



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.

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.