processes of park ecosystems that are selected to represent the overall health or condition of park resources, known or hypothesized effects of stressors, or elements that have important human values. The elements and processes that are

monitored are a subset of the total suite of natural resources that park managers are directed to preserve "unimpaired for future generations," including water, air, geological resources, plants and animals, and the various ecological, biological, and physical processes that act on those resources. Vital Signs may occur at any level of organization including landscape, community, population, or genetic level, and may be compositional (referring to the variety of elements in the system), structural (referring to the organization or pattern of the system), or functional (referring to ecological processes).