



The National Park Service
Inventory and Monitoring

Data Management Plan

Gulf Coast Network

2007



Version: 1.1

Status: Draft

Author: Whitney Granger

Table of Contents

_Toc167086834

Executive Summary	1
1. Introduction	7
1.1. The National Park Service Vital Signs Monitoring Program.....	7
1.2. Scope of the Data Management Plan	8
1.3. Data Management Goals.....	9
1.4. Audience	9
1.5. Revisions to the Data Management Plan.....	9
1.6. Types of Data Covered by this Plan.....	9
1.7. Sources and Priorities of Natural Resource Data.....	11
1.8. Organization of this Data Management Plan.....	11
1.9. Chapter Credits	12
2. Data Management Roles and Responsibilities	13
2.1. Data Stewardship Roles and Responsibilities	14
2.2. Data Management Coordination.....	15
2.3. Chapter Credits	19
3. Data Management Process and Workflow.....	20
3.1 Project Work Flow	20
3.2 Data Life Cycle	26
3.3 Integrating and Sharing Data Products	28
3.4. Review Cycle	31
3.5. Chapter Credits	31
4. Data Management Infrastructure and System Architecture.....	32
4.1 Computer Resources Infrastructure.....	33
4.2 National Information Management Systems	33
4.3 GULN Information Management System Architecture	37
4.4 Chapter Credits	40
5. Database Design Strategy	41
5.1. NPS and Program Standards	41
5.2. Introduction to Data Models	41
5.3. Fundamental Database Structure	45
5.4. Documenting and Integrating Databases	45
5.5. Chapter Credits	46
6. Data Acquisition and Processing.....	47
6.1. National Standards.....	47
6.2. GULN Network Standards	47
6.3. Data Sources	47
6.4. Data Discovery	48
6.5. Data Harvesting.....	48
6.6. Data Collection for Projects	49
6.7. Data Collection for Remote Sensing and LIDAR.....	51
6.8. Changes to Data Collection Procedures	52
6.9. Data Compilation, Processing, and Integration.....	53
6.10. Chapter Credits	53
7. Quality Assurance and Quality Control.....	54
7.1. NPS Standards	54
7.2. Quality Assurance and Quality Control Mechanisms.....	54
7.3. Roles and Responsibilities.....	56

7.4. Quality Assurance/Quality Control Goals and Objectives	56
7.5. Data Collection Quality Assurance/Quality Control.....	57
7.6. Data Entry	59
7.7. Verification and Validation Procedures	61
7.8. Version Control	64
7.9. Data Quality Review and Communication.....	64
7.10. Chapter Credits	66
8. Data Documentation.....	67
8.1. NPS Integrated Metadata System Plan and Tools	67
8.2. Metadata Process and Work flow	68
8.3. Vital Signs Protocol Documentation	69
8.4. Chapter Credits	70
9. Data Analysis and Reporting.....	71
9.1. Periodic and Annual Reporting	71
9.2. Long-term Trends and Analyses	71
10. Data Dissemination.....	72
10.1. Mechanisms for Distribution.....	72
10.2. Ownership, FOIA, and Sensitive Data.....	73
10.3. Feedback Mechanisms	76
10.4. Chapter Credits	76
11. Data Maintenance, Storage, and Archiving	77
11.1. Digital Data Maintenance	77
11.2. Storage and Archiving Procedures – digital data.....	80
11.3. Storage and Archiving Procedures – documents and objects	82
11.4. Role of curators in storage and archiving procedures	83
12. Implementation.....	84
12.1. Education and Training	84
12.2. Milestone Goals.....	84
12.3. Revisions.....	85
Literature Cited.....	86

List of Figures

Figure 1.1 Data Management Plan Overview	8
Figure 2.1 Core data management responsibilities for project managers and Network data managers.	16
Figure 3.1 Generalized project work flow model for short- and long-term data collection efforts. Bold face activities indicate core data management elements.	21
Figure 3.2 Diagram of the typical project data life cycle.	27
Figure 3.3 Storing and disseminating project information.....	29
Figure 3.4 Steps involved in product distribution.	30
Figure 3.5 Data flow diagram for water quality data.	31
Figure 4.1 Simplified comparison of infrastructure and system architecture.	32
Figure 4.2. Model of the national-level application architecture.	34
Figure 4.3 Common lookup tables and satellite databases	39
Figure 5.1 Schematic of Conceptual Data Model.	Error! Bookmark not defined.
Figure 5.2 Schematic of Logical Data Model.	44
Figure 5.3 Schematic of Physical Data Model.....	44
Figure 7.1 Schematic of the Quality Assurance/Quality Control procedures to be carried out during the project stages associated with the typical data life cycle.....	55

List of Tables

Table 1.1 Categories and examples of data products covered by the GULN DMP.....	10
Table 2.1 Categories of data stewardship involving all Network personnel and cooperators.	13
Table 2.2. Programmatic roles and associated data stewardship responsibilities.....	14
Table 3.1 List of tasks per stage and staff with primary and secondary responsibility. PL = Project manager; EC = Ecologist; DM = Data Manager; GIS = GIS specialist. * = Primary role; - = Secondary role.	22
Table 3.2 Repositories for GULN products.	29
Table 4.1 Groupings for common lookup tables.....	Error! Bookmark not defined.
Table 6.1 Satellite resolution, swath width, area, cost, and history of various remote-sensing imagery sources.	52
Table 8.1 Example of the Master Version Table used to track changes in a vital sign protocol narrative and/or standard operating procedures (SOP).	69
Table 10.1 Primary repositories for GULN information and associated specimens.....	73
Table 11.1 Backup schedule for GULN servers.....	81

List of Appendices

Appendix A- List of Network Computer Hardware

Appendix B- Standard Operating Procedures for Reference Manager

Appendix C- Final Product Specifications

Appendix D- GULN Photo Management Strategy

Appendix E- File Naming Conventions

Change History

Titles of updated or revised sections of the plan will be followed by brackets containing the date of the latest version of that section and a brief update history. For example, the title of Section 2.2 may be followed by “[06/02/2006; Update History: 1/15/2005, 10/23/2004]”. This indicates the current wording of that section was last updated on June 2, 2006 and it was previously revised on January 15, 2005 and October 23, 2004.

The following revisions have occurred:

Original Version #	Date of Revision	Revised By	Changes	Justification	New Version #
1.0	12/21/2006	Whitney Granger	Draft version	Addressed review comments	1.1

Executive Summary

Information is the common currency among the activities and staff involved in the stewardship of natural resources for the National Park Service (NPS). This Executive Summary chapter summarizes the GULN data management strategy. The Data Management Plan is a guide for current and future project leaders and GULN staff to ensure the continuity and documentation of data management methods and procedures over time. The DMP, in turn, refers to other guidance documents and standard operating procedures which convey the specific standards and steps for achieving the network's data management goals.

The Data Management Plan (DMP) focuses on the processes used to:

- Acquire, store, manage and archive data
- Ensure data quality
- Document and disseminate data
- Ensure the long-term access to and utility of data.

1. Introduction

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 our natural resource data.

The Data Management Plan goals are to ensure that:

- **Data managed by the network are of high quality**, including designing standardized data entry, importation, and handling procedures which effectively screen for inappropriate data and minimize transcription and translation errors;
- **Network data can be easily interpreted**, by considering the users' needs as the primary factor driving the design of summary reports and analyses; establishing rigorous data documentation standards; integrating common data tables and fields in NPS or regional standards; and making summary information available in formats tailored to the variety of audiences interested in I&M program results;
- **Data are secure for the long term**, including instituting standard procedures for versioning, data storage and archiving; and natural history archiving, curation and records management are provided to NPS curators;
- **Network data are readily available**, by implementing standard procedures for distributing data, while protecting sensitive data; and designing a standardized filing system for organizing I&M information;

2. Data Stewardship Roles and Responsibilities

Every individual involved in the I&M Program is required to understand and perform data stewardship responsibilities in the production, analysis, management, and end use of the data as described in the Data Management Plan and the specific vital signs monitoring protocols. Monitoring protocols will describe specific roles and responsibilities in detail.

Senior network staff shares responsibility in ensuring that data management procedures are followed.

3. Project Work Flow

Understanding the life cycle of data throughout a project will help to manage the staffing resources necessary to complete and support quality data. For data management to be effective, it must occur throughout the project life cycle.

A project is divided into the following stages:

1. Project Initiation
2. Planning and Approval
3. Design and Testing
4. Implementation
5. Product Delivery
6. Product Integration
7. Closure and Evaluation

GULN uses a Project Tracking database to document and support the progress of projects that collected information for vital signs monitoring. Most notably, this database tracks the status, changes, archival and distribution of deliverables.

4. Data Management Infrastructure and System Architecture

A modern information management infrastructure (e.g., staffing, hardware, software) represents the foundation upon which our network information system is built. Systems architecture refers to the applications, database systems, repositories, and software tools that make up the framework of our data management enterprise.

An important element of a data management system is a reliable, secure network of computers and servers maintained by national and local offsite IT specialists, assisted by the network data manager. These individuals attend to hardware replacement, software installation and support, security updates, virus-protection, telecommunications networking, and server backups. Our digital infrastructure consists of network data and GIS servers and servers maintained at the national level Each of these components hosts different parts of our natural resource information system.

The infrastructure needs to support these required functions:

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

5. Database Design Strategies

The project leader and the data manager will work together to develop conceptual data models to:

1. Understand conceptually the data life cycle flow of the data collection process. Where is the starting point of the data collection (for example, a visit to a site) and what happens next.
2. Determine the data relationships as the implementation progresses. For example, one site visited many times with many collections.
3. Determine how the information will be presented. Understanding the relationships between the data components collected is the key to the success of a database and its utility. If the relationships are misunderstood, the database may become tedious in data entry and cumbersome at data output.

The GULN Data Management Plan specifies the standards by which data will be handled. Data management elements or principles common to more than one vital sign will be managed in a conventional manner to allow for greater comparison of data across the network, as well as to ensure further general data integrity.

6. Acquiring and Processing Data

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

- Program data – 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 – produced by NPS entities without the involvement of the I&M Program (e.g., park or regional projects)
- Non-program external data – produced by agencies or institutions other than the NPS (e.g., weather and water quality data)

Most data acquired by the network will be collected as field data (inventories and monitoring studies) or discovered through data mining initiatives legacy/existing data). Methods of field data collection, such as paper filed data forms, field computers, automated data loggers, and GPS units will be specified in individual monitoring protocols and study plans. Field crew members will closely follow the established standard operating procedures (SOPs) in the project protocol. Data acquired by non-program sources, such as data downloaded from other agencies, will also be specified in individual monitoring protocols.

7. 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 products conform to data quality standards.

Although many quality assurance/quality control (QA/QC) procedures depend upon the individual vital signs being monitored, some general concepts apply to all. Specific procedures to ensure data quality must be included in the protocols for each vital sign. Examples of QA/QC practices include:

- Field crew training
- Standardized field data sheets with descriptive data dictionaries
- Use of handheld computers and data loggers

- Equipment maintenance and calibration
- Procedures for handling data in the field
- Database features to minimize transcription errors, including imports from data loggers, range limit, pick lists, etc.
- Verification and validation, including automated error-checking database routines

Quality assurance methods should be in place at the inception of any project and continue through all project stages to final archiving of the dataset. It is critical that each member of the team work to ensure data quality. The final step in project quality assurance is the preparation of summary documentation that assesses the overall data quality. A statement of data quality will be composed by the Project Leader and incorporated into formal metadata. Metadata for each dataset will also provide information on the specific quality assurance procedures applied and the results of the review.

8. Data Documentation

Documenting datasets, data sources, and methodology by which the data were acquired establishes the basis for interpreting and appropriately using data. At a minimum, all data managed by the network will require the following elements of documentation:

- Project documentation
- Formal metadata compliant with Federal Geographic Data Committee (FGDC) standards
- Data dictionaries and Entity Relationship Diagrams (ERDs) for all tabular databases

9. Data Analysis and Reporting

Providing meaningful results from data summary and analysis is a cornerstone of the I&M Program and characterizes the network's data management mission to provide useful information for managers and scientists. Each monitoring protocol establishes requirements for on-demand and scheduled data analysis and reporting. Based on these requirements, the associated databases for the protocols include functions to summarize and report directly from the database as well as output formats for import 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. Chapter 7 of the GULN Monitoring Plan provides more details regarding the network's analysis and reporting schedule and procedures.

10. Data Dissemination

The GULN data dissemination strategy aims to ensure that:

- Data are easily discoverable and obtainable
- 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

Access to GULN data products will be facilitated via a variety of means that allow users to browse, search and acquire network data and supporting documents. These means include, but are not limited to:

- Links to public data products will be maintained on the GULN public website
- NPS Data Store. Distribution instructions for each dataset 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
- FTP sites, CDs, DVDs, or hard drives, as appropriate

11. Data Maintenance, Storage, and Archiving

To ensure high-quality long-term management and maintenance of this information, the Network will implement procedures to protect information over time. These procedures will permit a broad range of users to easily obtain, share, and properly interpret both active and archived information, and they will ensure that digital and analog data and information are:

- Kept up-to-date in content and format so they remain easily accessible and usable
- Protected from catastrophic events (e.g., fire and flood), user error, hardware failure, software failure or corruption, security breaches, and vandalism

Technological obsolescence is a significant cause of information loss, and data can quickly become inaccessible to users if they are stored in out-of-date software programs, on outmoded media, or on deteriorating (aging) media. Effective maintenance of digital files depends on the proper management of a continuously changing infrastructure of hardware, software, file formats, and storage media. Major changes in hardware can be expected every 1-2 years and in software every 1-5 years. As software and hardware evolve, data sets must be consistently migrated to new platforms or saved in formats that are independent of specific software or platforms (e.g., ASCII delimited text files). Storage media should be refreshed (i.e., copying data sets to new media) on a regular basis, depending upon the life expectancy of the media.

Regular backups of data and off-site storage of backup sets are the most important safeguards against data loss; therefore, we have established data maintenance and backup schedules for data stored on the network data servers. Backups of data stored on personal workstations are the responsibility of each staff member. We strongly recommend that staff members store or regularly copy important files onto the network server. Backup routines represent a significant investment in hardware, media, and staff time; however, they are just a small percentage of the overall investment that we make in Program data.

12. Implementation

The GULN Data Management Plan (DMP) contains practices that may be new to staff and principal investigators. With a few exceptions, however, the DMP does not include any requirements that are new. Almost every requirement comes from law, Director's Orders, or the I&M Program. The DMP helps to put these requirements into context and in sequence, provides operational guidance for achieving these requirements, and outlines short- and long-term goals.

