National Park Service
U.S. Department of the Interior

Natural Resource Program Center

Gulf Coast Network Data Management Plan
Version 1.2

Natural Resource Report NPS/GULN/NRR–2007/00X

ON THE COVER
Top: (1) Great blue heron (*Ardea herodias*), Padre Island National Seashore. (2) Naval Live Oaks Reservation, Gulf Islands National Seashore. (3) Texas tortoise (*Gopherus berlandieri*), found in San Antonio Missions National Historical Park and Palo Alto Battlefield National Historic Site. (4) Natchez Trace Parkway (NPS photo).

Middle: (1) Dunes in November, Padre Island National Seashore. (2) Large ranid frog, Jean Lafitte National Historical Park and Preserve. (3) Vicksburg National Military Park.

Bottom: (1) Big Thicket National Preserve (NPS photo). (2) Acequia at San Antonio Missions National Historical Park.

All photos in this document are courtesy NPS/R. Woodman unless otherwise noted.

Gulf Coast Network Data Management Plan
Version 1.1

Natural Resource Report NPS/GULN/NRR–2007/00X

Whitney Granger, Science Information Specialist

National Park Service
646 Cajundome Boulevard
Room 175
Lafayette, Louisiana 70506

August 2007

U.S. Department of the Interior
National Park Service
Natural Resource Program Center
Fort Collins, Colorado

The Natural Resource Publication series addresses natural resource topics that are of interest and applicability to a broad readership in the National Park Service and to others in the management of natural resources, including the scientific community, the public, and the NPS conservation and environmental constituencies. Manuscripts are peer-reviewed to ensure that the information is scientifically credible, technically accurate, appropriately written for the intended audience, and is designed and published in a professional manner.

Natural Resource Reports are the designated medium for disseminating high priority, current natural resource management information with managerial application. The series targets a general, diverse audience, and may contain NPS policy considerations or address sensitive issues of management applicability. Examples of the diverse array of reports published in this series include vital signs monitoring plans; monitoring protocols; "how to" resource management papers; proceedings of resource management workshops or conferences; annual reports of resource programs or divisions of the Natural Resource Program Center; resource action plans; fact sheets; and regularly-published newsletters.

Printed copies of reports in these series may be produced in a limited quantity and they are only available as long as the supply lasts. This report is also available from the Natural Resource Publications Management website (http://www.nature.nps.gov/publications/NRPM) on the Internet or by sending a request to the address on the back cover.

Please cite this publication as:
Granger, W. 2007. Gulf Coast Network Data Management Plan. Natural Resource Report NPS/GULN/NRR–2007/00X. National Park Service, Fort Collins, Colorado.

NPS D-XXX, August 2007

Contents

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
1.1	12/11/20007	Whitney Granger	Add SOP Addressing Document Revisions	Create consistency for handling revisions to monitoring documents	1.2

Figures

Tables

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:

Project Initiation
Planning and Approval
Design and Testing
Implementation
Product Delivery
Product Integration
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:
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.
Determine the data relationships as the implementation progresses. For example, one site visited many times with many collections.
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.

1. Introduction

1.1. The Gulf Coast Network and the National Park Service Vital Signs Monitoring Program

The Gulf Coast Network (GULN) is one of 32 networks included in the Service wide Inventory and Monitoring program. The network approach facilitates collaboration, information sharing, and economies of scale in natural resource monitoring, and will provide parks with a foundational infrastructure for initiating natural resource monitoring that can be built upon in the future. The GULN (Figure 1.1) is located in portions of six states; spans from Brownsville, TX to Pensacola, FL, and north to Nashville, TN; and includes eight units of the National Park system: Big Thicket National Preserve (BITH), Gulf Islands National Seashore (GUIS), Jean Lafitte National Historical Park and Preserve (JELA), Natchez Trace Parkway (NATR), Palo Alto Battlefield National Historic Site (PAAL), Padre Island National Seashore (PAIS), San Antonio Missions National Historical Park (SAAN), and Vicksburg National Military Park (VICK).

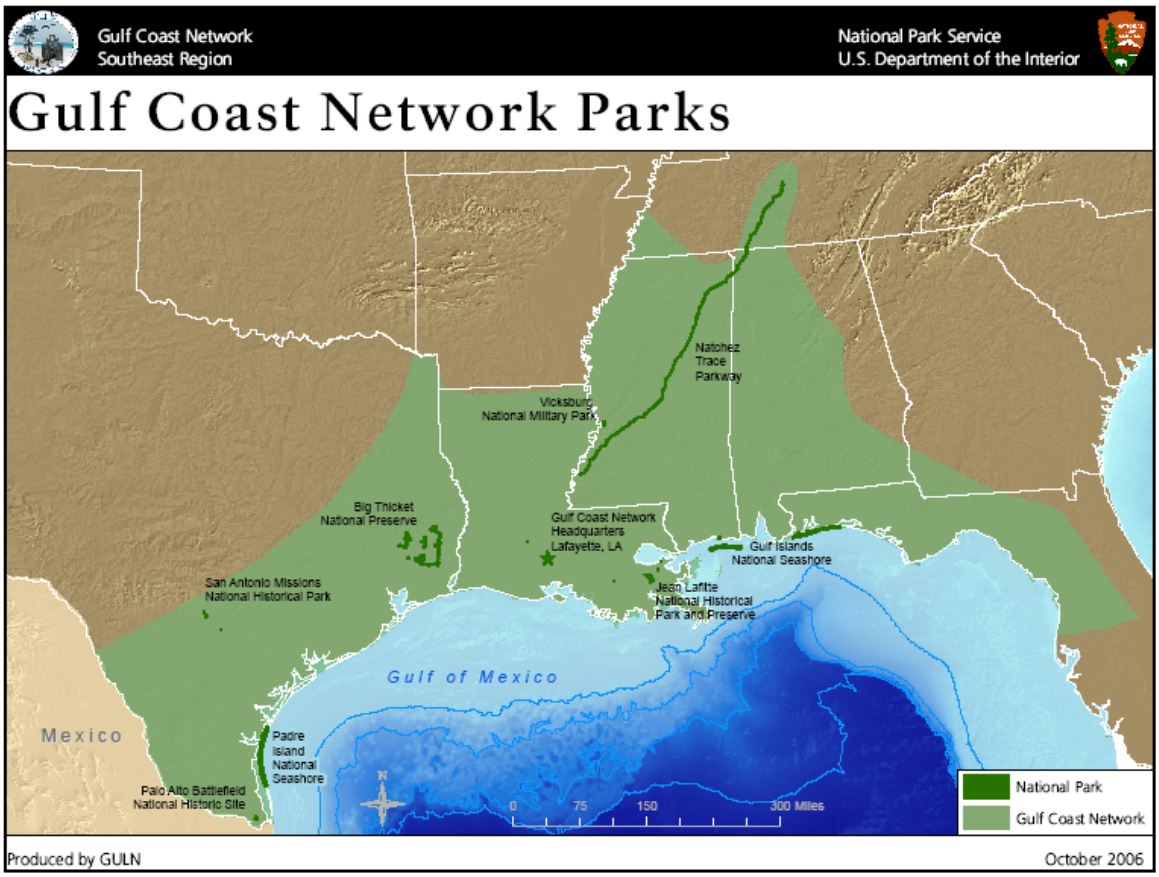


Figure 1.1. The Gulf Coast Network.

The Inventory and Monitoring (I&M) Program represents a long-term commitment by the National Park Service (NPS) to assess and document the status and trends of park resources. In 1998, the National Parks Omnibus Management Act established a framework for the I&M Program, which fully integrates natural resource monitoring and other scientific activities into the management processes of the National Park system.

The I&M Program's long-term data management goals are to:

1. Establish natural resource inventory and monitoring standards throughout the NPS that transcend traditional program, activity, and funding boundaries.
2. Inventory the natural resources under the NPS Stewardship to determine their nature and status.
3. Monitor park ecosystems to provide reference points for comparisons with other, altered environments.
4. Integrate natural resource inventory and monitoring information into NPS planning, management, and decision making.
5. Share NPS accomplishments and information with other natural resource organizations and form partnerships for attaining common goals and objectives.

To achieve the last two of these goals, a modern information management infrastructure must be developed. This infrastructure will include procedures to ensure that relevant natural resource data collected will be entered, validated, analyzed, reported, documented, cataloged, archived, and made available to others for management decision-making, research, and education.

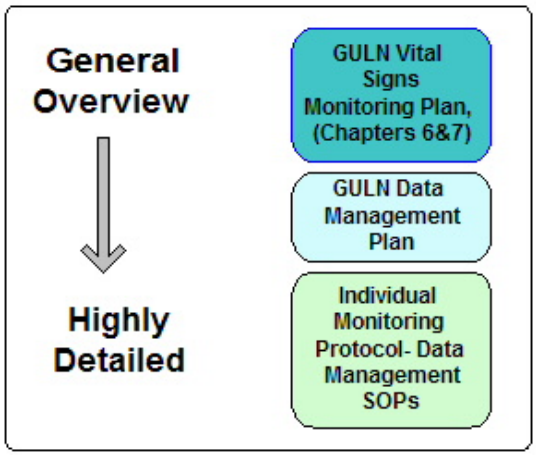


Figure 1.2. Data Management Plan overview.

1.2. Scope of the Data Management Plan

The Data Management Plan (DMP) is a guide for the Gulf Coast Network I&M Program to ensure the continuity and documentation of data management methods and procedures. It may provide a model for parks to follow with other projects, but its primary focus is large integrated datasets developed by the network. The DMP focuses on the processes used to:

1. Acquire, store, manage and archive data

Ensure data quality

2. Document and disseminate data

3. Ensure the long-term access to and utility of data.

Data management is currently being addressed at three levels of detail (Figure 1.2). Chapter VI (Data Management) and Chapter VII (Data Analysis and Reporting) of the Network Vital Signs Monitoring Plan are intended to provide summary information excerpted from the respective network's DMP. In addition, the DMP will be attached as an appendix to the Network Vital Signs Monitoring Plan. This approach ensures networks prioritize data management planning early on in program development.

It is beyond the scope of this document to provide detailed information on every data management procedure, database structure, etc., that will be implemented by the program. Much of that detail, as appropriate, will be found within the data management standard operating procedures (SOPs) as they are developed for the various monitoring protocols (i.e., the "third level" of data management guidance). The primary purpose of this DMP is to communicate an overarching data management strategy that establishes guidance and specific policy, as appropriate. This document:

- Establishes a framework within which more detailed data management procedures (i.e., SOPs, Databases, Protocols, Product Deliverables, etc.) and data management tools will be developed
- Defines data stewardship roles and responsibilities for those involved in GULN business
- Provides an overview of data management infrastructure
- Addresses specific requirements for the collection, management, maintenance, and dissemination of GULN data in its various forms and forma

1.3. Data Management Plan Goals

One of the five service-wide goals of the NPS I&M Program is to provide 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 DMP outlines the procedures and work practices that support effective data management and may also serve as a guide to convert legacy information to match formats and standards described in this document.

The DMP goals are to ensure that:

1. 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;
2. Network data are readily available, by implementing standard procedures for distributing data, while protecting sensitive data; and designing a standardized digital filing system for organizing I&M information;
3. 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; and integrating common data tables and fields in NPS or regional standards

1.4. Audience

The intended audiences of the DMP are:

- All Network staff

- All other NPS staff and agencies cooperating with the GULN I&M Program.

1.5. Revisions to the Data Management Plan

The DMP is scheduled for a full review and revision every five years, coinciding with the GULN I&M Program review schedule. Until then, minor revisions may be made as needed and will be documented in the corresponding chapters.

1.6. Types of Data Covered by this Plan

A distinction is usually made between the term 'data', which refers to assemblages of raw or un-interpreted facts, records or observations, and 'information', which is created from data through the process of analysis, synthesis, modeling, or other types of interpretation. For the sake of discussion in this plan, data will be used to signify both data and information.

Data must be accompanied by sufficient context about how and why they were collected, if they are to maintain their long-term value. A data management program cannot simply attend to the tables, fields, and values that make up a data set. There must also be a process for developing, preserving, and integrating the context that makes the data usable.

This plan encompasses a range of products that are coordinated or managed by GULN. These products fall into four general categories: data (raw, verified/validated and analyzed), documentation, reports, and administrative records (Table 1.1).

Table 1.1. Categories and examples of data products covered by the GULN DMP.

Category	Description	Examples
Raw Data	Data obtained from the environment and that has not been subjected to any quality assurance or control beyond those applied during field work.	• Field data sheets • Specimens • Remotely sensed data • Data gathered electronically on field computers • GPS rover files • Photographic imagery
Verified and Validated Data	Data that has been evaluated for completeness, correctness, and conformance/compliance of a specific data set against the standard operating procedure (verified), as well as reviewed for specific analytic quality (validated).	• Relational databases • Tabular data files • Laboratory results • GIS layers • Maps
Analyzed Data	Data that have been subjected to analytical routines after field collection and verification. This includes statistical operations conducted on the data for the purposes of arriving at a measure of the given ecological parameter or a compilation of analyzed data from different sources or time periods to derive new information.	• Summarized reports, data and maps from statistical or query operations

Category	Description	Examples
Documentation	Documentation provides the information required to understand the context of the data.	• Data collection protocols • Data processing/analysis protocols • Record of protocol changes • Data dictionary • FGDC metadata • Database design documents • QA/QC reports • Catalogs
Reports	Reports provide a means of presenting and publishing the methods and the results of analysis in the context of which it was intended.	• Annual progress reports • Final reports • Trend analysis reports • Publications
Administrative Records	Administrative records supplement the context of a project and should be considered part of the projects deliverables.	• Contracts and agreements • Study and work plans • Research permit • Critical administrative correspondence

1.7. Sources and Priorities of Natural Resource Data

Potential sources of important data and information about the condition of natural resources include:

- Inventories
- Vital Signs Monitoring
- Research done by GULN staff, contractors, or cooperators
- Monitoring or research done by other agencies or on adjacent lands

The priorities for network data management efforts are:

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

1.8. Organization of this Data Management Plan

Chapters 1 through 4 of this DMP are overview chapters describing roles, processes and framework. Chapters 5 through 11 discuss data management operations. When applicable, legal mandates are presented at the beginning of the chapter. Acknowledgements to authors are included at the end of each chapter. The writing of this DMP benefited immensely from discussions with other network data managers. The material presented here draws heavily from the writing of many of those data managers, and we've tried to recognize them all in the Acknowledgements section. The Literature Cited section at the end of the DMP provides a selected list of documents, guidelines or policies that may be helpful in further discussion and

development of data management operations. Appendices present additional materials that augment the data management strategies and procedures presented in the body of this plan.

1.9. Chapter Credits

Margaret Beer (Northern Colorado Plateau Network), Doug Wilder (Central Alaska Network), John Boetsch (Northern Coastal and Cascades Network), Sara Stevens (Northeast Coast and Barrier Network), and Alan Williams (Shenandoah National Park), Dorothy Mortenson (Southwest Alaska Network), Gordon Dicus (Pacific Island Network).

2. Data Management Roles and Responsibilities

Data management is about people and organizations as much as it is about information technology and database theory and application. Nearly every person in an organization manages data and information at some level. Good data stewardship can not be accomplished by data managers alone – it is truly a collaborative endeavor that involves many people with a broad range of tasks and responsibilities. The GULN contributes to the public service mission of the NPS and the I&M Program by providing status and trends information about park ecosystems. This requires every individual involved in Network business to understand and perform data stewardship responsibilities in the production, analysis, management, and/or end use of the data (Table 2.1). As coordinator of these activities, the fundamental role of the Network data manager is to understand program and project requirements, create and maintain data management infrastructure and standards, and educate, communicate, and work with all responsible individuals. This chapter discusses comprehensive data management roles and responsibilities that generally apply to all Network activities. Individuals who carry out monitoring protocols and inventory study plans are responsible for reading and understanding guidance at two levels – the general, overarching level of this DMP and the specific, task-oriented level of the respective protocol or study plan. This chapter clarifies roles and outlines the principal and ancillary data management responsibilities of network staff, cooperators, and other project participants.

Table 2.1. Categories of data stewardship involving all Network personnel and cooperators.

Stewardship Category	Related Activities	Position
Production	Collecting data or information from any original or derived source. This includes 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, and preparing and delivering informative products, such as summary tables, maps, charts, and reports.	Project Manager/Principal Investigator Project Crew Member Project Data/GIS Specialist or Technician
Analysis	Using data to predict, qualify, and quantify ecosystem elements, structure, and function as part of the effort to understand these components, address monitoring objectives, and inform park and ecosystem management.	Network Ecologist Park Resource Specialist
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 Regional GIS Manager Regional IT Specialist Project Database Manager Service-wide 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

2.1. Data Stewardship Roles and Responsibilities

A role is a function or position (e.g., Project Manager)
A responsibility is a duty or obligation (e.g., review data records)

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 steward data and information assets. Knowledgeable individuals from scientific, administrative, and technological disciplines must work in concert to ensure that data are collected using appropriate methods, and that resulting datasets, reports, maps, models, and other derived products are well managed. Datasets and the presentations of these data must be credible, representative, and available for current and future needs. Stewardship responsibilities apply to all personnel who handle, view, or manage data (Table 2.2). Vital sign monitoring protocols 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.
Computer Programmer	Apply knowledge and abilities related to database software and writing special application code.
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 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. Facilitate and integrate data and metadata. Oversee long-term data storage and maintenance. Design and develop databases and applications. Update software and hardware and implement secure file server backup scheme.
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.
GIS Specialist	Provide support for long-term storage of GIS data. Update and maintain GIS software and tools. Provide technical assistance.
Information Technology Specialist (USGS or Region)	Maintain LAN, establish and maintain system security.
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.

Among all of the data management responsibilities shared by project and Network personnel, the careful documentation of datasets, 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 products, 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. (Chapters 7 and 8 present important guidance and reference for documentation and metadata.)

2.2. Data Management Coordination

Network coordinators, project managers, and data managers (and potentially GIS specialists and other specialists) comprise the central data management team for inventory and monitoring projects. Each is responsible for certain aspects of project data and all share responsibility for some overlapping tasks (Figure 2.1). Because of the collaborative nature of project data management, good communication among these personnel is essential to meeting program goals. Communication is promoted by providing:

- Documentation
- Shared working files
- Centralized data management guidance and information discovery on the GULN website
- User needs assessments and surveys
- Newsletters and periodic e-mail updates
- Work groups discussions and meetings
- Presentations
- Training

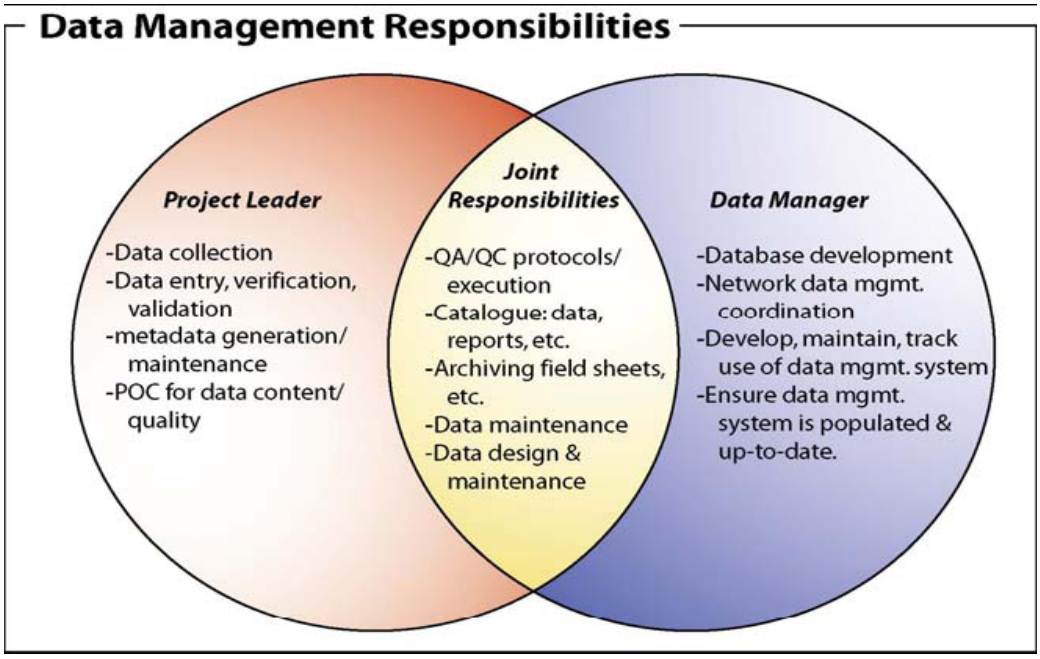


Figure 2.1. Core data management responsibilities for project managers and Network data managers.

2.2.1. Network Coordinator

The Network Coordinator supervises the project managers and has the ultimate responsibility for data entry, validation, verification, summarization/analysis and reporting. The Board of Directors and Technical Committee give final approval for products to be integrated, distributed to the public or parks, or protected in the case of sensitive information.

2.2.2. Project Manager

Project managers oversee and supervise all phases of an inventory and monitoring project from initiation to product delivery. Assuming "ownership" of the data, the project manager's active involvement in data management determines the quality and usefulness of the project data. The project manager provides project oversight, directs on-the-ground data collection efforts, and provides a cohesive link between data collection, synthesis, interpretation, and reporting. Project managers must act as a steward for all project data, and must work with project and GULN personnel to ensure the proper handling of project data. Project managers are responsible for designating an alternate leader who is capable of maintaining project operations in his or her absence, in order to ensure project continuity and data integrity.

Specifically, a project manager is responsible for:

- project documentation that describes the 'who,' 'what,' 'where,' 'when,' 'why,' and 'how' of a project
- documentation and implementation of standard procedures for field data collection and data handling
- implement quality assurance and quality control measures, which include the supervision and certification of all field operations, staff training, equipment calibration, species identification, data collection, data entry, verification, and validation
- maintenance of concise explanatory documentation of all deviations from standard procedures
- detailed post field-collection documentation
- maintenance of hard copies of data forms and archiving of original data forms
- scheduling of regular project milestones such as data collection periods, data-processing target dates, and reporting deadlines
- regular summary reports, periodic trend analysis of data, resulting reports, and their public availability
- identifying sensitive information that requires special consideration prior to distribution
- acting as the main point of contact concerning data content/quality

The project manager may also work closely with the data manager and/or a data specialist (such as a biometrician) to:

- develop quality assurance and quality control procedures specific to project operations
- identify training needs for staff related to data management and quality control procedures
- coordinate changes to the field data forms and the user interface for the project database (coordinate data entry procedures – data design maintenance)
- document and maintain master data (including metadata generation and maintenance)
- 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 the process of how project data will be transformed from raw data into meaningful information

- create data summary procedures to automate and standardize this transformation process
- identify and prioritize legacy data for conversion and convert priority datasets to a modern format
- increase the interpretability and accessibility of existing natural resource information

2.2.3. Network Data Manager

The data manager oversees the development, implementation, and maintenance of data infrastructure and standards for the Network. The data manager is responsible for ensuring the compatibility of project data with program standards and the long-term integrity and availability of project data. The data manager facilitates dissemination of project datasets and information products by working with project managers and other program staff to design databases, applications, and products,. The data manager exercises data stewardship to ensure data are archived, documented, and compatible with other program data. The data manager ensures the maintenance of mechanisms for making data discoverable and available. Both the data manager and the project manager ensure that the information conveyed via these mechanisms is up-to-date and accurate.

General data management duties for the data manager are:

- overall coordination of data management activities in the Network
- develop and maintain data and information housing and dissemination system for all program data and information
- work to improve the acquisition, accessibility and transparency of digital data
- ensure that the data and information system is populated and kept up-to-date with all relevant Program output
- develop and maintain logs recording both the changes/enhancements to the data and information-handling system/process and use/dissemination of system content (actual data and information)
- act as point of contact for access to Program output
- ensure data security (archiving operations, etc.)
- catalog project data and reports in nationally maintained I&M Program databases such as NPSpecies, NatureBib and NPS Data Store.

Data managers will also work closely with the project manager to:

- 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
- establish and implement procedures to protect sensitive data according to project needs
- collaborate with GIS specialists to integrate tabular data with geospatial data in a GIS system in a manner that meets project objectives
- 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

2.2.4. Network GIS Specialists

The Regional and Network GIS Specialist play a crucial supporting role for GULN projects. GIS support includes project planning to determine the GIS data and analysis needs for a project. GIS specialists will provide consultation to project managers on field collection of spatial data including the use of GPS and other spatial data collection techniques. They will also coordinate importing of spatial data into the GIS, work with project managers to analyze spatial data, and provide the results in map or tabular form. The GIS specialists will work with project managers to properly document data in compliance with spatial metadata standards and will be responsible for stewardship of GIS data and products.

GIS specialists will also work directly with data managers to design databases and applications, to create relationships between GIS and non-spatial data, and to create appropriate database and GIS applications to facilitate the analysis of both spatial and non-spatial data. Maintaining standards for geographic data, GIS specialists are responsible for sharing and disseminating GIS data throughout the network.

2.3. Chapter Credits

Rob Daley (Greater Yellowstone Network), Margaret Beer (Northern Colorado Plateau Network) and Doug Wilder (Central Alaska Network).

3. Data Management Process and Workflow

This chapter gives an overview of how the GULN natural resource data are generated, processed, finalized, and made available. The process is described in the context of the progressive stages of a project, and the life cycle of the resulting data.

3.1. Project Work Flow

From the perspective of managing work flow, there are two main types of projects:

- *Short-term projects*, which may include individual park research projects, inventories, or pilot work done in preparation for long-term monitoring.
- *Long-term projects*, which are mainly monitoring projects central to the I&M Program, but may also include multi-year research projects and monitoring performed by other agencies and cooperators. Long-term projects typically require a higher level of documentation, peer review, and adherence to standards to ensure consistency over time.

Projects can be divided into five primary stages: planning and approval; design and testing; implementation; product integration; evaluation and closure (Figure 3-1, Table 3.1). Each stage is characterized by a set of activities carried out by staff involved in the project. Primary responsibility for these activities rests with different individuals according to the different phases of a project. Additional discussion of the different roles and responsibilities of park and network staff can be found in Chapter 2 of this plan.

The Network uses a project tracking database (developed in Access by the Alaska Network) to document and support the progress of a project. This database provides an overview of the project, identifies primary contacts and due dates, lists deliverables, and notes the status and location of the deliverables. It also helps to generate reports used in the final archiving of a project. The database will be modified as necessary throughout the growth of the monitoring program.

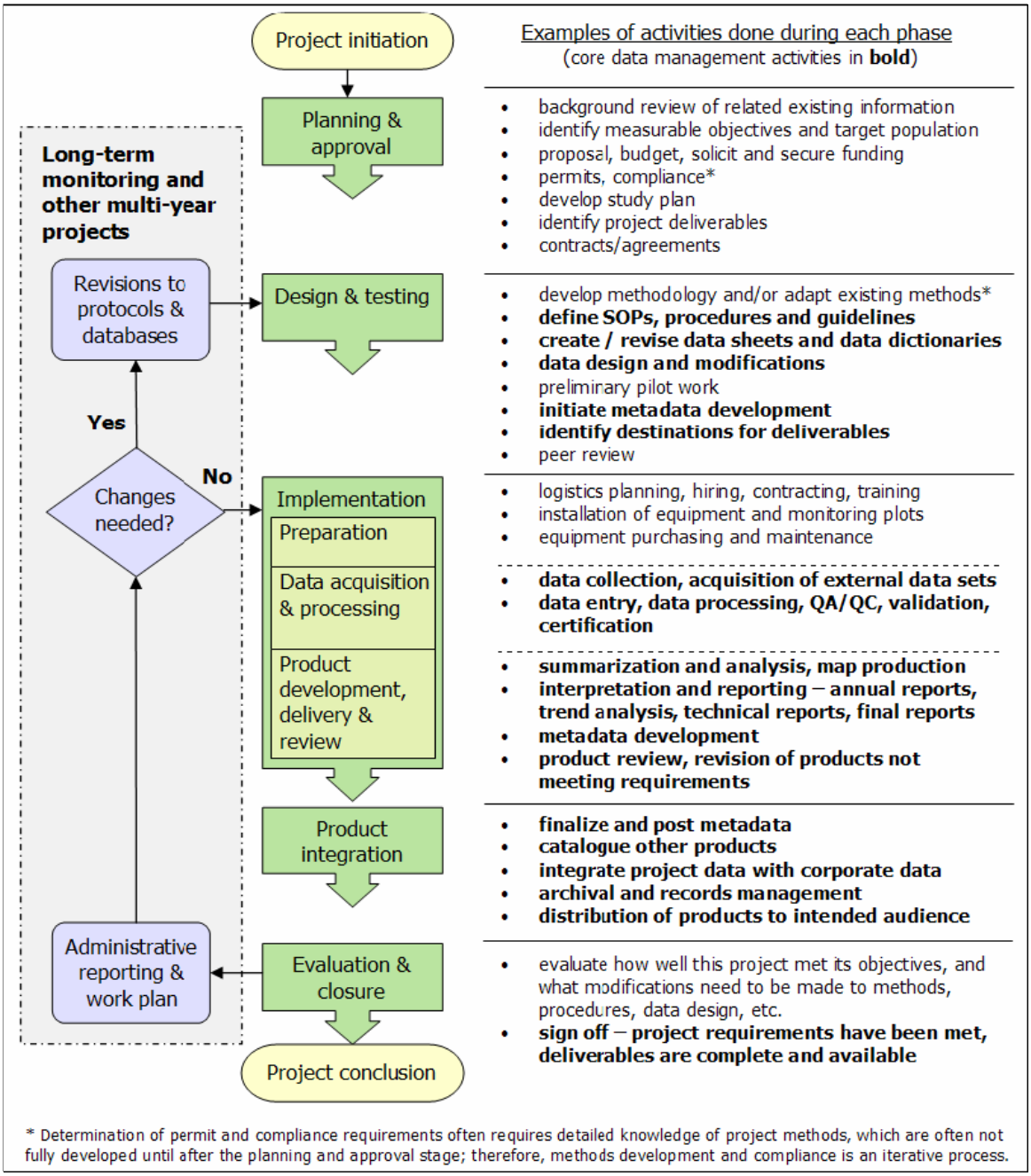


Figure 3.1. Generalized project work flow model for short- and long-term data collection efforts. Bold face activities indicate core data management elements.

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.

Project Stage	Task	Task Description	PL	EC	DM	GIS
Planning & Approval	1.1	notify data manager of new project in planning stage	*			
		Develop study plan and budget	*	-	-	
	1.2	create a project tracking record and assign a project code (status: planned)	-		*	
	1.3	identify project deliverables	*	-	-	-
	1.4	determine data sensitivity and ownership	*	-	-	-
	1.5	develop language for contracts/agreements and permits	*	-	-	-
		Apply for permits in the park and note in project tracking	*			
Design & Testing	2.1	notify data manager of changing status	*			
	2.2	update project tracking record (status: design or rejected)			*	
	2.3	clarify roles and responsibilities	*	*	*	*
	2.4	scoping for information systems design	*	*	*	*
	2.5	initiate metadata record	*		-	*
	2.6	create data dictionary and data forms	*		*	*
	2.7	define quality assurance measures	*	-	-	
	2.8	database and application development	*	-	*	*
	2.9	sample site selection and documentation	*	*	-	*
	2.10	provide training as needed and time permits	*	*	*	*
Implementation	3.1	notify data manager of changing status	*			
	3.2	Training	*	*		*
	3.3	update project tracking record (status: active or abandoned)			*	
	3.4	data acquisition, data entry, upload data logger files	*			
	3.5	upload GPS and other data logger data into database				*
	3.6	quality assurance: verification, validation and processing, data certification	*	*		
	3.7	update metadata and documentation	*			
	3.8	deliver data and metadata to DM/GIS (see Task 5.2)	*			
Product Development & Delivery	4.1	summarization, analysis, map production	*	*	*	*
	4.2	interpret results, reporting, peer review	*	*		*
	4.3	assemble products and deliver to DM	*	*		*
	4.4	route products to appropriate product handler(s)			*	
	4.5	check for completeness, proper format, metadata and certification			*	*
	4.6	rename files, store in digital library			*	*

Project Stage	Task	Task Description	PL	EC	DM	GIS
	4.7	update project tracking database (deliverable status: checked in)			*	
	4.8	upload certified data into master project database			*	-
	4.9	finalize and parse metadata record	-		*	*
Product Integration	5.1	publish products and integrate with national applications (NatureBib, NPSpecies, NPSTORET, NR-GIS Data Store)	*	*	*	*
	5.2	update project tracking database (deliverable status: published)			*	*
	5.3	Electronic permanent storage			*	*
	5.4	Hardcopy archival and distribution	*		-	
Evaluation & Administrative Reporting	6.1	verify that all products are accounted for	*		*	*
	6.2	administrative evaluation and reporting, decision on continuation and methods changes for long-term projects	*	-		
	6.3	communicate administrative decisions and information management needs to data manager/GIS specialist	*			
	6.4	update project tracking database (project status: complete, design, active or suspended)	-		*	

3.1.1. Planning and Approval

During planning and approval, a request for proposals package may be distributed to likely candidates or a proposal may be formulated in-house. The proposal provides the initial ideas of a project by providing enough description and budget information to determine if the idea should be continued. If it is, then a more detailed study plan is written and reviewed. Support is provided either in house or funded via a contract or cooperative agreement. A contract or cooperative agreement will make reference to the detailed study plan to describe the scope of work. The annual work plan is supplemental to the study plan and provides specific tasks and identifies deliverables for a specific year. Permits applications are completed after the study plan is approved and must be in hand before any work is started.

3.1.2. Design and Testing

During the design and testing phase, a project is ready to begin formalizing and testing of sampling methods, procedures, and analysis. A more thorough Data Discovery search is done using the many sources readily available. Methods and sampling design are developed or modified, field forms and datasheets/databases are created, and standard operating procedures are written. These methods, data, and SOPs are tested with “dummy data” or practiced, and are tested for quality control/quality assurance and preliminary integration into permanent storage. Many SOPs and guidelines will be written for the implementation, product delivery and product integration of the I&M Program. Review of these existing documents at this stage is advised. It is critical at this stage for the project manager and the data manager to work together. The dialog between these two will help to ensure information is consistently captured and entered into a database, and results can be retrieved. Scheduling of a meeting as early as possible (shortly after approval) is encouraged, and is the responsibility of the project manager.

3.1.3. Implementation

How the project will be implemented can vary greatly between projects, but all should comply with the methods described in the study plan and approved protocols. In many projects, preparation to go into the field is a significant portion of the project. Field work is done if the project requires it. Data are entered and analyzed following data guidelines. Materials collected are handled and stored appropriately following materials guidelines. Final or annual product deliverables are prepared. Time between collection, data entry, and analysis should be minimized.

3.1.3.1. Preparation

Preparation for a project includes logistics, staffing, training, acquiring and maintaining equipment, and preparation for accessioning. In addition the PI should ensure a permit has been acquired and is in hand. No data deliverables are anticipated at this time.

Logistics
It is advised to assign a contact person within the park who can provide logistical support. Logistics may include: flights, housing, and temporary storage of materials collected or of equipment.

Trainin
How well the data collectors are prepared and able to complete the tasks may be the most critical element to a successful project. Training may include technical data collection training, such as using a GPS, data logger, or specialty equipment. Other data collection training may include using field guides, entering field forms, or making estimates in classification. It is advised to check with the individual park Chief of Resources or Ranger for other required training. Such training of field staff may include ATV, first aid, small aircraft, small boat, or firearms safety.

Equipment
Equipment purchases should follow the standard NPS operating procedures. Equipment should be tracked and "checked out", then "check in" at completion. Ensure equipment is properly calibrated before and during its use.

Accessioning
All completed projects are to be archived. This is particularly important when materials from the park are collected, such as plant or animal specimens, as these are property of the park and must be traceable. Contact the park’s archives specialist to acquire an accession number. Be prepared to label all vials, notebooks, etc. with this number.

Data Collection
Acquisition of external data sets may be collected and analyzed. Field collection should be conducted as specified in the study and work plans, or as specified in protocols and supporting SOPs. The PI is responsible for overseeing the collection of data. Any deviation from these protocols should be carefully documented, such as in the case of equipment failure or adverse conditions.

Data Entry and Verification
After the data collection is complete, the information and materials collected need to be entered into a database or processed. This may include: data entry, labeling and storage of materials, preliminary reporting and analysis of the information, quality control/quality assurance procedures, and documentation. Methods for data entry and verification should adhere to protocols or to the study plan, and should reflect 95% or greater accuracy in data entry.

3.1.4. Product Development and Delivery

Product delivery is the completion of the deliverables as specified in the study and work plans, packaged and documented both in hard-copy and electronic format, typically as a stand-alone product. In almost all projects, a report is required to summarize and analyze the data and to make recommendations. Each product should follow the guidelines regarding that particular type of deliverable.

3.1.5. Product Integration

Products for a single project should be integrated into a larger repository of information. These include entering reports into the appropriate bibliographies and libraries, data into project databases, and materials into the archives. This will allow the information from the project to be searchable in the service wide search engines and retrievable by others, and will provide the long term storage of these products. A project may, however, have additional integration needs, such as when working jointly with other agencies for a common database. Ensuring this step of the project is done is a joint responsibility between the project manager and the data manager.

3.1.6. Evaluation and Administrative Reporting

For long-term monitoring and other cyclic projects, this phase occurs at the end of each field season, and leads to an annual review of the project. For short-term projects, this phase represents the completion of the project. After products are catalogued and made available, the program administrator, project manager, and data manager assess how well the project met its objectives, and determines what might be done to improve various aspects of the methodology, implementation, and formats of the resulting information. For monitoring protocols, careful documentation of all changes is required. Changes to protocol narratives, standard operating procedures (SOPs), and other procedures are maintained in version tables associated with each document. Major revisions may require additional peer review.

3.2. Data Life Cycle

During various phases of a project, data take on different forms and are maintained in different places as they are acquired, processed, documented, and archived. This data life cycle is characterized by a sequence of events that we can model to facilitate communication. These events involve interactions with the following objects:

- *Raw data* – analog data recorded by hand on hard-copy forms and digital files from handheld computers, GPS receivers, automated data loggers, etc.
- *Working database* – a project-specific database for entering and processing data for the current season (or other logical period of time). This might be the only database for short-term projects where there is no need to distinguish working data for the current season from the full set of validated project data.
- *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 manager that the data have passed all quality assurance requirements and are complete and ready for distribution. Metadata records include the detailed information about project data needed for their proper use and interpretation.
- *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, primarily for the purpose of integration among NPS units and for sharing information with cooperators and the public.
- *Local archives and digital library* – local storage of copies of data, metadata, and other products generated by projects. Archives are for hard-copy items and off-line storage media, whereas the digital library is maintained live on a server.

Although the data life cycle may vary depending on specific project needs and objectives, the typical life cycle for GULN projects proceeds as follows (Figure 3.2):

1) *Acquire data* – for data recorded by hand in the field, data forms should be reviewed regularly (at least daily) for completeness, legibility, and validity in order to capture errors as close to their origin as possible.
2) *Archive raw data* – copies of all raw data files are archived 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. Archiving 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.

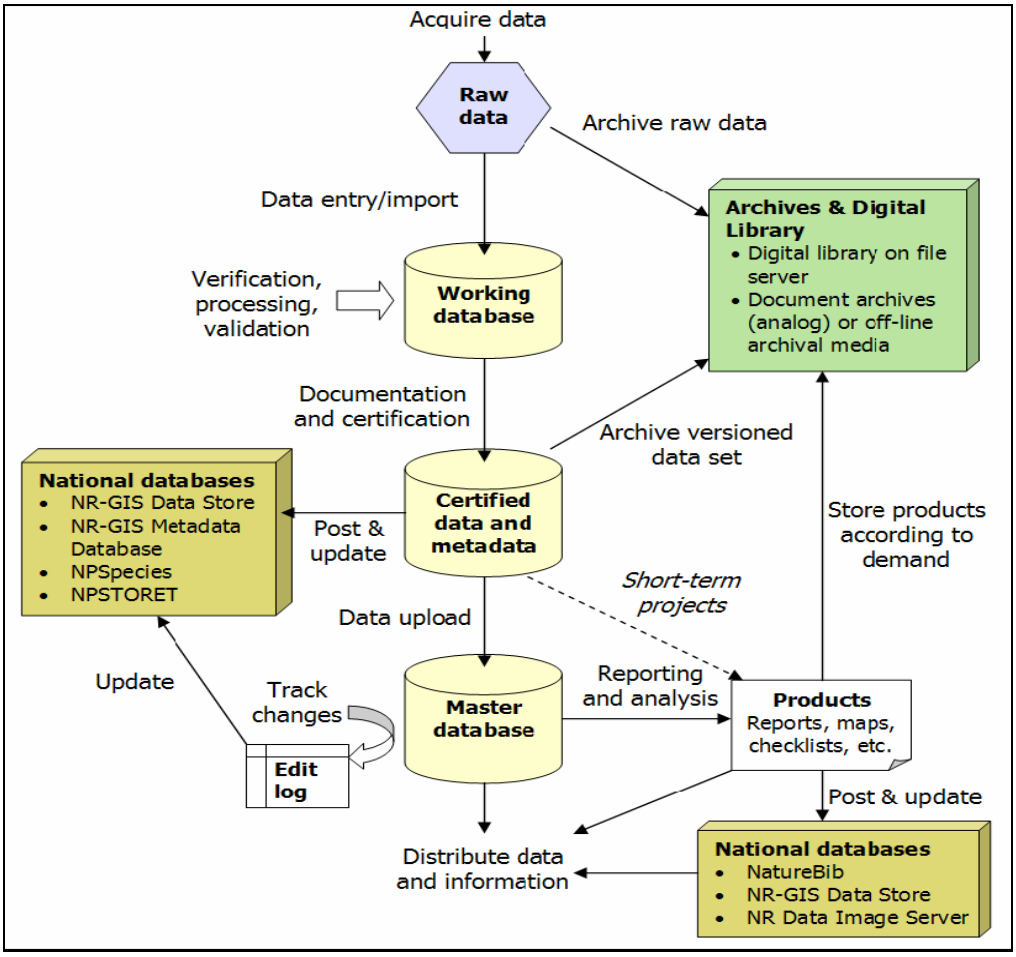


Figure 3.2. Diagram of the typical project data life cycle.

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; process data to remove missing values and other data flaws; and validate data using database queries to capture missing data, out-of-range values, and logic errors.
5) *Documentation and certification* – develop or update project metadata and certify the data set. Certification is a confirmation by the project manager 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.
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) *Post data and update national databases* – to make data available to others, certified data and metadata are posted to national repositories such as the NR-GIS Data Store. In addition, national databases such as NPSpecies, NPSTORET, and the NPS Data Store Database are updated. **Note:** Data and data products may not be posted on public sites if they contain protected information about the nature or location of rare, threatened or endangered species, or other natural resources of management concern.
8) *Upload data* – certified data are uploaded from the working database to the master project database. This step might be skipped for short-term projects where there is no need to distinguish working data for the current season from the full set of certified project data.
9) *Reporting and analysis* – certified data are used to generate data products, analyses, and reports, including semi-automated annual summary reports for monitoring projects. Depending on project needs, data might be exported for analysis or summarized within the database.
10) *Store products* – reports and other data products are stored according to format and likely demand – either in the digital library, on off-line media, or in the document archives.
11) *Post products and update national databases* – to make data available to others, reports and other products are posted to national repositories such as the NR-GIS Data Store or the NR Data Image Server. In addition, products are cataloged in NatureBib. Data products may not be posted on public sites if they contain protected information about the nature or location of rare, threatened or endangered species, or other natural resources of management concern.
12) *Distribute data and information* – data, metadata, reports, and other products 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 to cooperators. In all cases, distribution will follow legal requirements under the Freedom of Information Act, and limitations will be established to protect information about sensitive resources.
13) *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 the data and products to national databases and repositories.

This sequence of events occurs in an iterative fashion for long-term monitoring projects, whereas the sequence is followed only 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.

3.3. Integrating and Sharing Data Products

Once project data and derived products have been finalized, they need to be secured in long-term storage and made available to others. We use a range of information systems such as product repositories, clearinghouses, and web applications to accomplish this. Each of these systems has a different purpose and function, as shown in Figure 3.3.

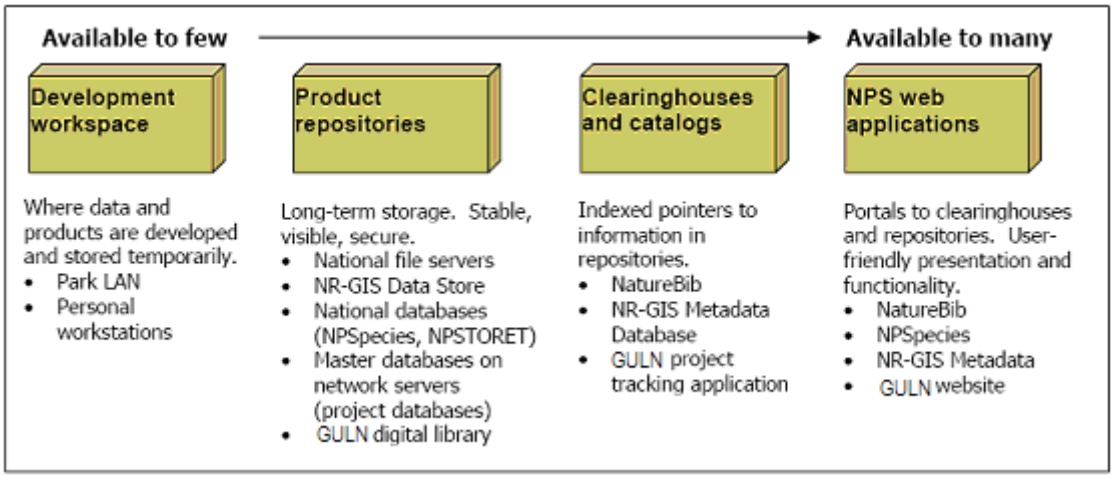


Figure 3.3. Storing and disseminating project information.

The specific repositories for most GULN products are indicated in Table 3.2.

Table 3.2. Repositories for GULN products.

Item	Repository
Reports	GULN digital library; posted to NR Data Image Server, linked and accessed through the catalog record in NatureBib; park collection (hard copy)
Digital data sets (non-sensitive)	NR-GIS Data Store Biodiversity Data Store
Digital data, metadata, and other products Raw and finalized data Metadata, protocols, SOPs Completed reports Digital photographs, derived products	GULN data servers and digital library; other cooperators for selected monitoring projects: Arizona Game and Fish Department, US Environmental Protection Agency, US Forest Service, etc.
Project materials Voucher specimens, raw data forms	Park archives and collections, or another park specified repository (e.g., Louisiana State University)
Administrative records	GULN offices and/or park offices, park archives, National Archives

3.3.1. Data Distribution

The process of product distribution involves several steps (see example for a GIS data set in Figure 3.4). As products are finalized, they can be sent to the appropriate person for integration, posting, and distribution. In most cases, it will be either the data manager or GIS specialist who reviews the product for conformance with format standards, then stores the product in the appropriate repository. Note that it is expected that all products will have already been reviewed for completeness and accuracy prior to delivery. Before storing the products on the GULN server, they are documented with metadata. At this point, data discovery is possible as metadata are then indexed by the clearinghouse function of the NPS Data Store Database. These metadata records

provide pointers to data and data products. Distribution then follows as data discovery allows potential users to find and either request or download the data sets from the NPS Data Store.

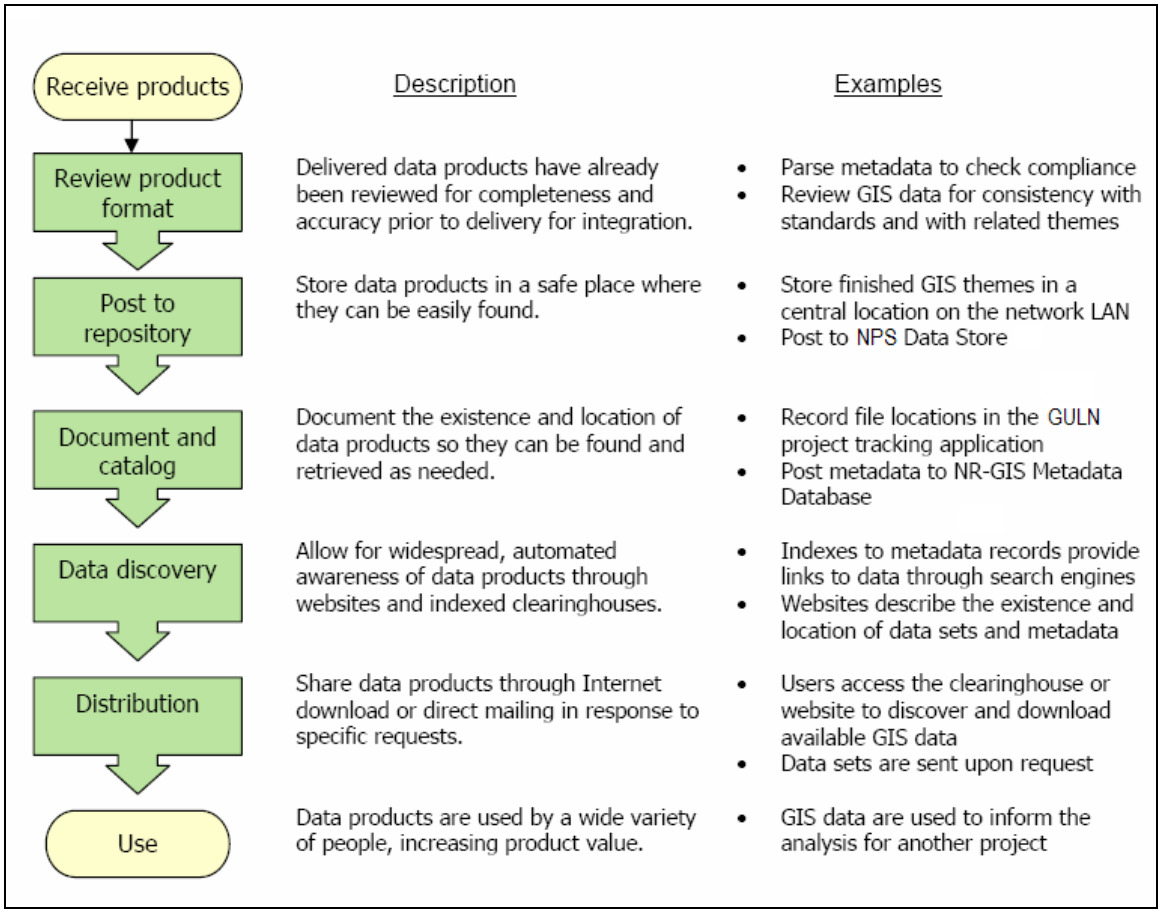


Figure 3.4. Steps involved in product distribution.

3.3.2. Integration with National Databases

In addition to storing and distributing data products, product integration also involves updates to national databases such as NPSpecies and NPSTORET. Both of these databases have local desktop versions that can be updated with data collected during the course of a project. Desktop databases are then uploaded and synchronized with the national databases on a regular basis.

To update NPSpecies, data on the distribution and occurrence of species in GULN parks will be compiled and added to the database upon delivery of data and data products. Updates will be performed in the master online version whenever possible. If the desktop version is used, synchronization with the master version of NPSpecies will occur at least twice annually, or more frequently depending on the timing and amount of updates.

For NPSTORET, any project collecting water quality data will be flagged in the project tracking application so that water quality data can be either extracted and uploaded or directly entered into NPSTORET. All water quality data collected by the Network will be managed according to guidelines from the NPS Water Resources Division. We will implement and maintain a desktop copy of NPSTORET and transfer its contents at least annually to NPS Water Resource Division for upload to the STORET database (Figure 3.5).

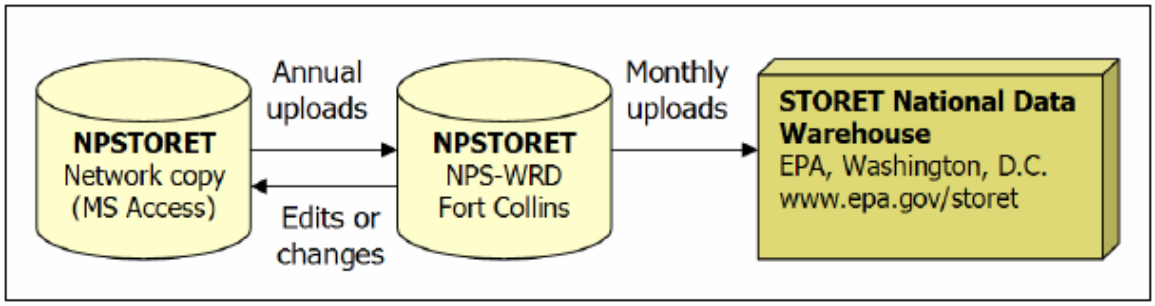


Figure 3.5. Data flow diagram for water quality data.

3.4. Review Cycle

Integration and use of products will identify holes or inefficiencies. A review of the information system as a whole every five years is being considered by the Network.

3.5. Chapter Credits

John Boetsch (Northern Coastal and Cascades Network), Sara Stevens (Northeast Coast and Barrier Network), Vel Decker (Cape Cod National Seashore), Doug Wilder (Central Alaska Network), Karen Oakley (USGS, Biological Division), and Sara Wesser (Alaska Region I&M Program).

4. Data Management Infrastructure and System Architecture

The ultimate measure of success for the data management strategy of the I&M Program is to supply quality information to the scientific community for research and analysis. The most common mechanism for delivering information to the scientific community is via the use of a data clearinghouse. The term 'clearinghouse' loosely refers to one-stop-shopping on a website. Clearinghouses provide a mechanism for searching, browsing and downloading information at the convenience of the user.

This chapter discusses the components of the I&M Program's overall information system architecture and the GULN computer resources infrastructure that allow data and information to be served to a broad audience, including park managers, scientists, and the public. Our infrastructure is composed of computers and servers that are functionally or directly linked through computer networking services, and represents the foundation upon which our network information system is built. System architecture refers to the applications, database systems, repositories, and software tools that make up the framework of the I&M Program's data management enterprise. See Figure 4.1 for a simplified comparison of infrastructure and system architecture.

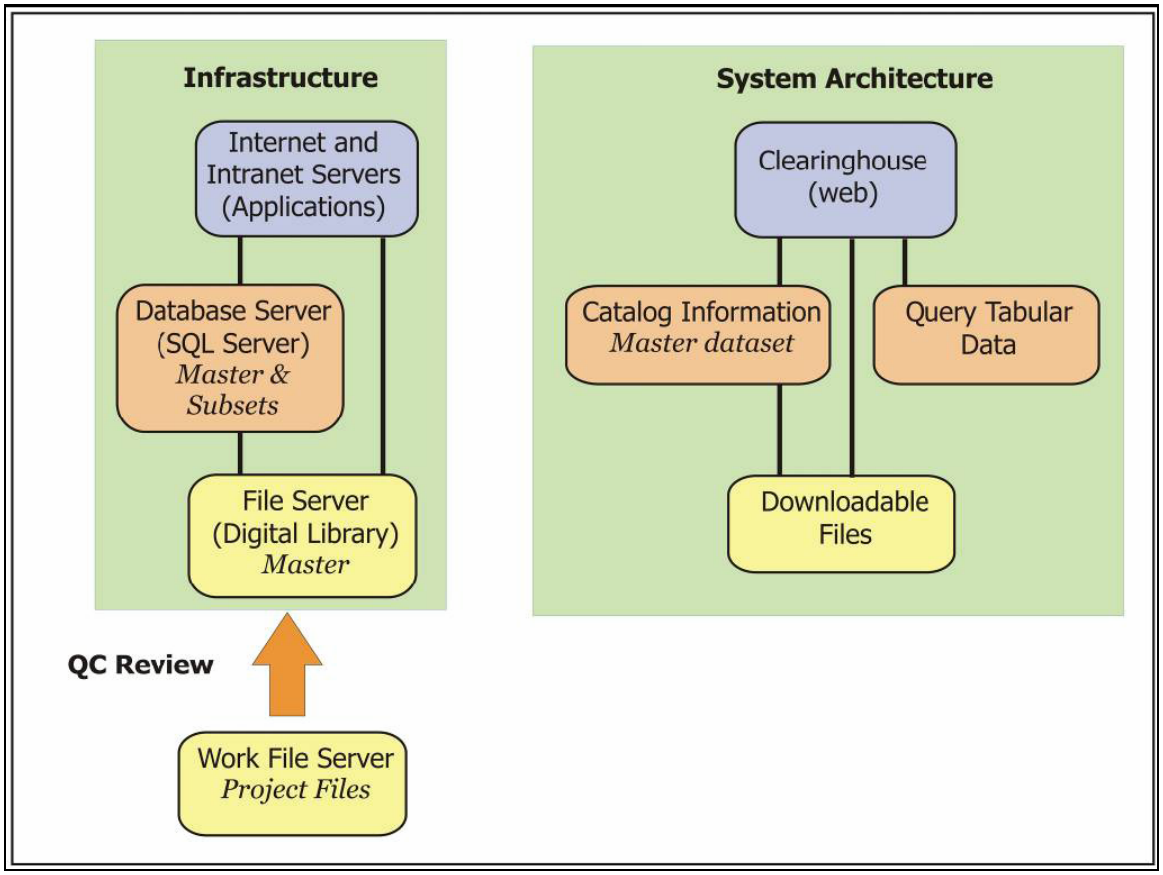


Figure 4.1. Simplified comparison of infrastructure and system architecture.

4.1. Computer Resources Infrastructure

"Infrastructure" refers to the system of computers and computer networks that our information management system is built upon. Our Network infrastructure works with two main components: 1) network-based local area network (LAN), including 6 network computer and 2 severs, maintained by the University of Louisiana at Lafayette (ULL) and the USGS's National Wetlands Research Center (NWRC); 2) wide area network (WAN) servers that are maintained at the national level by the Washington Support Office (WASO). These components each host different parts of our natural resource information system.

4.1.1. Network Data Server

- master project databases – compiled data sets for monitoring projects and other multi-year efforts that have been certified for data quality
- common lookup tables – park name, employees, species, etc
- project management application – used to track project status, contact information, product due dates
- digital library – network repository for read-only pdf versions of project deliverables, reference documents pertaining to NPSpecies, and general documents related to the network. (e.g., reports, methods documentation, data files, etc.)
- digital photo library-network repository for digital photos not related to a specific vital sign and are not considered data photos.
- local applications – desktop versions of national applications such as NPSpecies and Dataset Catalog
- working project materials – working databases and draft copies of reports
- archived final data sets for inventory and monitoring projects
- public file sharing directory

4.1.2. Network GIS Server

- GIS files – personal GIS projects, park base spatial data and imagery, project-specific themes, LIDAR data, and Remote Sensing analysis working and final data.

4.1.3. National Servers

- master applications – integrated client-server versions of NatureBib, NPSpecies, NPS Data Store
- centralized repositories – Natural Resource Data Store, Protocol Clearinghouse
- public internet access sites – portals to NatureBib, NPSpecies, NPSFocus, websites for monitoring networks

4.2. National Information Management Systems

The need for multi-discipline natural resource information management cuts across NPS divisional boundaries, and requires national-level information management strategies for success.

Inventory and monitoring information needs are broadly separated into two categories:

- *Detailed data needed for on-site resource management and protection*. The information used to guide natural resource management decisions must be specific and useful to management staff at parks and central offices.
- *Summary information needed to describe the resources and their condition*. This kind of information usually needs to be aggregated across the NPS for use by NPS and DOI managers and central office personnel to answer requests from Congress and for budget, program, and project planning.

The NPS Natural Resource Program Center (NRPC) and the national I&M Program actively develop and implement a national-level, program-wide information management framework. NRPC and I&M staff integrate desktop database applications with internet based databases to serve both local and national-level data and information requirements. NRPC staff members work with regional and support office staff to develop extensible desktop GIS systems that integrate closely with the database systems. Centralized data archiving and distribution capabilities at the NRPC provide for long term data security and storage. NRPC sponsors training courses on data management, I&M techniques, and remote sensing to assist I&M data managers with developing and effectively using natural resource information.

4.2.1. National-level Application Architecture

To achieve an integrated information management system, three of the national-level data management applications (NatureBib, NPSpecies, and NPS Data Store Database) use a distributed application architecture with both desktop and internet-accessible (master) components (Figure 4.2).

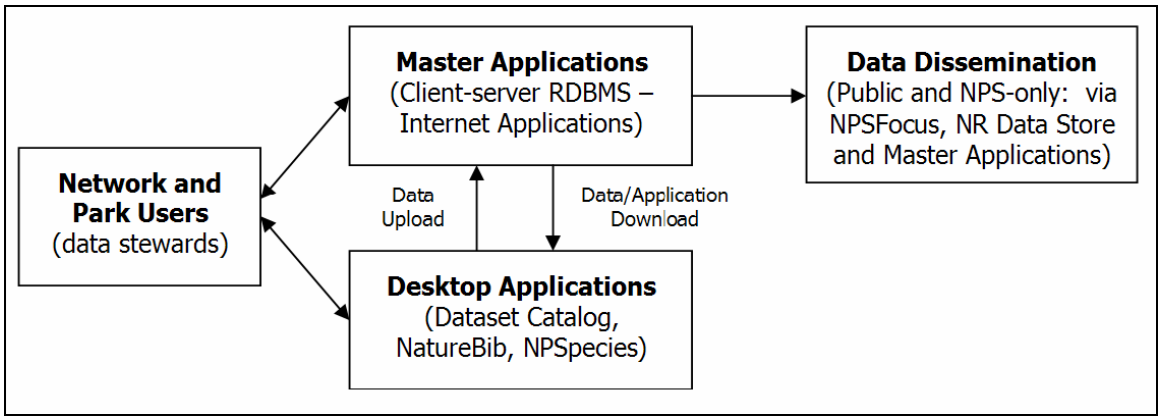


Figure 4.2. Model of the national-level application architecture.

4.2.1.1. 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 such as NPSpecies, Dataset Catalog, and the NPS Data Store Database. NatureBib currently focuses on park natural resource related references, but it may eventually be linked to references on cultural resources and other park operations. As with NPSpecies Database, it is possible to view and enter data in a MS-Access desktop version that can be used locally.

However, data must be park data already in the master version must be requested from the WASO Fort Collins, CO office. *(http://www.nature.nps.gov/nrbib).*

4.2.1.2. NPSpecies

NPSpecies is the master biodiversity database for NPS. The database lists the species that are known or suspected to occur in each park unit, and the physical or written evidence for the occurrence of the species (i.e., references, vouchers, and observations). Taxonomy and nomenclature are based on ITIS, the Integrated Taxonomic Information System.

The most current version of NPSpecies for each park or network can be downloaded from the master website into a local, MS-Access version. The Internet-based version is the master database, which can be accessed via password-protected logins administered by park, network and regional data stewards assigned for each park and network. 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, and will be linked to NPS Data Store Database to document biological inventory products. The MS-Access application and additional details can be found at the NPSpecies website *(http://science.nature.nps.gov/im/apps/npspp/index.htm).*

4.2.1.3. NPS Metadata Tools & Editor and NPS Data Store Database

NPS Metadata Tools & Editor provides the ability to create brief or FGCD compliant metadata. It is a stand-alone tool for creating and manipulating non-spatial metadata (outside the ArcCatalogTM environment) and provides authoring and editing spatial and nonspatial metadata records. As with other service-wide applications, the NPS Metadata Tools & Editor is used to upload metadata records to the master metadata database (NPS Data Store Database)on line and will be linked to NPSpecies (the NPS species database) and NatureBib (the bibliographic database). The tool can be downloaded from http://science.nature.nps.gov/nrdata/tools. The NPS Data Store can be accessed at http://science.nature.nps.gov/nrdata.

4.2.1.4. NPSTORET

STORET (STORage and RETrieval) 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. 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 funded WRD project be archived in STORET. NPSTORET (also known as Water Quality Database Templates) is the NPS master database designed to facilitate park-level standardized reporting for STORET. The database is still in development, but metadata, protocols, data dictionaries, and reporting capabilities are available through a front-end form. Upon implementation, network staff and cooperators will be able to use the MS-Access version of NPSTORET either as a direct database for data entry and management, or as a means of submitting data for upload to STORET by WRD staff. The MS-Access application and additional

details can be found at: http://www.nature.nps.gov/water/infodata.htm. Additional information on STORET can be found at: http://www.epa.gov/storet.

4.2.1.5. Natural Resource Database Template

The Natural Resource Database Template (NRDT) is a relational database in MS-Access for storing and managing inventory and monitoring data, and is a key component in standardizing the development of I&M program database structures. NRDT includes separate modules for different aspects of monitoring project implementation, from sampling design to data analysis and reporting, and includes data management components that describe database table structure, data entry forms and quality checking routines. Established monitoring protocols, including associated databases that are based on NRDT, are available through a web-based protocol clearinghouse (see below). A description of the NRDT application, a data dictionary, and examples are located on the NRDT website (http://science.nature.nps.gov/im/apps/template/index.htm).

4.2.1.6. Natural Resource Monitoring Protocols Clearinghouse

The Natural Resource Monitoring Protocol Clearinghouse (Protocol Database) is a web based clearinghouse of protocols that have been developed by the prototype monitoring parks or other established protocols used in national park units. The database includes summaries and protocols, and in some instances allows the download of associated database components (e.g., tables, queries, data entry forms) that are consistent with the Natural Resource Database Template. See the Protocol Database website for available protocols (http://science.nature.nps.gov/im/monitor/protocoldb.cfm).

4.2.1.7. NR-GIS Data Store

The NR-GIS Data Store is a key component of the data dissemination strategy employed by the I&M Program. The NR-GIS Data Store is a web-based search tool that links dataset metadata to a data server. The interface allows customized public or protected searches of natural resource datasets, inventory products and GIS data produced by the I&M and Natural Resource GIS Programs. Each park or network is able to post and curate its data on the server. The NR-GIS Data Store will be integrated with the master NPS Data Store Database application to streamline programmatic data documentation and dissemination processes. The simple browse function of this server can be accessed at: http://nrdata.nps.gov/. See the NR-GIS Data Store website for further information (http://science.nature.nps.gov/nrdata).

4.3. GULN Information Management System Architecture

GULN offices reside on the campus of the University of Louisiana at Lafayette and the NOAA Coastal Fisheries and Estuarine Habitats building both located in Lafayette, LA. The network computers are connected through a work group established on the University LAN where access is gained to the internet and the national WAN. The servers reside in the data manager/GIS specialist's office in the NOAA building where they are maintained by the data manager. Network staff has access to both servers allowing easy file sharing and access to network data. At the time of this document, the network LAN consited of:

- Dell PowerEdge 2800, 400 GB, RAID 5 server
- Dell PowerEdge 2800, 800 GB, RAID 5 server
- Dell PCs
- Workstation/laptops

A complete list of the network equipment can be found in Appendix A. Information Technology (IT) duties are currently being administered by the network data manager with assistance from NWRC IT staff. These duties include but are not limited to: hardware replacement, software installation and support, security updates, virus protection, networking, and backups of servers. The data manager works with regional IT specialists to ensure Network infrastructure meets required standards, and that the best data security measures are in place.

4.3.1. Team and User Project Files

Initial project files, for both short and long term, are typically stored in the Working Projects folder on the data server. The project files must:

- Contain a Project Organizer Document – a document containing contact information and an explanation of the directory structure and where files are stored
- Conform to file naming standards
- Conform to documentation standards (Chapter 8)

The data manager will create a working directory for the projects. Project managers will modify the directory structure as needed, such as deleting directories that are not being used. Project directories may be "checked out" to project managers located at parks, but should remain in a location accessible to the network data manager (i.e., on the park resources active projects directory). Project directories may also be "checked out" to contractors with the understanding a similar directory structure is required as part of the deliverables. Project template directory structures for short-term projects and protocol development have been developed. Contact the data manager for the most current version. Completed projects and products are moved to the data file server and redistributed as read-only files to all parks using the necessary media (e.g., CDs, DVDs, external hard drives). Electronic files of final products are uploaded by the data manager to the NPS Data Store application for public distribution.

4.3.2. Integration of Products to Master Data Libraries

Digital versions of all deliverables, as specified in the project tracking database, cooperative agreements, permits, study plans, or other administrative records, will be electronically stored on servers using the following general methods:
1. Project archive on GULN server, where all records are "packaged" as a stand alone. Types of information to be included are listed in Chapter 1, Table 1.1.
2. Product Library on GULN server will house all the final reports. Citation information and key words will be added to Reference Manager following the Reference Manger SOP (Appendix B) and to NatureBIB.
3. Final GIS data will be filed in the appropriate folder on the GIS Server to be used as working data layers for creating maps. The final archived GIS data will reside with the project data in the project archive folder on the Data Server.
4. Final products and long-term datasets will be uploaded to WASO Oracle database servers, in order to serve information to a broad audience (e.g., NPS Data Store) or to update an annual network dataset to a service-wide database (e.g., NPSTORET).

4.3.3. Project Database Standards

Project database standards are necessary for ensuring compatibility among data sets, which is vital given the often unpredictable ways in which data sets will be aggregated and summarized. Well thought out standards also help to encourage sound database design and facilitate interpretability of data sets. Databases that are developed for park and network projects will all contain the following main components:

- Common lookup tables – These are accessed via links to tables that reside in a centralized database, rather than storing redundant information in each database. These tables typically contain information that is not project-specific (e.g., lists of parks, personnel, and species).
- Core tables and fields based on network and national templates – These tables and fields are used to manage the information describing the "who, where and when" of project data. Core tables are distinguished from common lookup tables in that they reside in each individual project database and are populated locally. These core tables contain critical data fields that are standardized with regard to data types, field names, and domain ranges.
- Project-specific fields and tables – The remainder of database objects can be considered project-specific, although there will typically be a large amount of overlap among projects. This is true even among projects that may not seem related – for example, a temperature field will require similar data types and domain values. As much as is possible, efforts will be made to develop these project-specific objects to be compatible with those maintained by other networks and cooperators managing similar data sets – especially if integration with other data sets is important for meeting project objectives.

As much as possible, GULN standards for fields, tables and other database objects will mirror those conveyed through the Natural Resource Database Template. Where there are differences between local and national standards, documentation of the rationale for these differences will be developed. In addition, documentation and database tools (e.g., queries that rename or reformat data) will be developed to ensure that data exports for integration are in a format compatible with current national standards.

Certain key information is not only common to multiple data sets, but to the organization as a whole – lists of contacts, projects, parks, species are often complex and dynamic. It is a good strategy to centralize this information so that users have access to the most updated versions in a single, known place. Centralizing also avoids redundancy and versioning issues among multiple copies because there is only one copy to maintain. Centralized information is maintained in database tables that can be linked or referred to from several distinct project databases. Network applications – for project tracking, administrative reporting, or budget management – can also link to the same tables so that all users in the network have instantaneous access to edits made by other users (Figure 4.3).

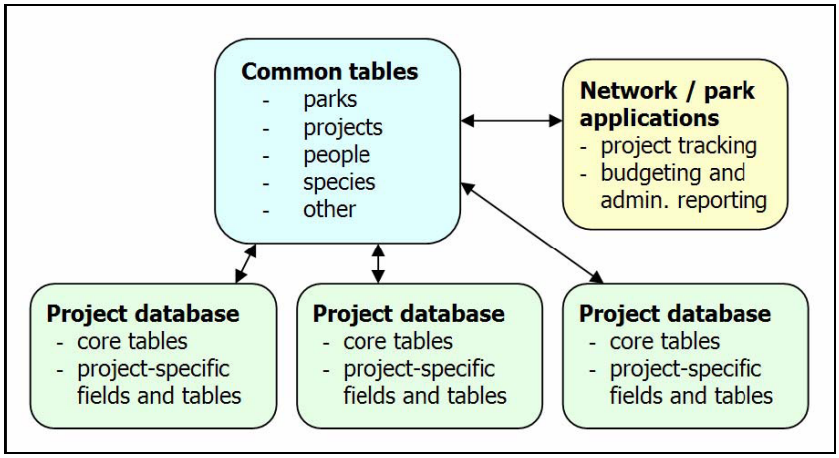


Figure 4.3. Common lookup tables and satellite databases.

At present, these common tables are grouped and maintained in separate Access (.mdb) files according to Table 4.1. Separating these tables by functional groupings is done primarily to reduce conflicts and performance losses associated with multiple users in MS Access. Databases associated with individual projects each access the common tables via links established in each project back-end data file.

4.3.4. LAN and WAN Infrastructure

Local-area network (LAN) and wide-area network (WAN) routers are monitored by the IT staff and are mostly invisible to employees. Though the logical structure is stable, the actual architecture may change from day to day. Any reduction in speed of performance should be reported to the IT Staff.

Table 4.1. Groupings for common lookup tables.

Grouping	Description
Parks	list of park units and networks
Projects	list of park and network projects, including inventories, monitoring, park-sponsored initiative projects, and external research projects
People	comprehensive list of contacts for parks and network, project-specific crewlists, lists of groups and users for tracking and managing access privileges
Species	comprehensive list of taxa for the park, linkage to NPSpecies taxonomic module, project-specific species lists
Other Lookups	lists of watersheds, drainages, place names, weather conditions, habitat attributes

4.3.5. IT Network Security

Information Technology security is managed by park, regional, and national DOI IT Specialists. GULN staff will comply with all required security training and procedures as advised by the USGS NWRC IT Specialists.

Over the past few years, enhanced security has been or is being developed. Security procedures are being directed not only service wide, but also from the Department of Interior. National solutions to security are currently under development and will be applied and enforced by the regional IT staff. These procedures will affect local access control to servers and their respective information.

4.4. Chapter Credits

John Boetsch (Northern Coastal and Cascades Network), Margaret Beer (Northern Colorado Plateau Network), Gordon Dicus (Pacific Island Network)

5. Database Design Strategy

5.1. NPS and Program Standards

Currently, there are no service-wide standards on database design. The I&M Program, however, recognized early on the need for database standards and detailed documentation. Database standards promote compatibility among data sets that will be aggregated and summarized in the future. 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
- Recommended Naming Standards
- Natural Resource Database Template Data Standards

Each database must ultimately meet the needs of the project managers. Considerations for these needs may include interactions with other agencies and ease of use, maintenance, integration, and customization. GULN will model and document databases as described in this DMP, will use the above recommendations in database design as a guideline where feasible, and will standardize attributes across databases where feasible.

5.2. Introduction to Data Models

Communication is a vital part of developing a suitable data design for individual projects. One mechanism for this communication is collaborative development of data models by project managers and data managers. Data models combine diagrams with associated descriptions. Data modeling is completed in three stages: conceptual, logical, and physical.

5.2.1. Conceptual Data Models

Conceptual data models (CDM) are constructed to graphically portray the processes specifically related to the implementation phase of a project – especially those that involve data acquisition, processing, QA/QC, and data reduction (Figure 5.1). These conceptual models are software-independent and free of database details, and instead focus upon capturing all of the information needed to accurately express the project data design. Conceptual data models are often created as the precursor to logical data models (LDMs), and may be discarded after the LDM is complete. For less complex projects, CDMs may be kept as alternatives to LDMs.
Conceptual Data Models should contain the following:

1. A short description in layman's terms of what is going to happen. Include key information to help put the database in perspective, such as environmental conditions while collecting, skill level of staff, etc.
2. A flow diagram of procedures, what information is needed and when, and what information is being collected or produced and when.
3. A description or mock-up illustration of how the data should be presented.

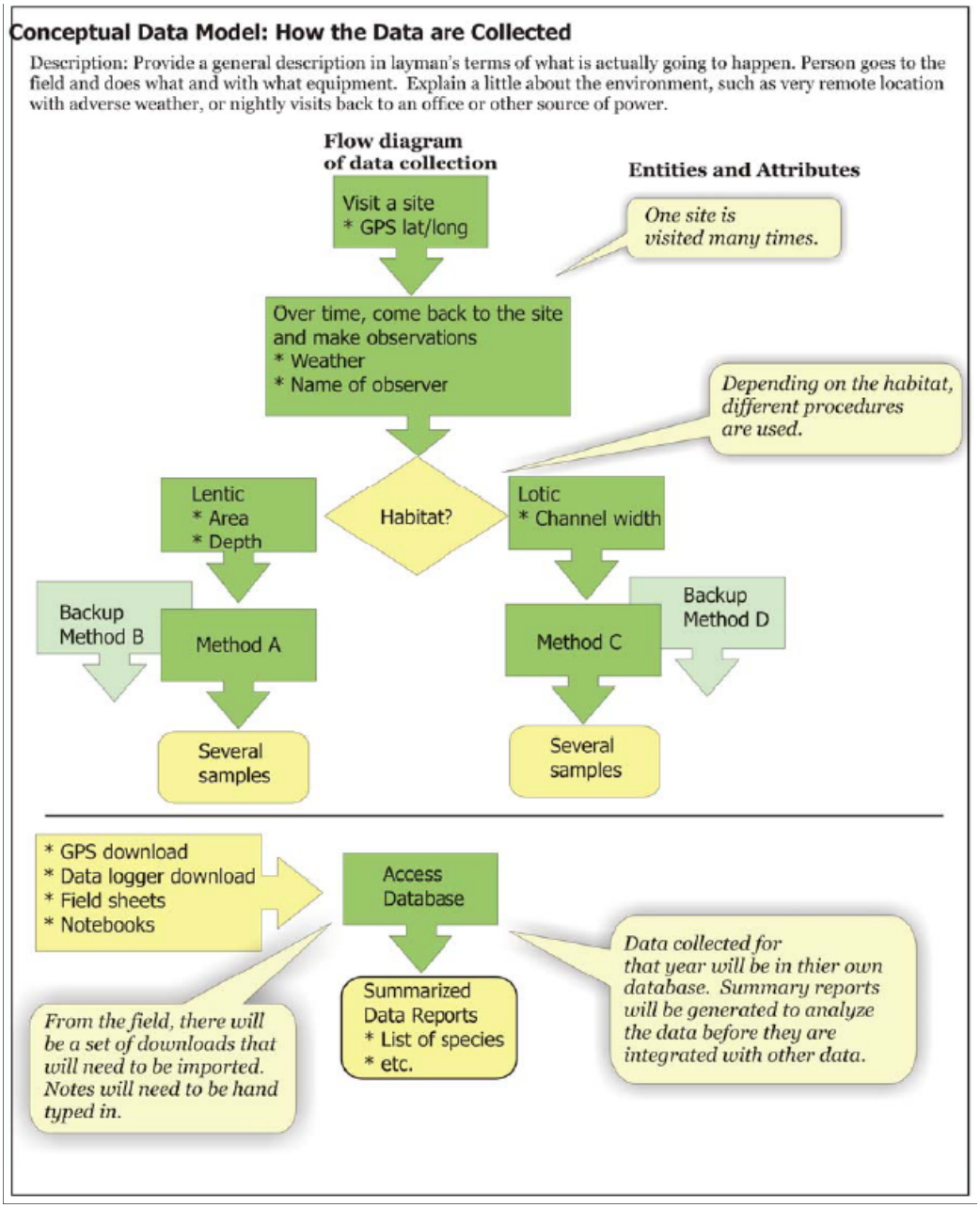


Figure 5.1. Schematic of conceptual data model.

5.2.2. Logical Data Models

A logical data model (LDM) is an abstract representation of a set of data entities and their relationship, usually including their key attributes (Figure 5.2). 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.

Logical data models are made up of five main components:

1. *Data entities* – distinct features, events, observations, and objects that are the building blocks of a data set, such as:
    - sample sites
    - sampling events
    - sampling units (transects, plots, etc.)
    - watersheds
    - species
    - habitat types
    - species observations
    - tissue specimens
2. *Entity attributes* – properties and rules of data entities, such as:
    - Sample sites have dimensions and a geographic position
    - Vegetation transects are 100 meters long
    - Temperature readings are recorded in Celsius to the nearest tenth degree
    - Elevation is recorded to the nearest foot, and cannot exceed 9,000 feet
    - Species abundance is recorded in terms of projected horizontal cover of all aboveground parts, as estimated by trained observers. Percent cover is estimated to the nearest whole number, ranging from 0 to 100%.
    - The degree to which vegetation obstructs the field of view around animal groups is classified in three categories: high (>75%), medium (25-75%) and low (<25%)
3. *Logical relationships* – how data entities are logically related
    - Each site will be visited numerous times
    - Each sampling event might have zero or numerous species observations
    - Each species can only be observed once per sampling event
    - Every sample must use one of three known sample methods
    - Every time a water sample is collected, temperature, pH and dissolved oxygen must also be measured
4. *Structural hierarchies* – the structure and order of relationships between data entities, which can be determined once the logical relationships are known
    - Site locations
        - ° sampling event
            - i. species observations
            - ii. water samples, temperature, pH, dissolved oxygen
5. *Views* – how the data will be viewed or what operations of the data will be performed
    - Summary list of bird species per park
    - Monthly average air temperature, wind direction, and precipitation

5.2.3. Physical Data Models

The physical data model (PDM) is used to design the actual database, depicting data tables, fields and definitions, and relationships between tables (Figure 5.3). Though the logical and physical data models are similar, the logical data model only provides enough detail to communicate the information to be stored in the database. The physical data model provides very specific details and definitions, such as primary keys and field types.

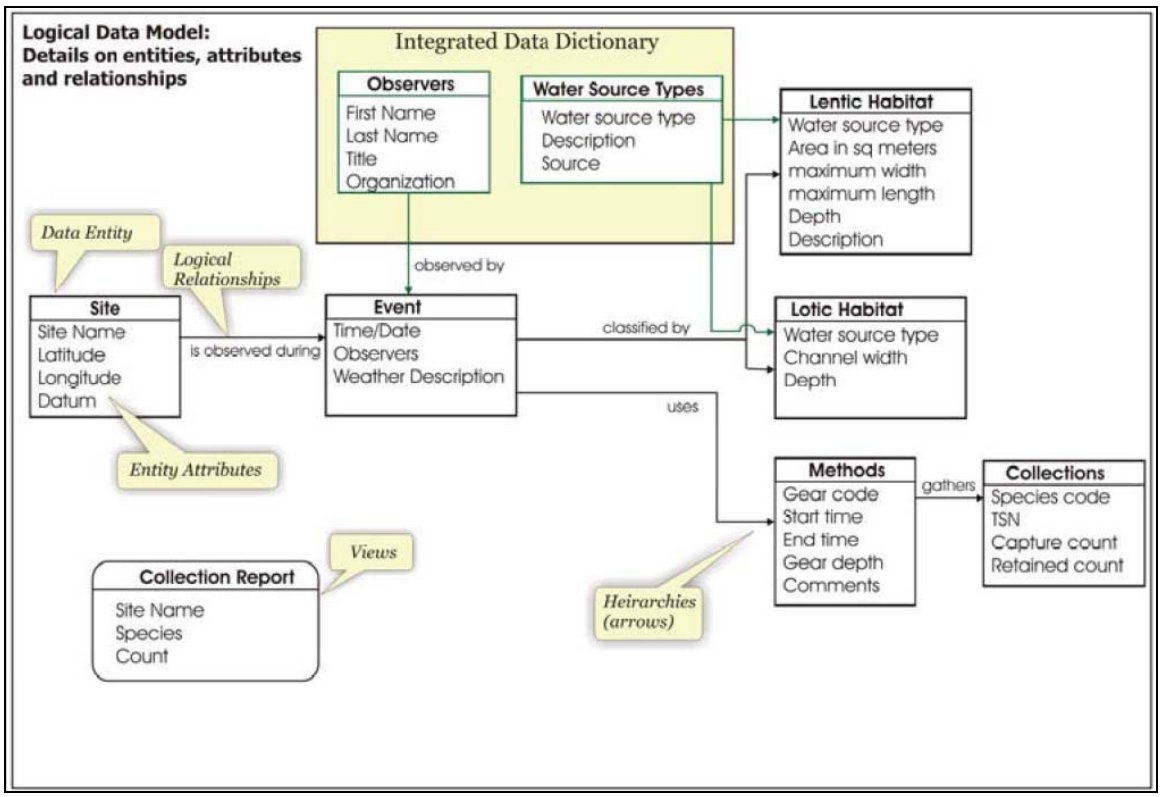


Figure 5.2. Schematic of logical data model.

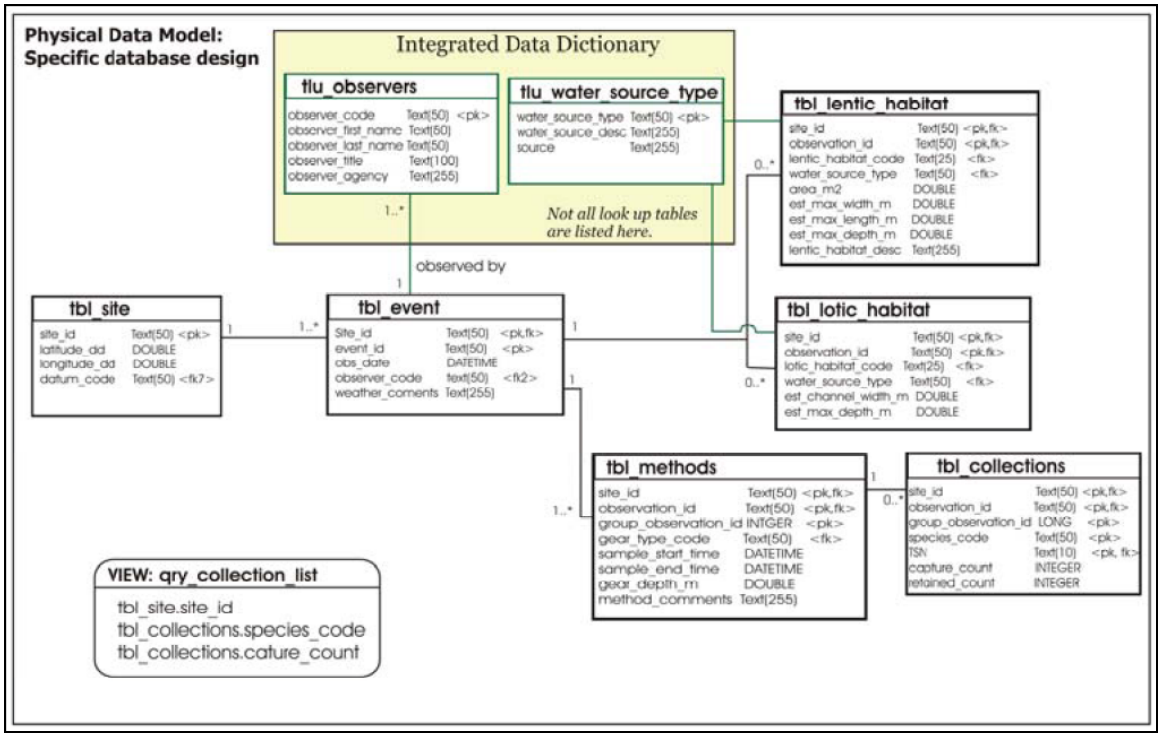


Figure 5.3. Schematic of physical data model.

5.2.4. Developing the Models

Four elements are required to develop the data models:

1. People – will need to be involved in the model development. Discussions may begin between a few, with review from a larger group or start with a large group working towards the specifics with a smaller group. Affected people may include the network coordinator, scientists, partners, field crews, biometricians, and data managers.
2. Protocols – The protocols for the vital sign will provide the greatest substance to the models. The protocols provide the goals, objectives, methods, standards, analysis, and reporting.
3. Reference materials – Reference materials such as field forms, drawings, mock-up reports, references to classifications to be used will play a significant role in the data models.
4. Frequent interactions – Frequent discussions between the project manager, the data manager, and others will be needed to develop successful data models. Detailed review of the protocols and reference materials will articulate the entities, relationships, and flow of information.

Data models are not flexible by themselves. It is how these techniques are applied that count. Not everything can be thought of from the start. To support change, data modeling should be iterative and interactive

5.3. 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 tables) tables and fields – Described in more detail below (Integrated Data Dictionary), these tables contain commonly used attributes, such as park code, 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.

The NRDT has been extensively reviewed by the I&M Program data managers and has been applied to some of the 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 GULN databases, its applicability towards vital signs monitoring may require specific mandatory tables to be reconsidered depending on the sampling methods and information collected.

5.4. Documenting and Integrating Databases

Database documentation is more fully described in Chapter 8 (Data Documentation).
Rather than developing a single, integrated database system, the GULN relies on modular, stand-alone project databases that share design standards. GULN will be developing a master look-up table database that will allow databases across projects and across disciplines to remain as consistent as possible. Likely standard attributes and values to be contained in this master look-up

table database include park codes, watershed codes, species identifiers, and vegetation or land type classifications.

5.5. Chapter Credits

John Boetsch (Olympic National Park) and Angie Southwould (Alaska Regional Office).

6. Data Acquisition and Processing

The National Park Service's Natural Resource I&M Program, in support of NPS' Natural Resources Challenge, is responsible for acquiring information required by park managers to properly manage and maintain the natural resources of their parks. To successfully accomplish this task, the GULN collects information from multiple sources and processes it so that it meets National and Network standards. This chapter describes the steps involved with acquiring data as well as the initial stages of data processing.
Steps for data acquisition are as follows:

- Data Discovery
- Data Harvesting
- Data Collection for Projects
- Data Collection for Remote Sensing
- Data Compilation, Processing, and Integration

6.1. National Standards

Executive Order 12906, Section 3 (d) states that "each agency must adopt internal procedures to ensure that the agency accesses the National Geospatial Data Clearinghouse before it expends Federal funds to collect or product new geospatial data, to determine whether the information has already been collected by others or whether cooperative efforts to obtain the data are possible".

6.2. GULN Network Standards

GULN will use the National Geospatial Data Clearinghouse for searching existing geospatial data, as stated above. In addition, GULN will use the internet, agency contacts, and other means to find non-spatial data that may benefit the Program.

6.3. Data Sources

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

1. 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.
2. Non-programmatic Data – includes data collected from NPS sources and data produced by external non-NPS sources.
    - Non-programmatic NPS Data – any data produced by the NPS that did not involve the inventory and monitoring program, such as park visitor use information.
    - Non-programmatic External Data – any data produced by agencies or institutions other than the NPS, such as USGS, NOAA, and various State agencies.

GULN vital signs monitoring will use a combination of both programmatic and non-programmatic data to meet its programmatic goals.

6.4. Data Discovery

Data discovery or data mining is the process of searching for existing data/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 that assist in the development of project methodologies and designs. The process involves reviewing many different sources for varying types of information. Data mining is an important part of any I&M project, with some vital signs monitoring projects depending largely on data collected by other agencies or institutions and harvested via 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.

6.4.1. Bibliographic/Literature

- National NPS Databases (e.g. NatureBib)
- Online literature databases (e.g. First Search or Biosis)
- Park archives through ANCS+

6.4.2. Geographic Data

- Regional centralized GIS data using NPS Theme Manager
- Federal Geographic Data Clearinghouse(s)
- Local, state, and federal government offices
- Regional GIS specialists

6.4.3. Biological/Natural Resources Data

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

Relevant information collected during a data discovery process is maintained at the Network either electronically and/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 datasets collected during this process are documented with FGDC compliant metadata.

6.5. Data Harvesting

The process of harvesting data from other sources should be standardized as much as possible. Considerations for harvesting are as follows:

- 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?

Vital signs monitored through data harvesting should address these questions in their data management protocols. Data harvested from another agency may not exactly match GULN data needs. 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 GULN accuracy standards. Project managers are encouraged to work with these agencies to incorporate GULN needs and standards as much as possible. Money can be saved and more efficient research can be done by improving existing systems.

6.6. Data Collection for Projects

Biological inventories and monitoring projects are the most common examples of projects conducted by the GULN I&M Program. The project manager is responsible for ensuring that data collection, data entry, verification, validation, storage, and archiving are consistent with GULN standards. Project managers will make use of GULN general standard operating procedures (SOPs) and guidance documents as applicable. In addition, protocol-specific SOPs will be developed to address data collection methods and procedures particular to a given protocol. 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 available for field data collection 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.

6.6.1. Field Forms

These are the most common method of recording field data. Use of formatted, project-specific data sheets are recommended, as opposed to using a field notebook. Field notebooks are important for entering additional notes and observations. It is recommended to use acid-free paper 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 will help ensure that all relevant information is recorded and subsequent data entry errors are minimized. Data sheets should contain as much basic preprinted project information as possible and sufficient space for recording relevant metadata such as date, collectors, weather conditions, etc. 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, copies of all original data sheets should be made and checked for legibility and completeness (i.e., no data cut off at the edges). The original datasheets should be used for data entry after which they will be stored in the fire proof cabinet. Copies of original datasheets should be archived offsite.

6.6.2. Tape Recorders

Handheld micro-cassette tape recorders are useful for recording field data. Recorded observations are subsequently transcribed to paper or directly entered into computer files. As with other technological solutions, there are drawbacks including battery and tape maintenance, low environmental tolerance, and risk of failure. However, if a single data collector is in the field, tape recorders can provide an easily operated, high quality, efficient method of collecting data. All audio tapes used for recording field data should be labeled appropriately (e.g. date, site, project) and stored in the fire proof cabinet. Analog audio cassettes degrade over time and are a media that is quickly becoming outdated and obsolete. If this is a desired method for field data collection, efforts should be made to transfer the audio data to a more permanent audio format such as CDs or MP3 files.

6.6.3. Cameras

Photographs provide an excellent visual record of field visits. Cameras are useful for capturing photo point records of long-term study sites. They also serve well for automated data collection by remotely recording information using web cams or trip cameras. The network will be using digital cameras. Specifications and requirements on digital camera types and management of digital photos are outlined in the GULN Photo Management Strategy, Appendix D.

6.6.4. Field Computers

These increase data collection and data entry efficiency by eliminating the need for paper field forms. Data can be downloaded directly from the field computers to the office desktops thereby eliminating the data entry step. Fewer chances for error because QA/QC checks can be built into the field database. Could be inefficient if copious amounts of notes or comments need to be recorded in the field.

Field computers, however, are subject to environmental constraints such as heat, dust, and moisture. 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. Printed data sheets should always accompany field teams on data collection trips in case the unit becomes inoperable in the field.

6.6.5. Palm-top computers (Personal Digital Assistants)

The small size and relative low cost of these devices make them attractive options for entering field data. Good for small field projects but not powerful enough for large data intensive field projects. PDAs can be ruggedized fairly easily and at a relatively low cost. Most run either Windows CE or Palm operating systems that may require additional processing/programming to transfer/create the database structure in the field units.

6.6.6. Tablet PCs

These 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 are harder to ruggedize than PDAs, but they are more powerful because these are fully functional PCs. Tablet PCs are ideal 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.

6.6.7. Automated Data Loggers

These are mainly used to collect ambient information such as air or water quality information. These devices can be calibrated and programmed to automatically record data and store them for later download directly to a computer, thereby eliminating the possibility for manual data entry errors. Must be properly calibrated and maintained so field crews must receive proper training and SOPs outlining the calibration/unit maintenance procedures. SOPs are currently in development and will be distributed to all field crews using these devices.

6.6.8. Permanently deployed devices

These are often cost prohibitive. Data from these devices must be retrieved and batteries changed on a regular basis. These intervals should be defined in the protocol.

6.6.9. Portable hand-held devices

These are deployed for sampling only during site visits. 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.

6.6.10. GPS Units

Two types of GPS units are often used during field work in GULN parks to collect location information.

6.6.11. Garmin Handheld Units

These are good for collecting general position information. Not recommended for obtaining high accuracy location information.

6.6.12. Trimble GPS Receivers

These are good for collecting highly accurate (submeter) location information.

6.7. Data Collection for Remote Sensing and LIDAR

Remote sensing technologies can be a powerful tool for characterizing and analyzing landscape data, as well as readily capturing data within areas of low accessibility. Considerations for selecting remote sensing imagery are as follows:

- Accuracy and resolution needed
- Frequency of measurement
- Costs
- Licensing for public use
- Ortho-rectified or not

Each remote sensing product is unique. It is imperative the user fully understands the product he or she is using. Products should be accompanied with well documented metadata. Imagery not rectified cannot be used for measuring distance or area, but may provide a low cost and timely overview of the landscape. Any protocols involving remote sensing should involve consultation with a professional remote-sensing specialist, and should consider the trade-off between accuracy and costs among different imagery sources (Table 6.1).

Although cost will be a primary decision factor, consistency among park units should also be a priority. Ideally, all imagery will be received in electronic and geo-referenced format.

The GULN's network GIS currently holds good coverage for all GULN parks in base cartographic layers, digital orthophoto quadrangles, and natural resource theme layers, as well as satellite imagery coverage for most of the parks. GULN will pursue cost-efficient options for acquisition of imagery in suitable formats as individual vital signs monitoring protocols identify additional remote-sensing imagery needs.

Table 6.1. Satellite resolution, swath width, area, cost, and history of various remote-sensing imagery sources.

Sensor	Multi-spectral Resolution (m)	Panchromatic Resolution (m)	Swath (km)	Area (sq km)	Cost/sq km ($)*	History
Quickbird [T]	2.4	0.7	16	272	29.73	2001-present
Ikonos [T]	4	1.0	11	121	27.03	1999-present
Spot5 [T]	10	2.5 and 5	60	3599	0.77	2002-present
Spot	66	10	60	3599	0.05**	1986-present
ETM+	30	15	185	34221	0.03	1999-2003
TM	30	NA	185	34221	0.03	1982-1999
MSS (ESTS)	79	NA	185	34221	NA	1972-1997
AVHRR	1100	NA	2700	7290138	NA	1978-present
MODIS	250	NA	2330	5429053	NA	1999-present

* cost uses the multi-spectral unit specifications

** cost for post-1998 imagery (pre-1998 imagery is half-price)

T requires tasking

6.8. Changes to Data Collection Procedures

Any changes to data collection procedures will be made based on valid reasons for altering the methodologies. Most issues will be identified during the design and testing stages of the protocols and changes will be implemented prior to the collection of field data. The protocol development process includes anticipating and addressing any foreseeable issues that might occur with data acquisition and processing. Unforeseen issues may arise after data collection has begun that require revision of procedures/protocols. Improvements in technology may also require changes to procedures. Significant changes to the protocols must be approved by the project manager, the GULN 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. Changes should be recommended if it is determined that the protocol has not achieved the desired results. All changes must be carefully documented within the SOPs and any associated databases.

6.9. Data Compilation, Processing, and Integration

All data, whether collected in the field by the I&M Program or downloaded from another organization, will need to be compiled and processed. This compilation should follow the data management guidelines developed for the specific data type (see Appendix C for GULN Project

Specifications). For example, tabular data will need to be entered or imported into a pre-approved MS Access database. Quality assurance and quality control (QA/QC) procedures are more fully described in Chapter 7.

After compilation and QA/QC is completed for data from a single field season or project milestone period, these data should be integrated with master datasets. Some extraction of information may be required. For example, species information may need to be extracted from various project databases for the purpose of updating the master NPSpecies database.

6.10. Chapter Credits

Geoffrey Sanders (National Capital Region Network).

7. Quality Assurance and Quality Control

The success of I&M networks is dependent on the quality of the data they collect, manage, and disseminate. Analyses performed to detect ecological trends or patterns require data that are recorded properly and have acceptable precision and minimal bias. Poor quality data can limit detection of subtle changes in ecosystem patterns and processes, can lead to incorrect interpretations and conclusions, and could compromise the credibility of the I&M Program. To ensure that GULN produces and maintains data of the highest possible quality, procedures have been established to identify and minimize errors at each project stage associated with the data life cycle (Figure 7.1). Data quality related to sampling design is covered in Chapter 4 of the GULN Monitoring Plan and in individual protocols, and is not discussed here.

7.1. NPS Standards

The National Park Service Director's Order #11B, "Ensuring Quality of Information Disseminated by the National Park Service," was issued in order to promote data quality (National Park Service 2002). It 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 an accurate, clear, complete, and unbiased manner within a proper context; and 2) *substance*, which focuses on ensuring accurate, usable, and reliable information.
>
> Utility refers to the usefulness of the information to its intended users.
>
> Integrity refers to the security of information; e.g., protection from unauthorized access or revision to ensure that the information is not compromised through corruption or falsification.

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

7.2. Quality Assurance and Quality Control Mechanisms

Palmer (2003) defines 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." He defines 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 introduces two fundamental types of errors into a dataset:

- Errors of commission include those caused by data entry or transcription errors, or malfunctioning equipment. They are common, fairly easy to identify, and can be effectively reduced up front 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.

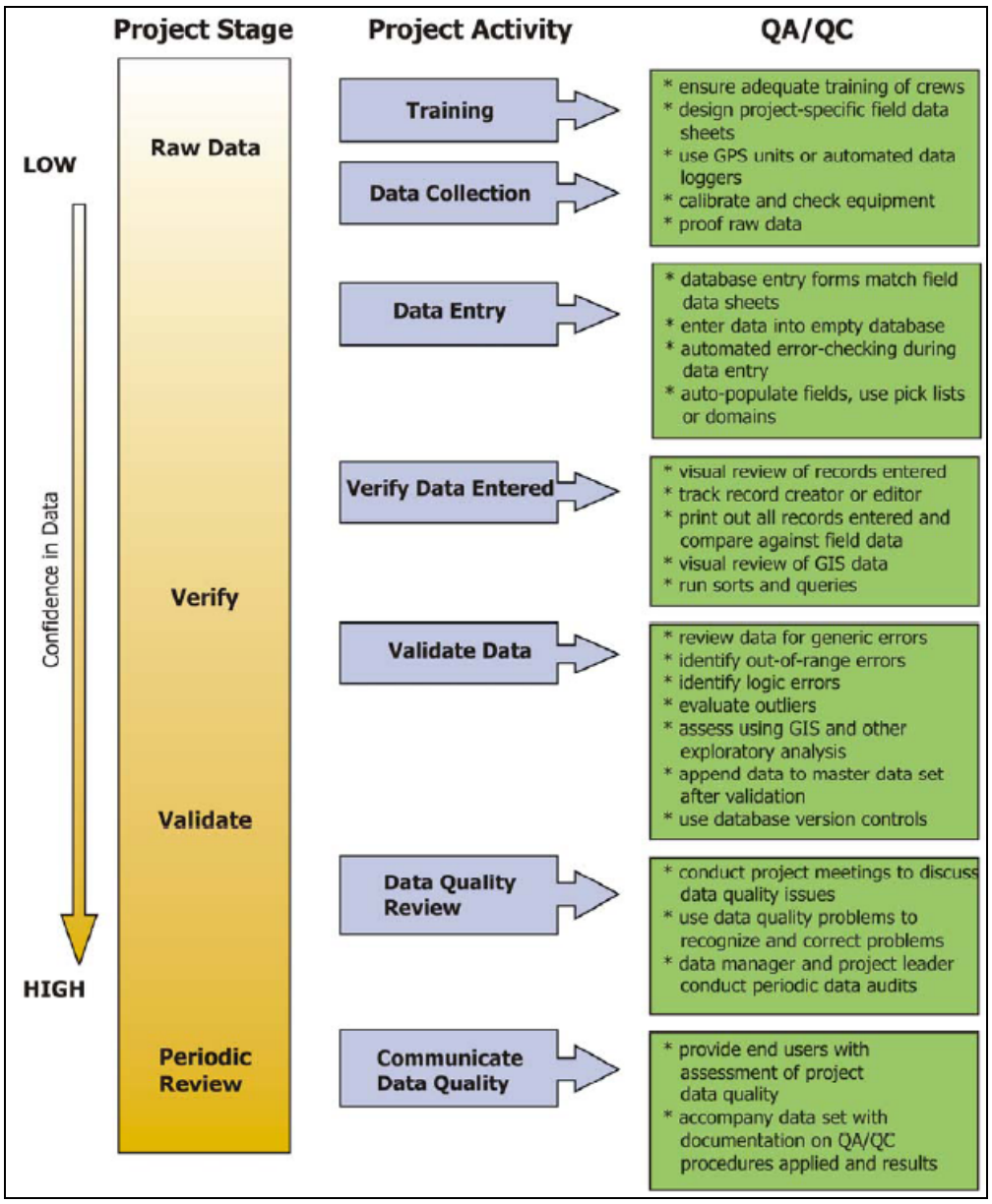


Figure 7.1. Schematic of the Quality Assurance/Quality Control procedures to be carried out during the project stages associated with the typical data life cycle.

QA/QC procedures applied to ecological data include four procedural areas (or activities), that range 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…), which begins with data design and continues through acquisition, entry, metadata development, and archiving. The progression from raw data to verified data to validated data implies increasing confidence in the quality of the data through time.

7.3. 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 essential that each member of the team have a stake in data quality, and is responsible for the quality of the results generated from his or her tasks, which are outlined below.
The data manager is responsible for:

- developing protocols and standard operating procedures (SOPs), in collaboration with the project manager, 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 the data quality program.

Project managers must:

- be aware of quality protocols and convey their importance to technicians and field crews
- ensure compliance with the protocols
- validate data after the verification process is complete
- review all final reports and information products.

Project technicians must:

- follow established protocols for data collection, data entry, and verification
- inform the project manager or data manager of quality-related problems or difficulties.

7.4. Quality Assurance/Quality Control Goals and Objectives

Although a data set containing no errors would be ideal, the cost and effort of attaining 100% accuracy would likely outweigh the benefits. Instead, two factors can be considered when setting data quality goals:

- the percent of entries that are incorrect (frequency of errors)
- the magnitude of the error (criticality of errors)

The significance of an error can vary depending on the data set and where the error occurs. A two-digit number with a misplaced decimal point is a significant error (e.g., 99 vs. 9.9). A four-digit number with an incorrect decimal value (e.g., 9999.99 vs. 9999.98) could retain an acceptable level of accuracy.

The most effective mechanism for ensuring that a project produces high-quality data is to determine procedures that direct project staff through accurate data collection, entry, and validation, and adhere to them. All monitoring projects undertaken by GULN will include a comprehensive set of SOPs that incorporate quality control in each stage of data collection and processing.

Although specific QA/QC procedures will depend upon the individual vital signs being monitored and must be specified in the protocols and SOPs, 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.

7.5. Data Collection Quality Assurance/Quality Control

Careful, accurate recording of field observations in the data collection phase of a project is the cornerstone for building a high-quality dataset. Unlike a typographical error that occurs when a recorded observation is incorrectly transferred from a paper field form to a digital database, an incorrect entry in the field is not easily corrected. Attention to detail during data collection is crucial to overall data quality.

Before the data collection phase of a project begins, the project manager and data manager determine data collection and storage protocols. Field sheets and field data recording procedures must be documented in the protocol SOPs and reviewed by the data manager. The project manager, in turn, ensures that field crews understand the procedures and closely follow them in the field. The project manager is responsible for training of field crews, with the data manager providing training assistance as needed and as time permits. Field technicians are responsible for proofing raw data forms in the field, ensuring their legibility, and verifying and explaining any unusual entries. They are expected to understand the data collection forms, know how to take measurements, and follow all procedures established in the protocol and SOPs.

7.5.1. Methods for Reducing Collection Errors

7.5.1.1. Ensure that field crews receive proper training

Protocols and SOPs alone cannot guarantee that high quality data will be collected. Adequate training is essential for field crews to understand and perform data collection procedures. A training manual may be helpful for long-term monitoring data collection efforts, and for those projects that will involve a large number of field staff, especially if staff turnover is anticipated.

7.5.1.2. Use a formatted, project-specific data sheet as opposed to a field notebook

Standardized data sheets that clearly identify the data to be recorded, and that reflect the design of the computer data entry interface, will help ensure that all data are recorded and entry errors are minimized. Acid-free paper is required to prevent degradation and subsequent data loss.

Field notebooks are valuable for recording additional observations, however, and are encouraged for this purpose. It is recommended to keep a daily journal and other peripheral information in this notebook. These notebooks should be archived and stored with the datasheets.

Data sheets should contain as much preprinted project information as possible, and include essential metadata such as the name of the data collector and date. Data sheets should clearly specify all required information, using examples where needed to ensure that proper format is used.

All data added to the data sheet must be printed and clearly legible. If alterations to the data are necessary, the original data should be crossed out with a single line and the new data written next to the original entry. Data should never be erased or overwritten.

After data entry, verification, and validation, copies of all original data sheets should be made and checked for legibility and completeness (i.e., no data cut off at the edges). The copies of the data sheets will be stored as specified in the protocol SOP, and the original data sheets should be archived.

7.5.1.3. Use an electronic device for data collection whenever possible

The use of handheld devices such as GPS units minimizes the need for manual data entry from field forms and associated transcription and data entry errors. Specially designed data dictionaries can be developed to fit project requirements and can incorporate on-the-spot QA/QC checks. Electronic devices are not a substitute for data hand-written on field sheets; rather, they are a tool to make subsequent database entry more efficient.

When electronic devices are used for data collection, data files should be downloaded daily to avoid potential loss of information. Thus, if a unit fails during data collection, only the current day's data are lost. Batteries should be checked prior to a data collection trip, and they should be charged at the end of every field day.

7.5.1.4. Use automated data loggers where appropriate

Instruments with their own data acquisition systems are useful for collecting some types of data, such as water and air quality data. These devices can be calibrated and programmed to automatically record data and store them for later download directly to a computer, thereby eliminating the possibility for manual data entry errors. Data loggers are an efficient method for recording continuous sensor data, but routine inspections are necessary, and environmental constraints, as well as power (e.g., sufficient battery charge) and maintenance requirements, are potential pitfalls when using these instruments. Regular downloads are required because physical memory is usually limited.

7.5.1.5. Consider calibration, maintenance, and training requirements of field equipment

Accurate field measurements are possible only if field equipment is regularly calibrated and maintained. Once in the field, allow sufficient time for field equipment to adjust to its environment so it will record accurate measurements (for example, when using water quality probes and GPS units). Researchers should maintain records of equipment calibration and equipment failures as an integral part of their field data.

7.5.1.6. 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.

7.5.1.7. Perform quantitative assessments of data quality

Repeating measurements is the primary tool for performing quantitative assessments of data. Project managers should periodically review the work of field technicians to ensure that their work does not drift from standards during the course of the field season.

7.6. Data Entry

Data entry is the process whereby raw data are transferred from paper field forms into an electronic data format. When data are gathered or stored digitally in the field (e.g., on a data logger), data entry consists of the transfer of data (downloading) to a file that can be moved into database tables.

Data entry should occur as soon as possible after data collection is completed, or as an on-going process during long projects, and by a person who is familiar with the data. The primary goal of data entry is to transcribe the data from paper records into the computer with 100% accuracy, although errors are unavoidable during data entry. Thus, all data are checked and corrected during the data verification process (see below).

The project manager, along with the GULN data management staff, provides training in the use of the database to all data entry technicians and other users. The project manager makes certain that data entry technicians understand how to enter data and follow the protocols. Data entry technicians are responsible for becoming familiar with the field data forms, the database software, database structure, and any standard codes for data entry used by GULN.

7.6.1. Methods for Reducing Data Entry Errors

7.6.1.1. Enter or download data in a timely manner

All data should be entered or downloaded into the project database as soon as possible after collection. Data entry should not be delayed until all the project data have been collected.

7.6.1.2. Design efficient data entry forms

A full-screen data entry form that mimics the field data forms can reduce manual data entry errors due to the 1:1 correspondence of the attributes.

7.6.1.3. Design data entry methods that distinguish between newly-entered and previously validated data

New records will be entered into an empty database. These records will be appended to the master data set only after formal verification and validation has been completed. When this procedure is not practical, a field in the database identifying the status of validation and verification for each record will be completed. (The process for validation and verification will be detailed in the data management SOPs associated with the project.)

7.6.1.4. Track record creation and edit details in the database

Fields that store the date a record is created or modified, and the initials of the person creating or modifying it, increases the level of personal responsibility for the accuracy of data entry. This feature also allows the project manager or data manager to determine if error patterns can be traced to a particular person, and follow up with additional staff training to correct the problem.

7.6.1.5. Build automated error checking features into the database

QA/QC measures for data entry will be built into the database design to perform automatic validation checks of data. Database entry forms can reduce transcription errors through auto-filled fields, range limits, pick lists, and spelling checks. These forms can also provide controlled access to the database (i.e., forms are set for data entry only, which prevents accidental deletion or alteration of existing data) and can control the sequence of data entry (i.e., certain fields require an entry before more information can be entered). Error messages can alert the operator when mistakes are made and require correction.

- *Auto-filled fields*. Whenever possible, the data in a field should be autofilled by the computer. For example, if a location ID is comprised 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.

7.6.1.6. Provide a clean, organized work environment

Desktop space near the computer should be free of clutter and distractions. 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 should also be available for making notes.

7.6.1.7. If possible, use two data entry technicians for data entry

When one technician reads the data from the field data forms and another enters them into the computer, the work is often faster and results in a lower error rate. Alternatively, one technician enters the data, while the second technician reenters the data (or a subset of the data). The two datasets are compared for errors. If only one person is available, he should work at a slower pace to avoid errors.

7.7. Verification and Validation Procedures

Data verification checks that data entered into a secondary (i.e., electronic) format match the source data, whereas data validation checks that the data make sense. Although data entry and verification can be conducted by personnel with a general familiarity of the data, the validation process requires in-depth knowledge about the project and data collected.

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

7.7.1. Methods for Data Verification

Data verification immediately follows data entry and involves checking the accuracy of the computerized records against the original source (usually hard copy field records), and

identifying and correcting any errors. GULN procedures are to verify all records entered against their original source. When the computerized data are verified as accurately reflecting the original field data, the paper forms are set aside for use later in the data validation process.

The following five verification methods will be used by GULN for all project data:

1. *Visual review at data entry*
   The data entry technician verifies each record after input and immediately corrects any errors. 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 entered during a data entry session are printed in a format that closely matches the original data source. Each data element on the printout is compared with the original values from the hard copy, preferably by a second person who did not perform the data entry. Errors are clearly marked and corrected in the database as soon after data entry as possible.
3. *Sorting and summary queries*
   Each project will have queries that can be run to detect broad errors such as inconsistent, duplicate, omitted, or unlinked records.
4. *Visual review of spatial data*
   Any spatial data that are collected as part of the project will be converted to GIS and visually inspected for accuracy (e.g., points located outside park boundaries, upland locations occurring near sea level).
5. *Duplicate data entry*
   The data entry person completes all data input, 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 query is run to automatically compare the duplicate records from the two datasets and report on any mismatches of data. These disparities are manually reviewed and corrected if necessary. This method involves the overhead of retyping the selected records, as well as the creation of a comparison query (which requires additional effort, but is not time-consuming). This method becomes increasingly successful as the value of **n** decreases.

Each method 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.

7.7.2. Methods for Data Validation

Although data may have been correctly transcribed from original field notes or forms, they still might be inaccurate or illogical. For example, values may be outside the physical range of possible values, or may fall outside the logical range of reasonable values. The process of reviewing computerized data for range and logic errors is called *validation*, and it can accompany data verification *only* if the reviewer has comprehensive knowledge about the data. Validation is a separate operation carried out *after* verification by a project specialist who can identify generic and specific errors in particular data types.

Invalid data commonly consist of slightly misspelled species names or site codes, the wrong date, or out-of-range errors in parameters with well defined limits (e.g., elevation). But more interesting and often puzzling errors are detected as unreasonable metrics (e.g., stream temperature of 70°C) or impossible associations (e.g., a tree 2 feet in diameter and only 3 feet high). These are *logical errors*. The discovery of logical 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 descriptive statistics can reveal possible logic and range errors.

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 to the field records 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 because field sheets or other sources of data could easily be overlooked.

General step-by-step instructions are not possible for data validation because each data set has its own unique contents and domains. Specific procedures for data validation will be delineated in data management SOPs. However, the following general methods can be used as guidelines:

7.7.2.1. Data entry application programming

Certain components of data validation are built into data entry forms. The simplest validation during data entry is range checking, such as ensuring that a user attempting to enter a pH of 20.0 gets a warning and the opportunity to enter a correct value between 1.0 and 14.0 (or within a narrow range appropriate to the study area). 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 is important.

Edwards (2000) suggests the use of 'illegal data' filters, which check a specified list of variable value constraints on the master data set (or on an update to be added to the master) and creates an output data set. This output data set includes an entry for each violation, along with identifying information and an explanation of the violation.
A caveat should be interjected regarding the operative word 'illegal.' Even though a value above or below a given threshold has never before been observed and the possibility that it could occur seems impossible, such an observation is not always an illegal data point. Edwards (2000) points out that one of the more famous data QA/QC blunders to date occurred when NASA's computer programs deleted satellite observations of ozone concentrations that were below a specified level, seriously delaying the discovery of the ozone hole over the South Pole.

7.7.2.2. 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' unusual.

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 perhaps a better way to explain this quality assurance goal. If contamination is not detected during data collection, it will be detected later only if an outlying data value results. When we detect an outlier, we should try to determine if some contamination is responsible.

GIS, database, graphic, and statistical tools for ad-hoc queries and displays of the data can be used 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 the documentation of the data set saves other users from checking the same unusual values.

7.7.2.3. 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. Normal probability plots, Grubb's test, and simple and multiple linear regression techniques may also be used (Edwards 2000).

7.8. Version Control

Version control is the process of managing copies of changing files over the course of a project. Change includes any alteration to the structure or content of the files. Good version control practices allow for full recovery of a dataset as it existed prior to changes being made. Before making any major changes to a file, a copy of the file should be saved with a unique file name. This process is particularly important when appending newly-validated data to a master dataset, or when upgrading a database to a new version. In addition, proper controls and communication are required to ensure that only the most current version is used for data entry or analysis.

Specific naming conventions and directory structures related to version control are detailed in individual monitoring SOPs. The data manager will determine the version control method that will be used, and other network personnel are responsible for accurately designating versions for any files with which they have worked.
These general version control guidelines apply to files maintained on the GULN file server. They do not apply to enterprise-level server applications such as the I&M Program's service-wide Oracle databases; updates to these master databases are accomplished either through online data entry/edit or through file submission to I&M Program WASO staff. General file naming guidelines are as follows:

<project_file>_<QA/QC_status>_<date>.<ext>
Where:
<project_file> = the name of the established base file name
<QA/QC_status> = the status of the information. For example, "_RAW_"; "_VERIF_"; "_VALID_"
<date> = The date of the file, as YYYYMMDD.
<ext> = the file extension, such as .mdb, .xls

7.9. Data Quality Review and Communication

Edwards (2000) suggests regular meetings of project managers, the data manager, and data management personnel for discussing data quality problems and issues. Participants become more aware of quality issues and learn to anticipate problems. Moreover, all participants realize their role in data quality and the entire monitoring effort.

7.9.1. Value of Feedback from QA/QC Procedures

Quality assurance procedures may need revision to improve quality levels if verification and validation processes reveal an unacceptable level of data quality. Quality checks should not be

performed with the sole objective of eliminating errors, as the results may also prove useful in improving the overall process. For example, if the month and day are repeatedly reversed in a date field, the data entry technicians may require retraining about the month/day entry order. If retraining is unsuccessful in reducing the error's occurrence, the computer program may need to be rewritten so that month and day are entered separately, field length limits are enforced, or a pick list is created. In this manner, the validation process will serve as a means of improving quality.

Field data forms can be modified to avoid common mistakes or logical errors. Often minor changes, small annotations, or adding check boxes to a field form can remove ambiguity about what to enter on the form. When the same type of validation error occurs repeatedly in different data sets, the field form–not the field crew–is usually at fault. Repeated errors found during validation can also mean that protocols or field training are at fault, which can then be recognized and corrected.

7.9.2. Monitoring Conformance to Plans and Standards

The data manager will perform periodic data audits to help maintain and improve GULN's data quality. The audits will verify that staff is adhering to data quality procedures specified in this plan and the protocol-specific data management plans, and will track and facilitate the correction of any deficiencies. These quality checks promote a cyclic process of continuous feedback and improvement of the both the data and quality planning process.
Audits include verification of the following:

- 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

The results of quality assessments are documented and reported to the research staff and the network coordinator. The project manager and coordinator are responsible for ensuring that non-conformities in data management practices are corrected.

7.9.3. Communicating Data Quality

The GULN will use data documentation and metadata to notify end users, project managers, and network management of data quality. A descriptive document for each dataset/database will provide information on the specific QA/QC procedures applied and the results of the review. Descriptive documents or formal FGDC-compliant metadata will document quality for spatial and non-spatial data files posted on the Internet.

7.10. Chapter Credits

Debbie Angell (Sonoran Desert Network) and Margaret Beer (Northern Colorado Plateau Network) Dorothy Mortenson (Southwest Alaskan Network).

8. Data Documentation

Data documentation is the most important step toward ensuring that data sets are useable well into the future. Data longevity is roughly proportional to the comprehensiveness of their documentation (Michener 2000). The term metadata is defined as information about the content, context, quality, structure, accessibility, and other characteristics of data. In addition to ensuring data longevity, metadata increase the possibility of data sharing and reuse for multiple purposes. Creating and maintaining comprehensive metadata is neither a simple nor quick process. It requires an up-front time investment for planning and organization, and an ongoing investment to keep it current.

While the importance of metadata is widely accepted within the data management community, the approaches for storing this information and the levels of detail can vary. However, some established metadata strategies apply to NPS data, including the following:

- Executive Order 12906, signed by President Clinton in 1994, mandates federal agencies to "...document all new geospatial data it collects or produces, either directly or indirectly..." using the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM). In addition, EO 12906 directs agencies to plan for legacy data documentation and provide metadata and data to the public.
- The FGDC Biological Data Profile contains all the elements of the CSDGM plus additional elements for describing biological data sets. Metadata created in compliance with the Biological Data Profile can be added to the National Biological Information Infrastructure (NBII) Clearinghouse. Although not a requirement, completion of the Biological Data Profile for appropriate data sets is recommended by NBII.
- The NPS Geographic Information System (GIS) Committee requires all GIS data layers be described with FGDC standards and the NPS Metadata Profile.

8.1. NPS Integrated Metadata System Plan and Tools

Of the numerous tools available for developing metadata, the NPS Integrated Metadata System Plan recommends two desktop applications for storing metadata. These include Dataset Catalog (developed by the national I&M Program) and ArcCatalog from ESRI. These tools are briefly described, below.

8.1.1. NPS Metadata Tools and Editor

Metadata Tools and Editor (MTE) is a custom software application for authoring, editing, and managing National Park Service (NPS) metadata. The MTE operates either as an extension to ArcCatalog versions 8.3/ 9.x or as a standalone desktop application. The MTE is intended to be the primary editor for metadata that will be uploaded to the **NPS Data Store** and produces XML metadata based on the NPS Metadata Profile. Metadata editing is accomplished with editing stylesheets that transform an XML metadata file into an editable record.

8.1.2. ArcCatalog

ArcCatalog is contained within the ArcGIS suite of applications and supports several metadata standards that allow users to create, edit, and view information about the data. Metadata within ArcCatalog are stored exclusively as XML files. The NPS Integrated Metadata System Plan recommends ArcCatalog for gathering GIS-integrated geospatial metadata. GULN will be using

the MTE imbedded within ArcCatalog to generate spatial metadata and for uploading to the NPS Data Store.

8.2. Metadata Process and Work flow

Data used or documented by GULN can be grouped into three broad categories based on data origin: GULN project data; other NPS data; and data from external (non-NPS) sources. The level and extent to which metadata can be completed varies depending on these data categories, as does the work flow process.

- GULN Project data – These are projects undertaken by the GULN with metadata considerations beginning at the onset of the project. Complete and compliant metadata are required for each project.
- Other NPS data – This category includes natural resource-related data sets typically obtained during the data mining process at parks. In many cases, legacy datasets are missing pertinent information, including the originator who may no longer be in contact. While the desirable level of documentation may not be possible, all available data and supporting documentation are assembled and reviewed. Ongoing resource management projects in network parks are the source of many valuable natural resource-related data sets. While many of the data sets from these projects have associated documentation or knowledgeable project managers who are still at the parks, well-structured and compliant metadata are frequently not created or maintained.
- External Data - Other agencies and organizations gather data that are relevant to GULN or park projects. GULN will obtain, to the best extent possible, all available metadata with these datasets. As with legacy data, gaps may exist in non-programmatic data documentation.

8.2.1. NPS Metadata Tools and Editor Use

These records can be imported into the online NPS Data Store System. As a general rule, if a data set exists primarily in tabular format (e.g., MS-Access database, Excel spreadsheet, assemblages of paper or other materials), Dataset Catalog is used as the metadata tool. If a data set exists primarily in spatial format (e.g., GIS coverage or shapefile of trails, boundaries), ArcCatalog with MTE extension is used (see below).

GULN uses MTE to develop metadata for all data sets as XML or TXT files that are named the same as the file being documented with _metadata. For example, the NPSpecies file for VICK mammals named VICK_NPSpecies_Mammal_Certified_010505.mdb would have a metadata file accompanying it names VICK_NPSpecies_Mammal_Certified_010505_metadata.xml.

8.2.2. ArcCatalog Use

ArcCatalog is the tool used by GULN and its network parks for creating FGDC-compliant spatial metadata. All spatial metadata will be written using ArcCatalog (and the NPS Metadata Tools extension), and will be stored as either .xml or .txt files.

8.2.3. Making Information Available

MTE and ArcCatalog metadata will be submitted to the NPS Data Store. Metadata records may be withheld from public posting, or may be abridged, if their content is classified as sensitive and their release could potentially jeopardize a protected resource (see Chapter 9, Data Dissemination).

8.2.4. Updating Metadata Records

For GULN projects, the project manager is required to keep the data manager and GIS specialist informed of any data or format changes so that associated metadata can be verified and updated.

8.3. Vital Signs Protocol Documentation

8.3.1. Master Version Table

Vital signs protocols (protocol narrative and accompanying SOPs) constitute essential project documentation that must accompany the distribution of monitoring data. Over time there will be instances when the protocol narrative and SOPs will need to be updated. Narrative and SOP updates may occur independently. That is, a change in one SOP will not necessarily invoke changes in other SOPs; a narrative update may not require SOP modifications.

Every protocol and SOP will have version numbers in a Master Version Table (MVT) at the beginning of the document. The MVT contains a Version Key Number that designates the narrative and SOP versions that are in use at a specific time (Table 8.1). Every protocol contains an SOP entitled "Revising the Protocol Narrative and SOPs." The MVT is contained in this SOP and is required to be updated when any protocol revisions are made. The protocol narrative, SOPs, and data will not be distributed independently of the MVT. Incorporation of the MVT into individual protocol databases will require active database links that assign the SOP version number(s) to each sampling event, so that data records can be associated with a specific SOP version.

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).

Version Key #	Version Key Date	Narrative	SOP #1	SOP #2	SOP #3	SOP #5	SOP #6
VK1	12/15/2004	1.00	1.00	1.00	1.00	1.00	1.00
VK2	5/10/2005	1.00	1.01	1.01	1.00	1.01	1.00
VK3	11/18/2005	1.00	1.01	1.01	1.01	1.02	1.00

8.3.2. Additional Vital Signs Metadata

Long-term monitoring projects present a different set of metadata questions and requirements that may extend beyond the scope of the project tracking database, established FGDC standards, or a descriptive document for a tabular database. Essential documentation such as algorithms, output files, or spatial analyses may reside in different systems and formats, and could potentially be overlooked when distributing or applying the data. Depending on the project, documentation may need to include details on data models or algorithms used, procedures for data synthesis, and associated input and output files. Data use and data request histories, and secondary research or publications resulting from long-term monitoring projects, may also need to be tracked.

As the protocols are developed, vital signs documentation will be tested and will evolve to combine metadata needs and ease of use.

8.4. Chapter Credits

Teresa Leibfreid (Cumberland Piedmont Network), Bill Moore (Mammoth Cave National Park), Doug Wilder (Central Alaska Network), and Margaret Beer (Northern Colorado Plateau Network).

9. Data Analysis and 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. Chapter 7 (Data Analysis and Processing) of the GULN Monitoring Plan contains the background and overall approach to data analysis and processing for our network. 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. This chapter discusses data management activities related to using GIS and database application software for data summary and analysis, and to prepare data for analysis using statistical software applications.

9.1. Periodic and Annual Reporting

The data manager will work with the project managers, network ecologist, 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 SAS (commercial) and R (freeware). Some basic summary and reporting functions required by a vital sign monitoring protocol can be developed within Microsoft Access database applications where data are stored. Examples include descriptive statistics (mean, standard deviation, sample size). 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 network and/or park affiliate staff. Ad hoc queries and reports will be handled on a case by case basis due to their dynamic nature. GULN will enhance its web site over time to deliver reports and provide supplemental background data and information.

9.2. Long-term Trends and Analyses

Most long-term data analysis will involve statistical software applications, for which required data formats often involve arrays of binary or discrete values that represent one or more parameters. 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. Methods may be developed to visualize time-series data, perform geostatistical functions, and do spatial network analysis with hydrography features. Tabular and spatial results can be shared in reports and made available on the GULN web site and will be posted to the NPS Data Store.

10. Data Dissemination

GULN data management 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

The GULN will continue to provide links to GULN public data products via the GULN website and the NPS Data Store. These will be the network's portal for the most current data distribution as the Monitoring Program develops. Distribution instructions for each dataset will be provided in the respective metadata.

10.1. Mechanisms for Distribution

GULN data products (final deliverables or periodic milestones) will be "packaged," where possible, and made available for distribution as a complete set. This "package" is similar to the materials consolidated for natural history archiving and curation (See Chapter 11). 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 (Table 10.1).

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. 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).

Biodiversity Data Store – a digital archive of documents, GIS datasets and non- GIS dataset files that document the presence/absence, distribution and/or abundance of taxa in NPS units. (http://www.nature.nps.gov/biology/biologicalinventories/biodiversitydatastore.cfm)

NPS Data Store – an online repository for spatial and non-spatial metadata and associated data products. (http://science.nature.nps.gov/nrdata/)

GULN 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. (http://www1.nature.nps.gov/im/units/guln/index.cfm)

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 GULN oversees the use of NPSTORET according to the network's integrated and regulatory 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. (http://www.nature.nps.gov/water/infoanddata/index.cfm)

Table 10.1. Primary repositories for GULN information and associated specimens.

Item	Repository
Reports (public) digital	NatureBib, NPS Data Store, GULN Data Server
hard copy	Park and network libraries, park archives
bibliography	NatureBib, Reference Manager
Network-generated digital datasets and data products (public, non-sensitive) • certified data and data products (including photos) • metadata	NR-GIS Data Store, GULN website, NPSpecies, Biodiversity Data Store, NPSTORET, GULN Data Server
Network-generated digital datasets and data products (NPS staff, sensitive) • raw, validated, and analyzed data • metadata • submitted reports • digital photos • digital presentations	GULN intranet website; selected vital sign data may be housed externally with an established MOU; GULN Data server
Project product materials • specimen vouchers • photograph film	Park archives, Louisiana State University Museum, or other curation facility (according to park or project protocol)
Project administrative records or miscellaneous items (hard copy and digital)	GULN office, GULN Data Server

10.2. Ownership, FOIA, and Sensitive Data

GULN products are considered property of the NPS (OMB, Circular A-110, Section 36). The Freedom of Information Act, 5 U.S.C. § 552, referred to as FOIA, stipulates that the United States Government, including the NPS, 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 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. Information regarding the following resources is recognized as sensitive by the NPS:

- Endangered, threatened, rare, or commercially valuable National Park System 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 deemed 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. Network staff will classify sensitive data on a case-by-case, project-by-project basis and will work closely with project managers s 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.

The following are suggested guidance for determining 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 park resource 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 that 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?

10.2.1. Access Restrictions on Sensitive Data

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
- Any information classified as protected should not be released in any format except

as approved in advance by the NPS

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

When preparing information for any repository, 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 or research partners. However, limits to how these data are subsequently released must also be clearly defined. It is crucial that the person uploading records to the online applications (repositories) is familiar with the procedures for identifying and entering protected information.

10.2.2. NPS Only versus Public

Only data subjected to complete quality control are released, unless 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 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 are still valuable in-house. Provided these do not contain sensitive data as described in this chapter, these data may be released to the public upon specific FOIA request. They must be accompanied with qualifying documentation. I&M Program applications, such as NatureBib and NPS Data Store, provide a means to flag these data as "NPS Only" or "Public". Additional notes describing the quality should be added to the records in these applications, as resources permit.

10.3. Feedback Mechanisms

Overall comments and questions concerning GULN I&M Program are welcome at any time and may be submitted via e-mail or telephone to the Network Coordinator. The GULN 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.

10.4. Chapter Credits

Sara Stevens (Northeast Coast and Barrier Network) and Wendy Schumacher (Washington Support Office) Dorothy Mortenson (Southwest Alaska Network).

11. 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 properlyinterpreted 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.

11.1. Digital Data Maintenance

In general, digital data maintained over the long term will be one of two types: short term data sets, for which data collection and modification have been completed (i.e., inventory projects); and long-term monitoring data sets, for which data acquisition and entry will continue indefinitely.

Following the lead of the NPS and the national I&M Program, GULN has adopted MS-Access as its database standard and ArcGIS as its spatial data management standard. GULN will remain current and compatible with NPS or national I&M version standards for these software programs. Technological obsolescence is a significant cause of information loss, and data can quickly become inaccessible to users if stored in out-of-date software programs or on outmoded media. Maintaining digital files involves managing the ever-changing associated 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 (Vogt-O'Connor 2000). As software and hardware evolve, data sets must be consistently migrated to new platforms, or they must be saved in formats that are independent of specific platforms or software (e.g., American Standard Code for Information Interchange (ASCII) delimited files).

Any data set for which data entry or updates is still occurring will be stored in subdirectories under the "I&M" directory on the GULN server (e.g., I&M\vital_signs). The "Archive" directory and subdirectories are reserved for data sets that will no longer change.

11.1.1. Short-term data sets

For short-term data sets created or managed by GULN, upon project finalization a set of ASCII comma-delimited text files will be created for each data table comprising the data set. These files will be accompanied by a readme.txt file that explains the contents of each file, file relationships, and field definitions. The ASCII files are in addition to the native version of the dataset (typically in database or spreadsheet format). Creating these text files will help ensure the data are usable in a wide range of applications or platforms.

In addition to creating ASCII files, GULN will also update completed and archived data sets that may be in older versions of MS-Access, with the goal of having no dataset more than two versions behind the current version used by GULN. There is the risk of losing a certain amount of

performance in the process of conversion; for example, complex data entry forms or reports may not function properly in an upgraded version. To the extent possible, proper functionality of data entry forms and reports will be maintained; however, the priority will be to ensure basic table and relationship integrity. All previous versions of the data set will be saved.

11.1.2. Long-term monitoring data sets

Long-term monitoring data sets require regular updates and conversion to current database formats. All active or long-term databases will conform to the current NPS and I&M software version standards.

Monitoring projects will have variable long-term data archiving requirements. Raw data sets that are later manipulated or synthesized may need to be stored in perpetuity. Modifications to protocols will typically require complete data sets to be archived before modifications are implemented. Depending on the monitoring project, it may be necessary to preserve interim data sets (data "milestones") over the long term. Archived data sets or subsets destined for long-term archiving will be saved, whenever possible, in their native formats in addition to ASCII text files. Specific data archiving requirements for ongoing projects are spelled out in the data management standard operating procedures (SOPs) for each monitoring project.

11.1.3. QC of converted data

All ASCII files created from databases will undergo quality control (QC) to ensure that the number of records and fields correspond to the source data set, and that conversion has not created errors or data loss. A second reviewer (preferably a program scientist) will evaluate the ASCII files and documentation to verify that tables, fields, and relations are fully explained and presented in a way that is useful to secondary users. Databases that are converted from one version of MS-Access to an upgraded version will require additional QC' particularly if the databases are being actively used for data entry or analysis. Forms, queries, reports, and data entry all will be thoroughly tested.

11.1.4. Version control

Previous versions of databases will be saved in their native format and archived in addition to the current version. Documentation of version updates and associated details will be part of the archive metadata document, with revision information and history be included in tables within the database files. Directory and file names should be assigned according to the following conventions:

11.1.4.1. Raw Data

Raw data sets obtained from secondary sources are archived in their native format before any manipulation is done to the data. These data should be stored in a raw_data directory, with file names consisting of the data source name and the download date.

Example:
\I&M\Vital_Signs\monitoring_archive\cimate\raw_data\NWSCoop
NWSCoop_04022007.ext
where .ext is the extension of the native format.

11.1.4.2. Working Backup Files

Databases containing processed data should be backed up before new data sets are appended, and regularly during data entry. These data sets should be stored in a working_backup subdirectory, with file names consisting of the database name, version, and date.

Example:
\Vital_Signs\peregrine_falcon\data\working_backup\peregrine_falcon_2.3_02042007.mdb

11.1.4.3. Version Archives

Long-term archive files are created before any database version or software upgrade occurs. These data sets are stored on the GULN server in their native MS-Access format:

Example:
\archive\monitoring_archive\climate\version_archive

A subdirectory is created for each version number. File names of archive files will contain the database name, the database version number, and the word Archive.

Example:
Version_1-00\SnotelClimate_1-00_archive.mdb

In addition to the copy of the database in MS-Access format, all tables will be archived in a comma-delimited ASCII format. These ASCII files will reside in the version archive subdirectory and will be named according to the database version number.

Example:
Version_1-00\SnotelClimate_1-00_archive.txt

11.1.5. Spatial data

Spatial data sets that are essential to GULN will be maintained in a format that remains fully-accessible by the current ArcGIS version. ArcGIS has maintained compatibility with previous data formats. Shapefiles have retained all functionality in ArcGIS, however coverages may require conversion to ArcGIS format if this format is no longer supported. At this time there is no practical way to save GIS data in a software or platform-independent format.

Both uncorrected and corrected global positioning system (GPS) data (e.g., .ssf and .cor files) will be archived in their native format. Corrected files will be converted into GIS files.

11.2. Storage and Archiving Procedures – digital data

11.2.1. Directory structure for electronic archives

Digital data needs to be stored in a repository that ensures both security and ready access to the data in perpetuity. As of 2006, GULN relies on a 400 GB server and an 800 GB server, both with a level-5 RAID (redundant array of independent disks) for data storage, combined with a schedule of full, incremental, and differential backups using 8 mm tape. The servers are located in a locked, climate-controlled room and are managed by the network data manager.

11.2.2. Directory structure for individual projects

The organization and naming of folders and files should be intuitive to users unfamiliar with a specific project. Because each project will have its own idiosyncrasies, a standardized structure isn't realistic; however, all project archives will include several to most of the following elements:

- administrative documents such as agreements, correspondence, research permits
- programmatic documents including protocols, procedures, supporting documents
- interim data sets or "milestones"
- data sets submitted by contractors
- data sets reformatted or manipulated by GULN (e.g., data converted to Natural
- Resource Database Template format, data sets migrated to current software formats)
- data sets – ASCII format
- conceptual or statistical models used for data interpretation
- final report
- readme files -- includes an explanation of directory contents, project metadata (including a dataset catalog report), and version documentation.
- Metadata for all GIS and non-GIS datasets.

Once final data and reports have been submitted, draft products will no longer be need to be maintained and should removed from the archive folder to prevent confusion.

11.2.3. Backup procedures for digital data

The risk of data loss can come from a variety of sources, including catastrophic events (e.g., fire, flood), user error, hardware failure, software failure or corruption, and security breaches or vandalism. Performing regular backups of data and arranging for off-site storage of backup sets are the most important safeguards against data loss. GULN is integrated into the ULL/NWRC local area network and benefits enormously from the expertise of the NWRC system administrators.

As of December, 2006, data residing on the GULN server are backed up onto 8 mm tapes using a DLT tape backup system (Table 11.1). Two routines are used:

1. Backup schedule A is designed for frequently-changing files and allows for total recovery of data from the previous 4 weeks. Because these are actively-used files, most data loss or corruption can be detected within this time.
2. Backup schedule B is designed for archival data or voluminous data files that are relatively static, and allows for total recovery of data from the previous 12 months.

Off-site storage of backup tapes are located on the ULL campus in a fireproof data safe. All backups are performed and monitored by the network data manager.

Backups of data that reside on the personal computers of staff are the responsibility of each staff member. Every computer is equipped with an external hard drive and utilizes Retrospect HD software to perform a daily back-up of selected files and folders. Because the amount of GULN data is increasing rapidly, this backup schedule and system will need to be reevaluated at least annually. Backup routines represent a significant investment in hardware, media, and staff time; however, they are just a small percentage of the overall investment that has been made in program data.

Table 11.1. Backup schedule for GULN servers

Schedule A – frequently-changing files; estimate total of 75-100 GB by 2006. Used for all files located on GULN Data Server\I&M	
5 Tape Sets	Data server is backed up to tape daily using a differential back-up scheme with a full back-up performed every Friday. Each week, a backup tape is transferred to an offsite location.
Schedule B – large files, relatively static; estimate total of 400GB by 2006. Used for all files located on GULN GIS Server	
2 + 4 Tape Sets	Base park GIS raster and vector data sets are backed up once using a full back-up scheme. Project specific GIS data sets that may be modified or edited often by the network will follow schedule A to ensure changes are saved daily.

11.2.4. Data and network security

GULN's LAN is routed and administered through the USGS National Wetland Research Center and conform to Department of Interior security guidelines.

Only I&M staff and system administrators have permission to access files on the I&M servers, and restrictions have been established on archived data files. Directories containing completed project data or interim versions of ongoing projects are designated as read-only for all staff with the exception of the data manager. In this way, any changes must be routed through the data manager, who is responsible for ensuring that documentation and readme files associated with the data set are updated.

11.3. Storage and Archiving Procedures – documents and objects

This section applies to documents such as final reports prepared by staff or contractors, program administrative documents, contracts and agreements, memoranda of agreement, and other documents related to GULN administration, activities and projects. This section also applies to physical items such as natural history specimens, photographs, or audio tapes. In most instances these documents and objects are essential companions to the digital data archives described above.

Direction for managing these materials (as well as digital materials) is provided in NPS Director's Order 19: Records Management and its appendix, NPS Records Disposition Schedule (National Park Service 2003). NPS-19 states that all records of natural and cultural resources and their management are considered mission-critical records; that is, they are necessary to fulfill 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 Appendix B, which provides guidelines on natural resource-related records (including the results of I&M Programs), indicates that all natural resource records are considered "permanent," that is, are to be transferred to the National Archives when 30 years old. 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.

11.3.1. Documents

All paper documents managed or produced by the GULN will be housed in one of two locations.

11.3.1.1. GULN central files, Lafayette, Louisiana.

These files contain project files, administrative documents, and non-record copies of documents that are archived at the regional office or at the park. Examples include: meeting minutes, correspondence, memoranda of understanding, contracts and agreements, research permits, interim and selected final reports produced by the program or under its auspices. GULN will use acid-free paper and folders for all permanent records in the central files. In addition to maintaining these paper records, GULN will maintain electronic versions, when possible, on the data server. The central hard copy files are maintained by the GULN coordinator, and digital copies are maintained under the guidance of the data manager and program coordinator.

11.3.1.2. Park-based Museums, libraries, central files

Hard copies of park-related documents resulting from park data mining efforts and network projects have been digitally scanned into pdf format and cataloged in Reference Manager. These digital copies will be provided to park resource management staff. Parks may choose to accession these materials into their museums, incorporate them into their central files, or house them in their resource management library, as they deem appropriate. GULN will not manage documents at the park level.

11.3.2. Photography

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 to be used by any given project are developed by the project manager and GULN staff. GULN accepts both digital and analog photographs. Under the project manager's direction, only pictures serving as data (e.i. photo used to determine a change in a specific location over time) or as vouchers will be archived as a hard copy photographs. If data or voucher pictures are taken with a digital camera, images should be printed using a LaserJet printer on high quality photo paper. Regardless of whether the picture is a digital printout or a 35mm photo, the rear of the picture should be a labeled to include: a unique ID, project name, accession number, photographer name, photo date, a brief identification of contents (e.g., species name, plot ID), and geographic location (coordinates and datum, or a description). See Appendix D for the GULN Photo Management Strategy.

11.3.3. Specimens

Specimens collected under the auspices of GULN will be provided to the network park in which they were collected, or to a repository approved by a park (where the specimens are considered on loan). GULN will provide park curators with associated data required for cataloging each specimen. These data will be in comma-delimited (.csv) format for automated uploading into ANCS+. Data provided to non-NPS curators will be in Excel format.

11.4. Role of curators in storage and archiving procedures

Curators for parks within GULN are an ongoing source of expertise, advice, and guidance on archiving and curatorial issues, and they have a role in almost all projects undertaken by the network. Project managers should involve park curators when projects are in the planning stage, to ensure that all aspects of specimen curation or document archiving are considered, and that any associated expenses are included in project budgets.

12. Implementation

The data management plans for each of the 32 I&M Networks are the first comprehensive documents of their kind in the NPS and contain practices that may be new to staff and cooperators. However, almost every requirement stems from federal law, Executive Orders, Director's Orders, or national I&M Program guidance. The DMP helps put these requirements into context, and provides operational guidance for achieving them.

12.1. Education and Training

Implementation will require education and training in order to familiarize park staff and cooperators with the tools, procedures, and guidelines outlined in the plan. Formal (training sessions) and informal (one on one communication and assistance) methods will be used. These efforts will begin in 2007 and be led, at least initially, by GULN data manager and GIS Specialist, with participation by interested parties at all parks actively encouraged.

12.2. Milestone Goals

Goals for the first 3 years should include

- Acceptance and understanding by all staff of targeted programs and their cooperators of the fundamentals of data and information management, including
    - File management
    - Documentation
    - Quality assurance and quality control
    - Electronic storage
    - Archive storage
- Improvement of data management practices by implementing
    - accepted database design standards
    - thorough testing of databases, data collection methods, and their integration prior to field work
    - quality assurance and control procedures at every stage of project development
- Common SOPs and guidance documents for multiple protocols
- Inclusion of detailed specifications for data management consistent with the DMP in all protocols developed for the Vital Signs Monitoring Program
- Development of procedures and outlets for communication within and among Network parks and with the public as described in this plan

Beyond the first three years, goals should include the development and assessment of

- Methods for improving file management (e.g., a content management system), database administration and security (e.g., migration to SQL-Server) and integration into the network of off-site users
- Procedures to facilitate the summarization and reporting of monitoring data to staff, cooperators, and the public in formats appropriate to each audience
- Framework and gateway for integration of monitoring data with other initiatives both within the NPS and with outside agencies
- Methods for improving information management (e.g., basic content management services including library services, search tools, and decentralized web administration), database administration and security (e.g., migration to SQL-Server), integration into the network of off-site users, and other needs identified in the DMP

Full implementation of this plan by the GULN I&M Program is expected to be achieved by 2010. Implementation at the broader park level is expected to be an ongoing process and will most likely occur in stages. Realistically, it will not be possible for all programs to fully adopt the practices and procedures recommended without an increase in current levels of funding and staff support. Currently, none of the Network park resources divisions have a data manager on staff. Programs wishing to participate should identify those areas that can be implemented given existing resources. For example, it should be relatively easy for most programs to begin implementing standards and guidelines for organizing and managing electronic files, including digital photos. Pragmatism should dictate prioritization and implementation. Critically important to successful implementation at the broader park level will be the active support of park resource management teams.

12.3. Revisions

The next plan revision should be completed within three years of plan approval, or by October 1, 2010, and then every five years afterward. Plan appendices, including SOPs, detailed guidelines, reference manuals, policy statements, etc., will likely require more frequent updates to account for changes in technology or availability of better information. These updates will be completed as needed.

13. 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* W. K. Michener and J. W. Brunt, eds., Ecological data: Design, management and processing. Malden, Mass.: Blackwell Science Ltd. Methods in Ecology Series.

National Parks Omnibus Management Act of 1998. 105th Congress. Public Law 105-391. Title II, Section 204. 16 USC 5934.

National Park Service. 2002. Director's Order #11B: Ensuring quality of information disseminated by the National Park Service.

——. 2003. NPS records disposition schedule. NPS-19, Appendix B – Revised, 5-03.

——. 2006. Gulf Coast Network Vital Signs Monitoring Plan Phase III Report.

Michener, W. K. 2000. Metadata. Pages 92–116 *in* W. K. Michener and J. W. Brunt, eds., Ecological data: Design, management and processing. Malden, Mass.: Blackwell Science Ltd. Methods in Ecology Series.

Olson, R. J., and R. A. McCord. 1998. Data archival. Pages 53–57 *in* W. K. Michener, J. H. Porter, and S. G. Stafford. Data and information management in the ecological sciences: A resource guide. Albuquerque, N.M.: LTER Network Office, University of New Mexico.

Palmer, C. J. 2003. Approaches to quality assurance and information management for regional ecological monitoring programs. Pages 211–225 *in* D. E. Busch and J. C. Trexler, eds., Monitoring ecosystems: Interdisciplinary approaches for evaluating ecoregional initiatives. Washington, D.C.: Island Press.

Palmer, C. J., and E. B. Landis. 2002. Lake Mead National Recreational Area resource management division: Quality system management plan for environmental data collection projects (draft).

Tessler, S., and J. Gregson. 1997. Draft data management protocol. NPS I&M Program. http://www1.nrintra.nps.gov/im/dmproto/joe40001.htm. Last accessed November 28, 2004.

Vogt-O'Connor, D. 2000. Planning digital projects for preservation and access. National Park Service Conserve O Gram **14**:4. NPS Museum Management Program, Washington, DC. 4 p.

14. Appendix A: List of Computer Electronic Hardware for GULN

GULN Property Tag	Serial Number	Item	Location	Date Acquired
	Covered by prop tag	Camera	RW office - Abdalla	2001
	320307120	Canon EOS 30D Digital SLR camera body (w/ factory accessories- battery, charger, strap, software, manuals, cables) (Model code DS126131)	Woodman –Abdalla Hall	Aug-06
	2728202724	Canon PowerShot S3 IS digital camera (w/ strap, software, manuals, cables) ("point & shoot") (Model Code PC 1192)	Woodman –Abdalla Hall	Aug-06
719	5rk50b1	Dell Desktop	WG office – NOAA Lab	5/9/2006
688	9TTN171	Dell Computer Server	WG office – NOAA Lab	2/25/2005
690	45JD881	Dell Computer Server	WG office – NOAA Lab	7/12/2005
721	3HVDG41	Dell Desk top	MS office	2/13/2004
718	CVKSY71	Dell Desktop	WG office – NOAA Lab	7/12/2005
701	BCM94309MP	Dell Laptop	WG office – NOAA Lab	2001
702	TW03J010-12961-251-1060	Dell Laptop	WG office – NOAA Lab	2003
		DELL Monitor	RW office - Abdalla	10/9/2003
681	CN-0CC299-64180- 64P-06ML	DELL monitor	WG office – NOAA Lab	
687	CN-0C0646-46633-569-3JLL	DELL Monitor	WG office – NOAA Lab	
686	CN-0C0646-46633-569-3JUL	DELL Monitor	WG office – NOAA Lab	
682	CN-0C0646-46633-412-0V5L	DELL Monitor	WG office – NOAA Lab	
683	KR-09J367-47602-240-AGXL	DELL Monitor	WG office – NOAA Lab	
684	CN-0DC323-71618-63E-AC3E	DELL Monitor	WG office – NOAA Lab	

GULN Property Tag	Serial Number	Item	Location	Date Acquired
685	CN-0DC323-71618-63E-AFLL	DELL Monitor	WG office – NOAA Lab	
689	9419-HB773P38908895451	DELL Monitor	WG office – NOAA Lab	
691	9419-HB773P38908808807	DELL Monitor	WG office – NOAA Lab	
	JPHR000611	HP Color Laserjet 5500	Intern office Abdalla	8/8/2003
	SG42H24004	HP designjet Plotter	WG office – NOAA Lab	3/17/2003
	KL-XZR52	IBM laptop	MS office	1/19/2004
	LP-BZGX3	IBM laptop	Robert Woodman	5/4/2005
	I-15110411J8GL48KX6W267132	Jeep Liberty	CUPN Office	4/15/2006
	Y24SAEYE	Maxtor 80Gb	WG office – NOAA Lab	2003
	TW32500566	HP Projector	Intern office Abdalla	2003
696	REG0049006417	Dlink Ethernet Hub	WG office – NOAA Lab	2004
	141007021	Lacie 250Gb External HD	WG office – NOAA Lab	2004
	CV315319	Canon Scanner	Robert woodman	2004
	CV315810	Canon Scanner	WG office – NOAA Lab	2004
	B41PXVLH	Maxtor 200Gb External HD	Robert Woodman	2005
	RB2-3001	HP LaserJet Printer	WG office – NOAA Lab	2006
	WMAMY1041080	Cavalry 400GB External HD	WG office – NOAA Lab	2006
	Y476674E	Cavalry 160 GB External HD	MS Office	2006
	4017632	Lens; Sigma 24-70 mm 1:2.8 EX DG Macro (w/ case)	Woodman –Abdalla Hall	Aug-06
	2338843	Lens; Sigma 70-200 mm 1:2.8 EX DG HSM Telephoto (w/ case)	Woodman –Abdalla Hall	Aug-06
694	BQ0713B001545	DUBH4 Ether net HUB	WG office – NOAA Lab	

GULN Property Tag	Serial Number	Item	Location	Date Acquired
695	Q1251-90011	Hp Designjet 5500ps Printer	WG office – NOAA Lab	
730	C400932202	Belkin Hi-Speed USB 2.0 4 port hub	WG office – NOAA Lab	

15. Appendix B: Standard Operating Procedures for Using Reference Manager 11 Software for the GULN

Version: 1
Status: Final
Author: Luis German
Revision Date: June 14, 2006
Update Frequency: As needed; following new versions of Reference Manager 7.0

15.1. Definitions

Field – a specific piece of information that correlates to one record from a database (ex. The author field and the title field are two separate fields both relating to a single record in a database)

Record – a single record in Reference Manager consists of multiple fields, some of which are left blank more often than not
(ex. A record will have fields such as author, title, date published, keywords, link to .pdf, web link, etc. which all relate to a single file in the database)

Database – a database is a collection of records that deals with similar information (ex. Reference Manager is used exclusively to maintain a database of only bibliographic information. NPSpecies can be used to develop databases of voucher specimen and master lists of park-native species)

15.2. Getting started

15.2.1. File format

Before you can enter the first record into a new database or add one into an existing database you must follow the proper naming conventions for storing the particular file in on the data server. Before following naming conventions, make sure the file has either been scanned in pdf format or has been converted into a pdf from a word document. To convert a word document to a pdf you must have Adobe Acrobat, note that Adobe Reader does not support the converting feature. Once installed, an Adobe toolbar will be made accessible in Microsoft Word (Figure 1.0) that allows the one button conversion.

Figure 15.1. Conversion tool in MS Word.

15.2.2. Naming Convention

The requirements for our naming convention are as follows:

- File names should consist of the primary author's last name, the year published, and as much of the title as possible

- Avoid spaces, dashes, apostrophes, parentheses, and other special symbols in the file name
- Use underscores (_) to separate author, date, and title sections instead of spaces
- Capitalize all words in the title in order to maintain readability
- If the file name is too long, omit the tale portion of the title
- If the document has no author, use “Unknown” as author
- If the document has no publication year, skip over this part of the file name and proceed straight from author to title
- If the document is authored by author and agency, use the agency acronym as the author

Example:

Given the following citation:

(Smith, L.L., T.S. Jones, and R.J. Ramirez. 2002. Effects of the office environment on sanity levels.)

The file name would be:

(Smith_2002_EffectsOfTheOfficeEnvironmentOnSanityLevels.pdf)

15.2.3. File Placement

In order to find files quickly and efficiently we’ve come up with a file structure that enables us to easily process files which have and haven’t been entered into the Reference Manager database. On all NPS Gulf Coast Network Headquarters’ computers the Npsdataserver, where all non-GIS data is stored, should be assigned as the J: drive. This is essential for the *links to pdf* field to work properly.

The file path to store files that have yet to be been entered into Reference Manager is designated as ‘J:\GULN_Scan_PDF_Files\Incoming_Documents_RefMan\.’ Once a file has been entered into Reference Manager, the pdf file must be moved into the proper park folder and category subfolder. For instance a document about Jean Lafitte’s recent fish survey would be placed into ‘J:\GULN_Scan_PDF_Files\JELA\Fish\.’ If a file does not relate to any natural resources then it should be placed into the ‘General’ subfolder for the proper park.

15.3. Windows and Panes

Pictorial guide to the different viewing windows in Reference Manager.

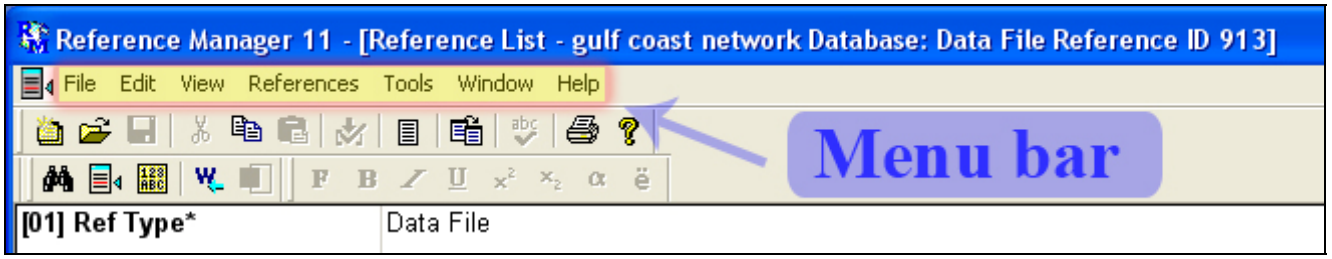


Figure 15.2. Menu bar.

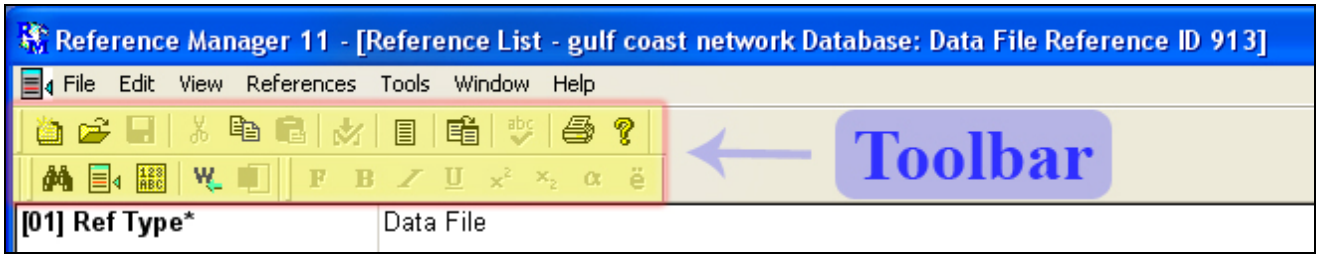


Figure 15.3. Tool bar.

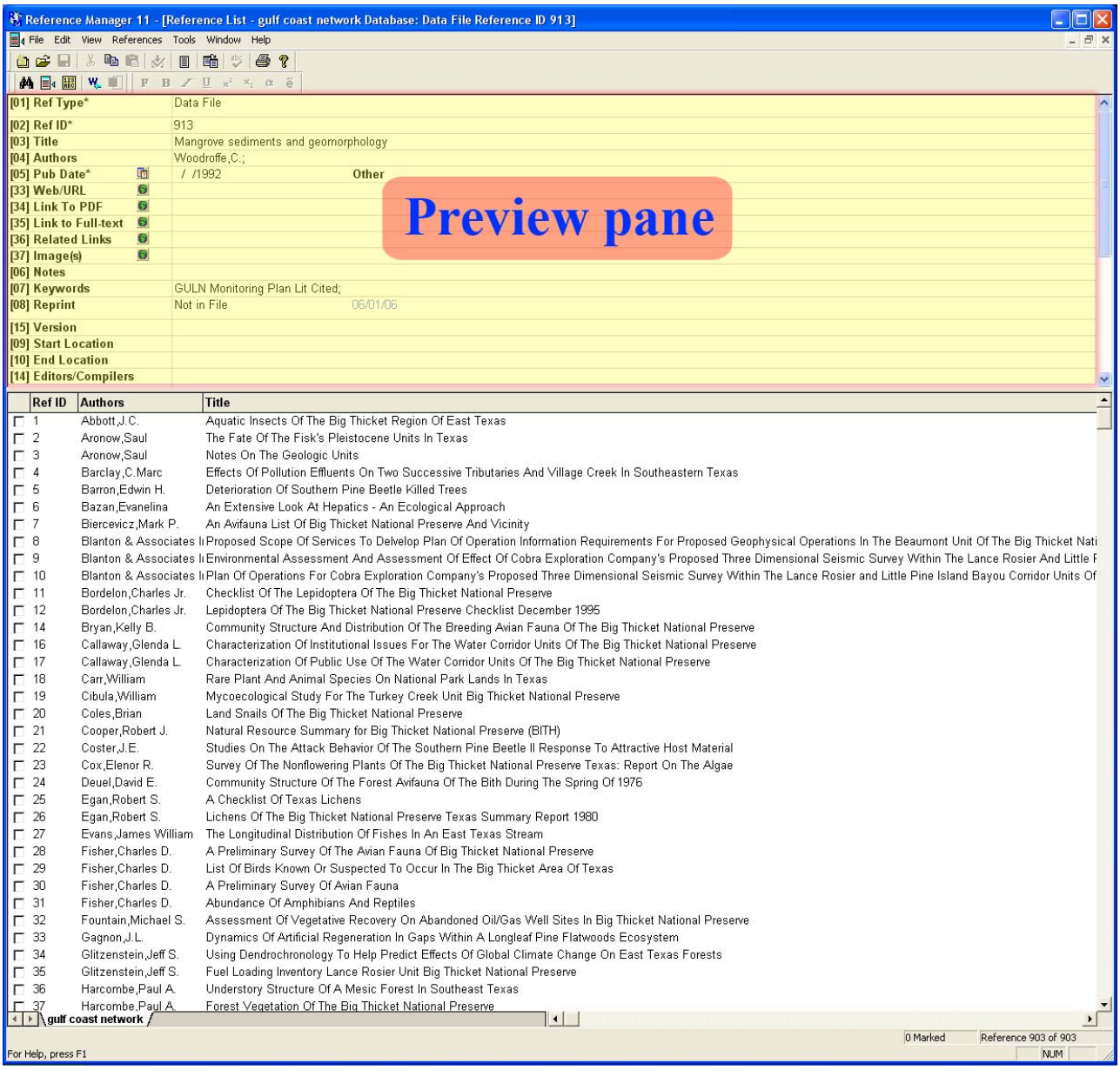


Figure 15.4. Standard Window (preview pane).

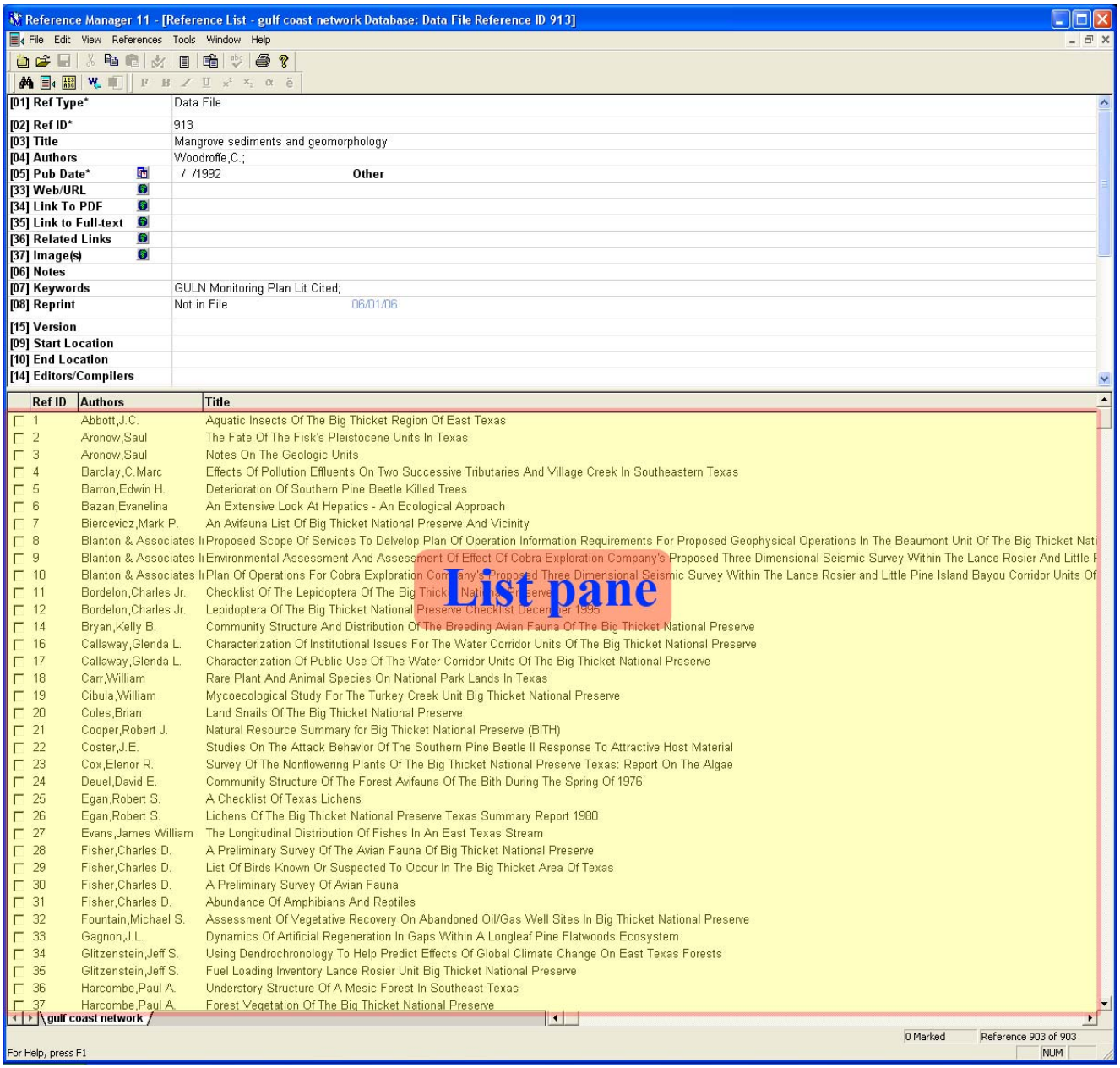


Figure 15.5. Standard Window (List pane).

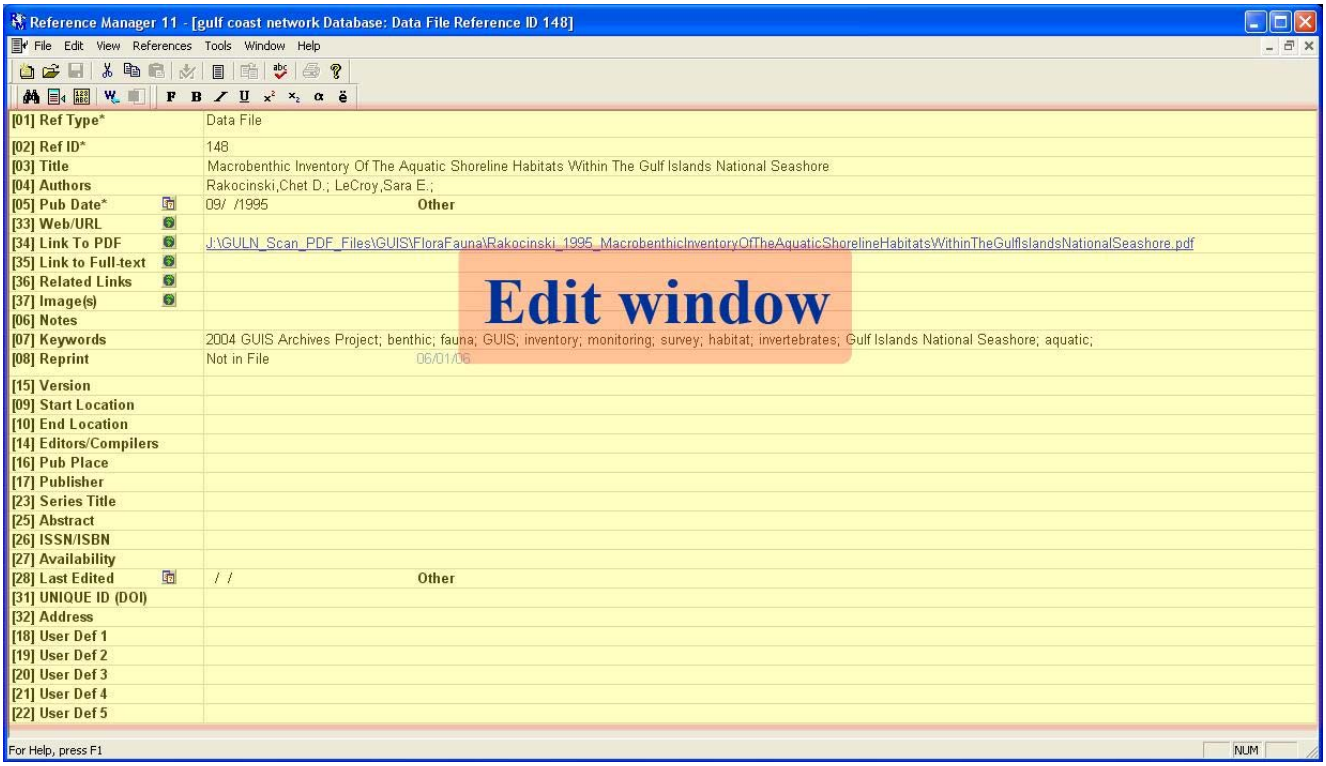


Figure 15.6. Edit Window.

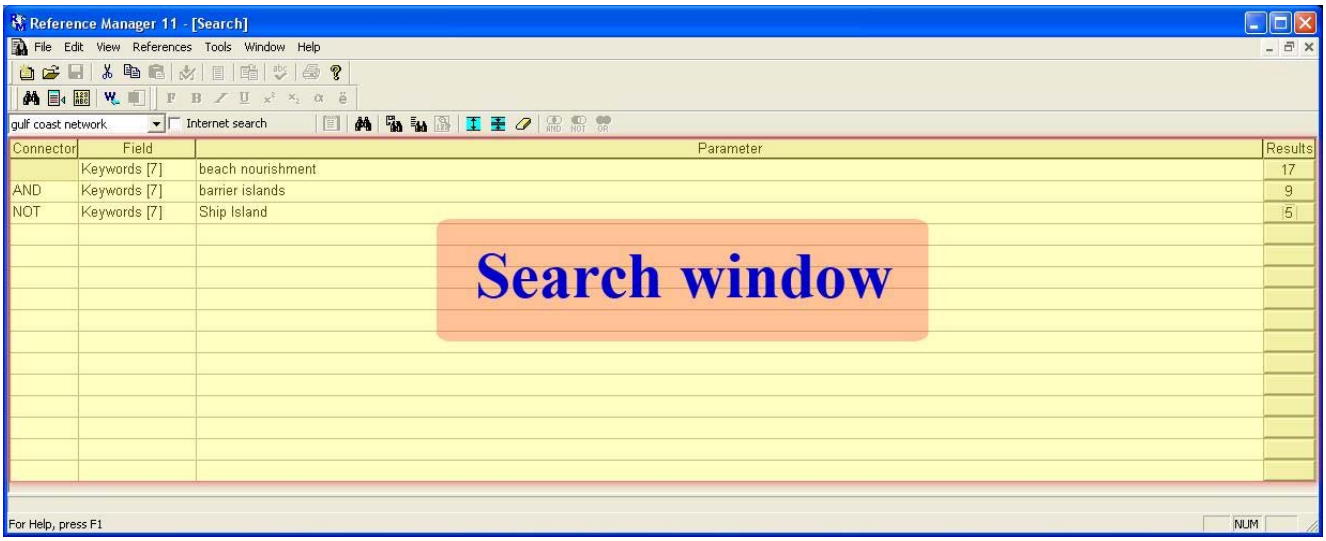


Figure 15.7. Search Window.

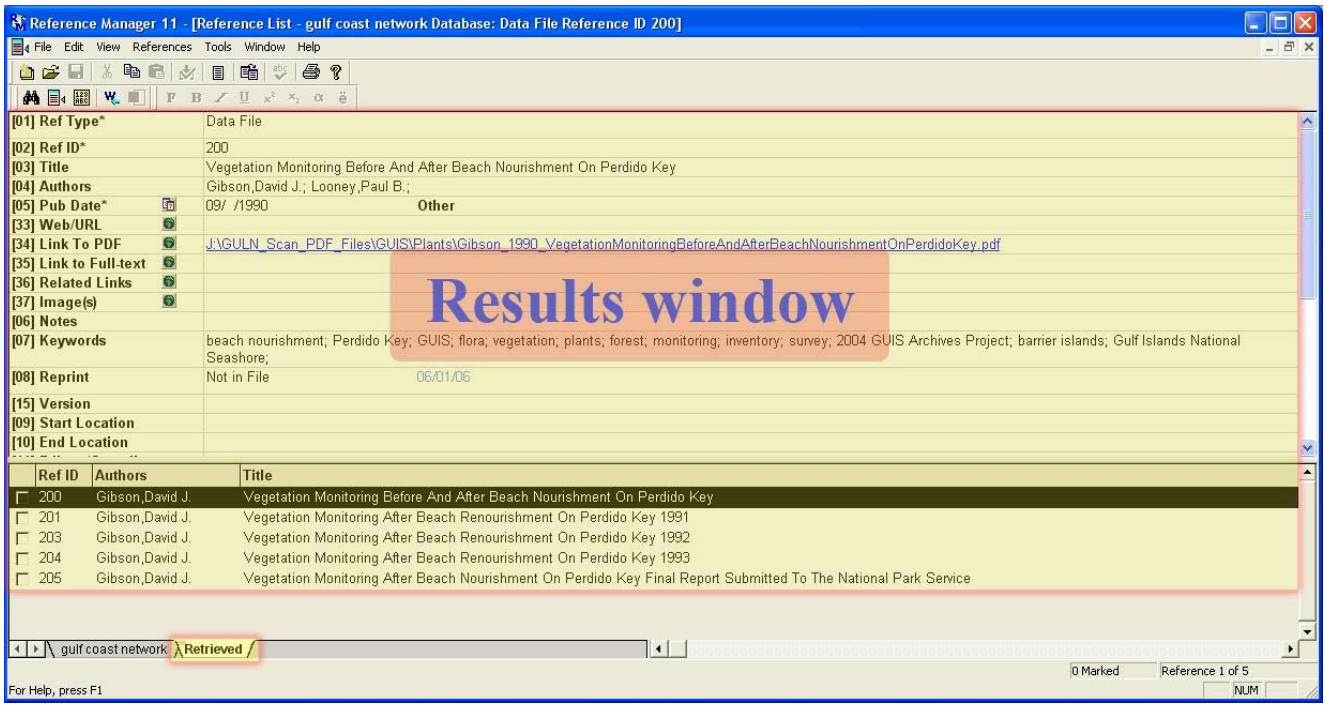


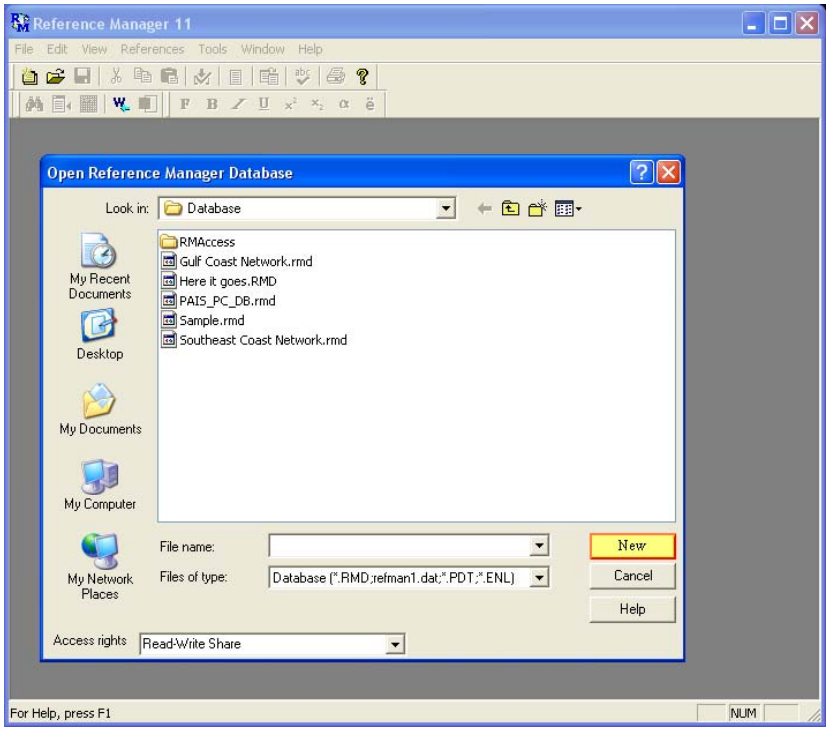
Figure 15.8. Results Window.

15.4. The Basics

15.4.1. Creating databases

This will make a database with no previously entered records.

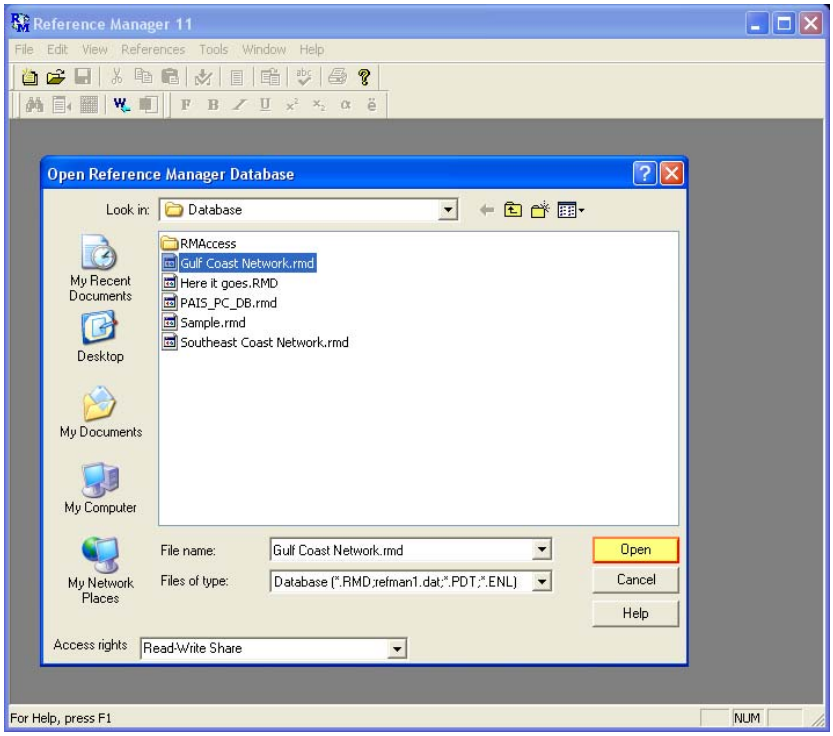
- Navigate to ‘File>New Database’ or use the keyboard shortcut ‘Ctrl+N’
- Enter your desired database name
- Click ‘New’



15.4.2. Opening databases

This is to open a database that already has some records in it.

- Navigate to ‘File>Open Database’ or use the keyboard shortcut ‘Ctrl+O’
- Find your database on your local or networked hard disk and select it. We’re using ‘J:\Program_Files\Reference_Manager_11\Database\Gulf Coast Network.rmd’
- Click “Open”

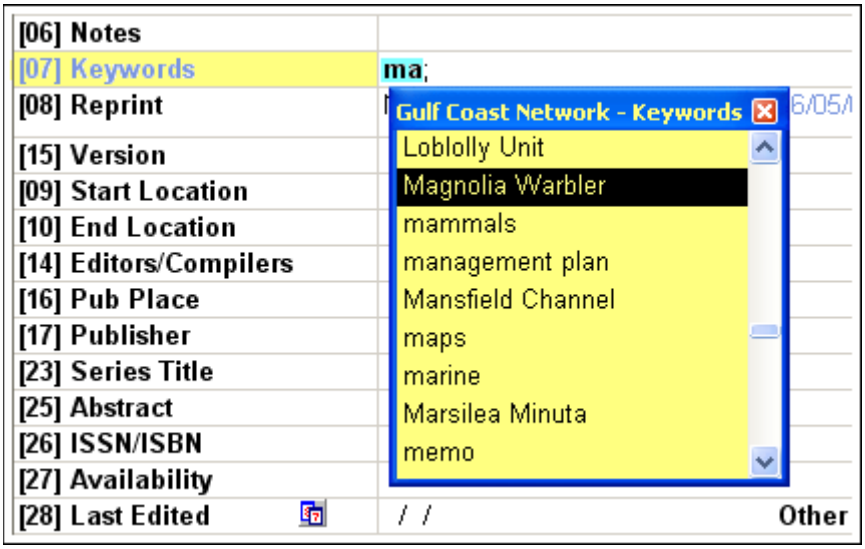


15.4.3. Creating Records

This is how to add new references to our database:

- After opening our database, you should see the standard window, consisting of both the list pane and the preview pane.
- From the menu bar, select ‘Reference>New’ or use the keyboard shortcut [Insert key].
- This will take you to the Edit Window.
- You can copy/paste the file’s name into the title field, delete the author’s name and date from the name, and add spaces to the title portion of the file name or you can retype the title into this field.
- Open the document and locate the bibliographic information. This is often in various locations of different documents.
- Using ‘last name, first name or initials’ format, place all authors into the author field. Using just the primary author is not enough. If no authors are available use ‘Unknown’ as the author. If an organization published the document and no single author is available use the organization’s name for the author field, such as ‘National Park Service’ and ‘NPS.’ Note that when there is an organization that has an abbreviation use the full name first and then the abbreviated name.
- The publication date field refers to the original publication not to a document’s reprint publication date. Normally the date field is an eight digits long number separated by backslashes: two digit month, two digit day, and four digit year. If no date could be located, substitute the entire field with the two digit year ‘ / /01’ because Reference Manager cannot search for a document with no date. We decided to do this so in the event you wish to search for only documents with no dates, you can enter ‘ / /01’ as a search within the date field.

- This is when you move the file to its final location, move it from the 'J:\GULN_Scan_PDF_Files\Incoming_Documents_RefMan\' folder to the one that it belongs. In the link to pdf field you can now copy/paste the file's permanent file location and file name including the file extension.
- The last of the five fields that we require to be filled is the keywords field. Read the title of the document and excerpts to familiarize yourself with the content and apply as many keywords from the Standardized Keywords section of this SOP that you believe applies to the document. If the keywords have been used before, Reference Manager will recognize them within a few letters. You may then use the arrows keys to highlight the proper word from the pop-up keyword list and hit the enter key to select it.



15.4.4. Editing Records

These are the two ways to edit a record previously entered into a database

- Find the record which you wish to edit in the list pane of the standard window
- Highlight it and double-click to select and bring it to the edit window
- Make changes to the desired fields
- Close the window, not the program, to return back to your standard display
- It will ask if you want to save changes to the record
- Click 'Yes'

OR

- Find the record which you wish to edit in the list pane of the standard window
- Highlight it to select it and display in the edit pane
- Make changes to the desired fields in the edit pane of the standard window

- Highlight a different record in the list pane of your standard display to unselect the current record
- It will ask if you want to save changes to the edited record
- Click 'Yes

15.4.5. Standardized keywords

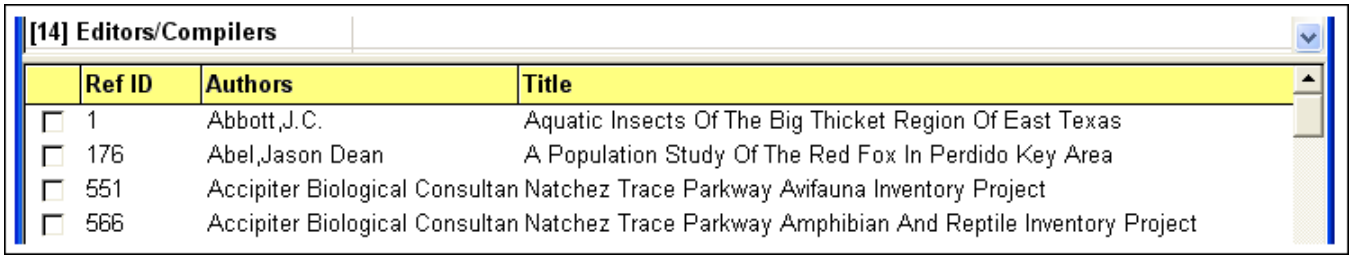
We have standardized keywords to assist in our ability to search for documents concerning certain areas of interest. However, do not feel restricted to use only these keywords. If a document contains subjects not represented in this standardized list, then add the appropriate keywords. I like to refer to these as Flag words, referred to in the last bullet point.

- Park unit (include both abbreviated name and full name): BITH, GIS, JELA, NATR, PAAL, PAIS, SAAN, VICK, Big Thicket National Preserve, Gulf Islands National Seashore, Jean Lafitte National Historic Park and Preserve, Natchez Trace Parkway, Palo Alto National Battlefield Historic Site, Padre Islands National Seashore, San Antonio Missions National Historical Park, Vicksburg National Military Park
- Document type: article, letter, memo, assessment, draft, interview, notes, proposal, conference, field observations, power point presentation
- Report type: study, monitoring, inventory, checklist, survey, management plan, progress report
- Interest type: general, historic resources, cultural, architectural, archaeological, water quality, human impact, land use, habitat
- Natural resource type: fauna, aves, avifauna, birds, fish, ichthyology, amphibia, amphibians, reptilia, reptiles, mammals, fungi, mycology, flora, vegetation, forests, plants, invertebrates, insects, aquatic, endangered, threatened, invasive species, benthic
- Location (if document deals with specific locations, bodies of water, or park units include them as keywords): examples… Hardin County, Gulf Coast, Horn Island, barrier islands, Lake Salvador, Lance Rosier Unit
- Flag words (this includes words that tell a little extra about the document but do not fall into the previous categories): examples… beach nourishment, recovery, preservation, hurricane, erosion, etc.

15.4.6. Sorting records

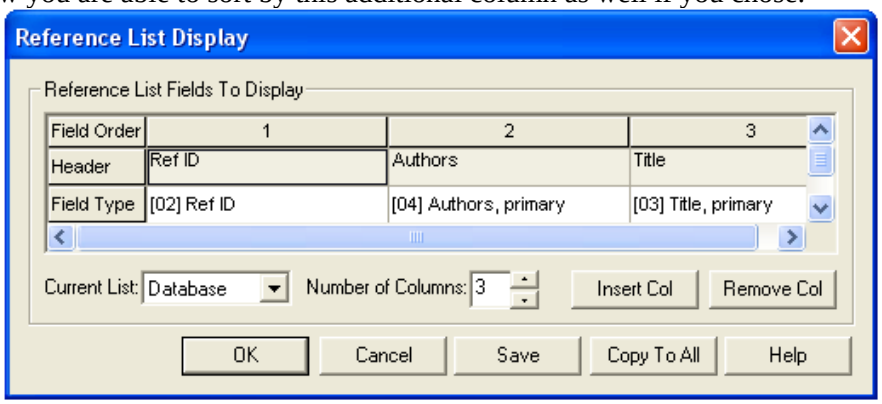
The entire list can be overwhelming if not sorted by a particular field. Across the top of the list pane of the standard window, there's a header that by default contains three columns: Ref ID, Authors, and Title.

- To sort by author, simply click the 'Author' header
- To sort by any other attribute, click on that attribute



If you want to sort by fields other than those displayed by default then you must add more fields to the list pane manually

- To do this, right click anywhere in the list pane and select 'Reference List Display'
- Click the 'Insert Col' button
- Select which additional field you wish to add in the drop down menu
- Click the 'OK' button
- It will then ask if you'd like to copy this setting to other items on the retrieved list. Click 'Yes'
- Now you are able to sort by this additional column as well if you chose.



If you later chose you want to remove a column that you have added to the default display, you must also do this manually.

- Right-click anywhere in the list pane of the standard window and select 'Reference List Display'
- Instead of clicking 'Insert Col', click the 'Remove Col' button after highlighting the column which you want to remove.

15.4.7. Searching records

The flexibility to search records based on any field is convenient when pulling up information about a particular topic or finding a document by a specific author.

- To access the search page, you can click on the binoculars button on the toolbar, or go to 'References>Search References' on the menu bar, or use the keyboard shortcut [F4].
- If you're used to a search bar like Google's you may be a bit unfamiliar with this search page. It's divided into four columns: Connector, Field, Parameter, and Results.
- The best way to understand these columns is for me to explain them out of order. The Parameter column is where you input the actual words you want to search for in the database. The Field column is represented by a drop down menu where you choose which field the parameter searches. The Connector column is only used when doing a combination search. This column uses Boolean phrases to combine searches. If you've never used "or, and, not" as words in a search then perhaps it would be beneficial to read this article (http://www.health.library.mcgill.ca/course/boolean.htm) on using Boolean phrases to enhance searches. You should keep in mind that there's an order of operation when using Boolean phrases (AND, NOT, then OR). The order in which you place your searches and connectors will affect the search results. The Results column shows the

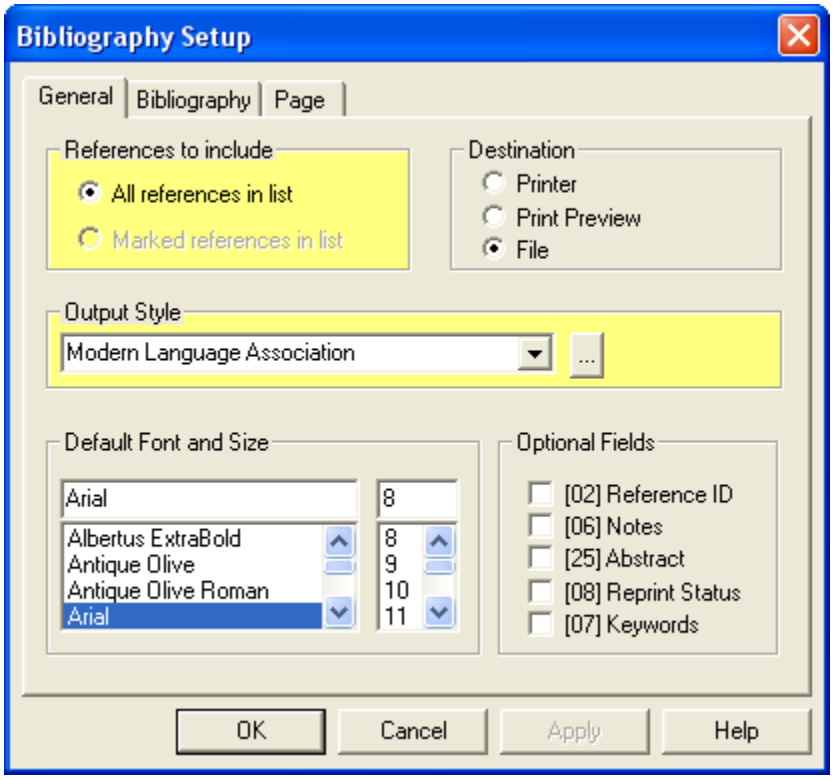
number of records returned from the search and once clicked brings you to the actual Returned Results List window. This window looks exactly like the Standard List pane and appears as a tab on the bottom the Reference Manager window.

- Let's walk through a complex search with three search parameters to get a good grip on the search tool. On the first line for the Field column, use the drop-down menu to select 'Keywords' and type 'beach nourishment' into the Parameters column. Now on the second line and third line select 'Keywords' for the field and type 'barrier islands' and 'Ship Island' as the parameters respectively. The results will be displayed at the end of each line, but why? The first line's result number refers to the number of documents that were found to have 'beach nourishment' as a keyword. The second line's results number is the amount with both 'beach nourishment' and 'barrier islands' as keywords. The third line has results that have 'beach nourishment' and 'barrier islands' but not 'Ship Island' as a keyword. This means each line displays the results up to that point in the search. CAPTION: In this case (see above) 17, 9, and 5.
- In most cases, three or four search parameters will suffice for finding want you need, but if you have very specific search criteria and only if you are proficient in logical thinking, you may benefit from knowing that it is possible to simulate embed Boolean phrases with parenthesis by using the operator buttons on the toolbar. This will insert the Boolean operators into the actual Parameters column. By doing utilizing Boolean operators within the Parameters column and in the Connector column, the search tool simulates the structure of parenthetical logic. This is the only way to navigate around the traditional Boolean hierarchy.
- The search tool is very flexible and only as complicated as you make it to be.

15.5. Advanced Features

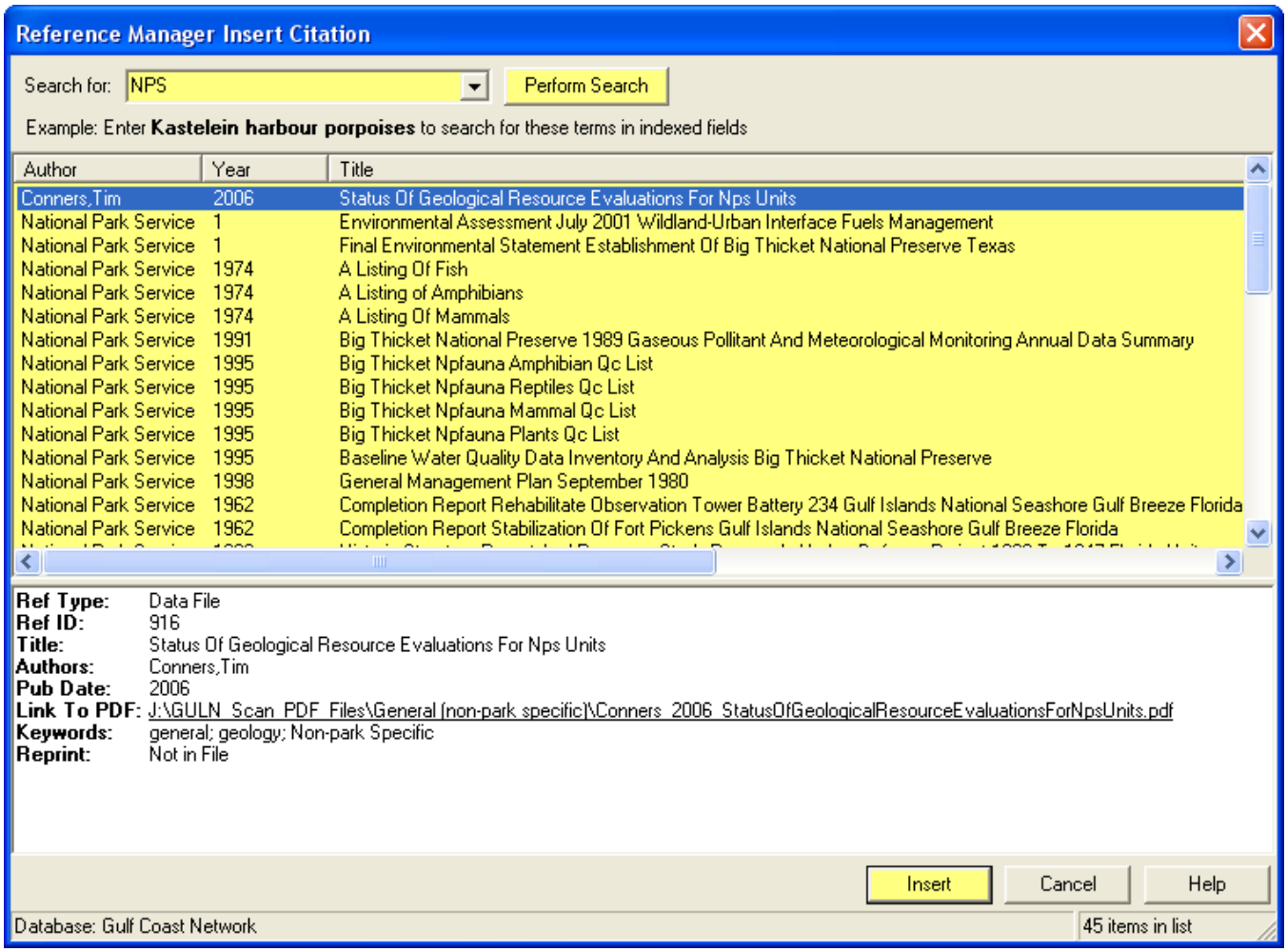
15.5.1. Generating lit cited pages

- You can generate a bibliographic cited page containing as many of the records in your database as you want.
- To generate a list cited from your database, navigate to the menu bar to 'Tools>Bibliography>Generate From Reference List' or use the keyboard shortcut 'Ctrl+L.'
- This will bring up the Bibliography Setup Window. Make sure that you've selected the 'All references in list' option if you want to generate a list cited of all items in your current pane (whether it is from a search pane or standard list pane). You also have the choice to only cite selected, check-marked records from a pane with the 'Marked references in list' option.
- Apart from choosing what will be cited, you can also determine the format in which each individual citation will be presented on the generated rich text format file. This is done by choosing the output style. By default it is set to 'Modern Language Association' but by clicking on the triple dotted button next to it you have the ability to choose others. I recommend the 'NPS.os' format style file if you have it.
- After manipulating those two options (references to include and output style), you can further customize but it is not necessary. Click the 'OK' button and choose where you wish to save the generated rtf file.



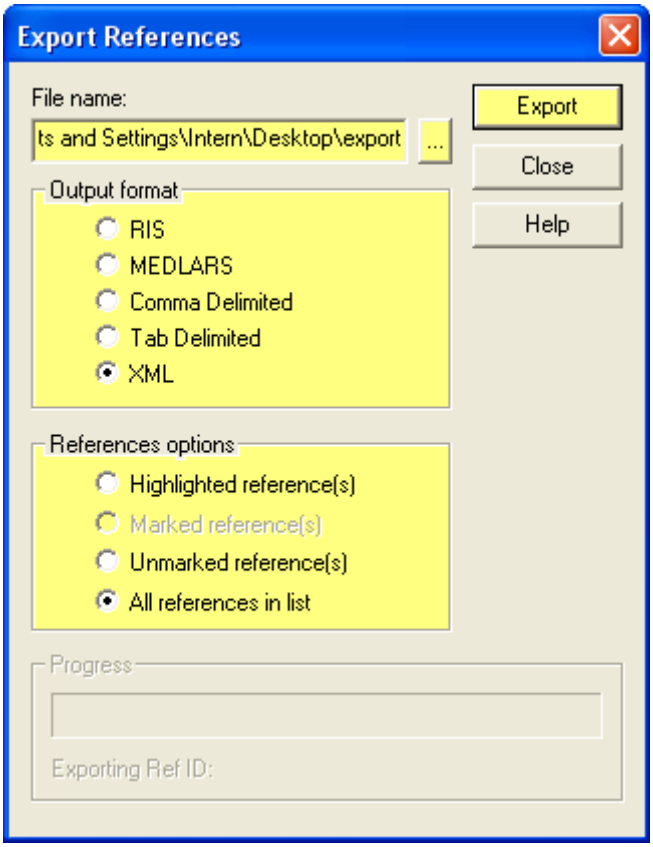
15.5.2. Accessing Reference Manager through Microsoft Word

- After Reference Manager has been installed on a computer with Microsoft Word, a new item will appear under 'Tools' in the menu bar of MS Word. It will read as 'Reference Manager.'
- You can have in-text citing that's automatically formatted and added to a document's Lit Cited page by navigating to 'Tools>Reference Manger>Insert Citation'
- This brings up a Reference Manger Insert Citation window. In this window, you'll have a search bar that isn't nearly as powerful as in the actual Reference Manager's search feature. Type whatever you wish to search and hit enter. It will find all matches regardless of in which field the match was found. Using the Control key on the keyboard and clicking allows the use to highlight multiple citations at once. After selecting which of the records you want to cite simply hit the 'Insert' button.



15.5.3. Exporting databases

- If you ever need to pull your database into another database management program all you need to do to convert your file's extension is export the original from within Reference Manager as an XML file.
- Navigate to 'File>Export' from the menu bar.
- This brings up an Export References window that has your three important options. 1- File Name – where the file will be saved. 2- Output Format – the way your database will be exported. 3- Reference Options – which records of your database will be included in the conversion to a new format.
- Tips: Save the file in an easily accessed location, I recommend the desktop, until you move it to its final location.
- Tips: When exporting you have to be familiar with what you plan on doing with the converted database. If you want to switch to a different database manager you must check what formats that program can import and use that format as the Output Format.
- Tips: Make sure you've looked carefully at what you've selected to export under the Reference Options. For creating an exact copy of your database you must select 'All References In List.' For picking and choosing you need to use one of the other options based on your preference.



15.5.4. Importing/Merging databases

If you're switching from another database manager (i.e. Procite) to Reference Manager, you may want to import your old databases into Reference Manager and save to its format so you don't have to re-enter every entry into it. You may also want to merge to RM databases to make one master database. Use these instructions to do both.

- First, export the file you would like to import as a xml file from your old database manager.
- After that's done, open Reference Manager and navigate to 'File>Import XML File' on the menu bar.
- It will bring up your classic open dialogue window so you may locate the xml file to pull into Reference Manager. Find the file and click Open.
- Next just use the Save option and you'll have a fully compatible format.

Original Version #	Date of Revision	Revised By	Changes	Justification	New Version#
1.0					

16. Appendix C: Product Specifications Standard Operating Procedure

16.1. Purpose

One of the primary goals of the National Park Service (NPS) Inventory and Monitoring (I&M) Program is to make existing and new natural resource information more available and useful to park managers, scientists, and educators for planning, management, research and education. The specifications presented here were developed to provide NPS staff and cooperators with a standard set of guidelines to use for the submission of all products produced during NPS sponsored natural resource studies.

16.2. Scope and Applicability

The National Park Service has adopted certain word processing, database, and GIS software as standards to promote compatibility and sharing of data among parks. In addition to software standards, the NPS I&M program has adopted a set of products with standardized formats to be submitted along with all NPS, I&M sponsored studies. This document provides details on these required software and product standards.

16.3. Reference Documents

- Code of Federal Regulations, Title 36 (2003) Document Online at: http://www.access.gpo.gov/nara/cfr/cfr-table-search.html#page1
- Product Specifications document available online at: http://science.nature.nps.gov/im/datamgmt/docs.htm
- National Park Service (2000) Director's Order 24, NPS Museum Collections Management. Document available online at: http://www.nps.gov/policy/DOrders/DOrder24.html
- National Park Service (2003) Museum Handbook. Document available online at http://www.cr.nps.gov/museum/publications/handbook.html
- National Park Service (in Development) Natural Resource Management Reference Manual #77. Document available online at: http://www.nature.nps.gov/RM77/
- National Park Service (2003) Northeast Region I&M Program Product Specifications. 28 pages.
- Wotawa, Mark A. (2003) Guidelines for Submitting Biological Inventory Products. Document available online at: http://science.nature.nps.gov/im/inventory/biology

16.4. Procedures and General Requirements

16.4.1. Software Standards

- Word Processing: Microsoft Word. All reports and documents must be delivered in electronic format in the latest available version of MS Word unless otherwise specified in advance.

- Databases: Microsoft Access. All database products must be delivered in electronic format in the latest available version of MS Access unless otherwise specified in advance. All database products will be compatible with the Natural Resource Database Template (described below).

- GIS Products: Must be compatible with ArcINFO or ArcView and be accompanied by FGDC compliant metadata. Additional information regarding GIS specifications is described in the GIS Specifications SOP document.

- Metadata: For all spatial datasets, metadata must be FGDC compliant. Please refer to the Metadata SOP and http://www.fgdc.gov for additional information. There is currently no standard metadata software platform, however, SMMS or ArcCatalog are typically the platforms of choice. For non-spatial data (e.g. databases, data files, photographs, field data sheets etc.), a data dictionary must be provided – please refer to the Data Dictionary SOP.

- Bibliographic Citations: All reference or citations must be delivered in electronic format in the latest version of ProCite and will also be recorded in NatureBib, the NPS bibliographic database (described below).

- Species Information: All species information must be entered into NPSpecies (the NPS biodiversity database, described below) and information for voucher specimens must be entered into ANCS+ (the NPS Automated National Catalog System, described below) software.

16.4.2. National Park Service Information Management Tools

- Natural Resource Database Template: The Natural Resource Database Template is a flexible, relational database in MS Access for storing inventory and monitoring data (including raw data collected during field studies). This relational database has a core database structure that can be modified and built upon by different parks and networks depending on the components of their inventory and monitoring program and the specific sampling protocols they use. A description of the template and data dictionary is included as Appendix XX. The most recent version and details can be found at http://science.nature.nps.gov/im/apps/template/index.htm and specific implementations of the database template will be described on a protocol by protocol basis.

- NPSpecies: NPSpecies is the master species/biodiversity database for the NPS. The database lists the species that occur in each park, and the physical or written evidence for the occurrence of the species (e.g., vouchers and reports). Taxonomy and nomenclature are based on the ITIS, the interagency Integrated Taxonomic Information System. The master version for each park or network of parks is stored in Oracle and is a password protected, online database. Copies of each park's master database can be downloaded from the secure website into the MS Access (desktop) version of NPSpecies. NPSpecies is linked to NatureBib for bibliographic references that provide written evidence of a species' occurrence in a park. A data dictionary is included in Appendix XX. The MS Access application and additional details can be found at http://science.nature.nps.gov/im/apps/npspp/DesktopApp.htm

- NatureBib**:** NatureBib is the master, Oracle database that merges a number of previously separate bibliographic databases such as NRBib, GeoBib, and others. As with NPSpecies, it will be possible to download data from the master web version into the MS Access version that can be used locally on computers without an internet

connection. Details are available at http://science.nature.nps.gov/im/apps/nrbib/index.htm

- ANCS+: ANCS+ is the Automated National Catalog System software distributed by the Museum Management Program of the NPS Cultural Resource Stewardship and Partnerships. All voucher specimens, their associated data and cultural objects collected in national parks are required to be entered into ANCS+, regardless of where they are stored. For voucher specimens, ANCS+ requires the documentation of more information regarding the specimen than does the voucher table in NPSpecies. It is possible to electronically transfer the required fields from ANCS+ to NPSpecies, and parks and their contractors are encouraged to first enter voucher data into ANCS+.

16.4.3. NPS Natural Resource Project Product Specifications

There are 4 main types of products that will be generated by most NPS natural resource projects: documents, maps (GIS layers and metadata), data sets, and possibly vouchers. Documents include technical reports, annual and final project reports, and (possibly) field forms or notebooks if they are scanned into a digital format. Maps are geo-referenced GIS layers and the supporting FGDC compliant metadata. Data sets are all tabular data stored in databases or spreadsheets. Vouchers include physical specimens (whole or parts of organisms), audio recordings (digital or physical) or photographs (digital or physical).

16.4.4. Final Reports

At the completion of a project, the investigator must submit a draft final report that meets the guidelines presented below. Upon submission of the draft final report, the designated NPS Key Official will review the manuscript and seek additional comments from NPS regional staff, park personnel, and peer scientists to ensure the technical quality and accuracy of information. All appropriate comments from draft final report reviews should be addressed and incorporated during the preparation of the final report – and must be approved by the NPS Key Official before accepting the final report.

While the format may vary, the final report should include the following content:

- Title Page – Should include: title, parks, study dates, report date, authors, and other credits (contract numbers etc.) as appropriate.
- Introduction – Should include the hypotheses and purpose of the investigation, research objectives, conditions under which the study was conducted, the general plan of treatment of the subject, and summary of previous work (literature review) that relates to the project.
- Study Area – describe the area(s) and habitat(s) included in the study.
- Sample Design – describe how the sample locations were chosen.
- Sample Locations – include locations in the sampling design as well as any incidental locations (presented as GIS maps).
- Methods – Present a detailed explanation of the methods, materials and analytical techniques that were used in the field, laboratory, and office during the study. The methods should be documented so that the investigation could be exactly repeated, if necessary. Be sure to include how the data were analyzed and what statistical tests were used, if any. Describe the processes used for determining

whether the data met data quality objectives and, if not, what corrective actions were taken. Detailed QA/QC procedures for the data collection, verification and validation should be placed in an appendix (see Data Entry and Data Validation and Verification SOP's).

- Results – In a logical sequence, present the findings of the study the either support or provide contrary evidence against the hypotheses or that answer questions posed in the introduction. Basic, descriptive statistics (sample size, mean, max/min etc.) are appropriate when clearly presented. Avoid highly technical discussions of complex statistical testing; instead refer readers to a separate appendix for additional information.
- Discussion – The "Discussion" and the "Conclusion" sections are the most important sections of the report. Present a clear interpretation of the data that addresses the hypotheses, objectives, or purpose for which the study was conducted. Be sure to include how this research is applicable to the park(s) where the study occurred and to other studies that have been conducted in that specific area of research. Other findings may be reported that would be of general interest to the scientific community.
- Conclusions – Provide a specific and detailed summation of the conclusions of the project. If the study was initiated due to specific park management needs, management implications should be emphasized and discussed.
- Literature Cited – List all references in the "Literature Cited" section of the report, using the Council of Biology Editors (CBE) bibliographic style as outlined in Appendix X-7.
- Appendices – For biological/species inventories, include a table of Organisms Found and a table of Organisms Expected, But Not Found. Otherwise, include any supplementary information that supports the main body of this report.

16.4.5. Progress Reports

While the exact format may vary, progress reports should include the following content:

- Title Page – Should include the words "Progress Report; title of the project; investigator names, affiliations and addresses; NPS contract, purchase or agreement number; date of submission and time period of the report.
- Findings – A quantitative description of overall progress and significant findings to date and a status report of cataloging vouchers and associated data.
- Problems – An indication of any current problems that may impede performance and proposed corrective actions.
- Next Steps – Discussion of work to be performed during the next reporting period.

16.4.5.1. GIS Layers and Metadata

Detailed information regarding GIS layers is presented in the following SOP documents: GPS Specifications and GIS Specifications, and thus will not be duplicated here. All digital geospatial data should have Federal Geographic Data Committee (FGDC) compliant metadata in digital form developed by the producer. Metadata should be supplied as ASCII text with a .txt extension, hypertext markup language with an .html extension, and standard general markup language with a .sgml extension. Cooperators should contact the project manager or data manager for metadata development instructions and refer to http://www.fgdc.gov regarding metadata standards.

16.4.5.2. Database Specifications

As previously stated, all databases should be in MS Access and be compatible with the Natural Resource Database Template (NRDT) data structure. Due to the project specific requirements of each database, a lengthy discussion of database specifications will not be presented here. However, a meeting between the project and/or data manager and cooperators, prior to beginning data collection, is required to ensure that the database is properly developed and meets the needs of all participants. Additional information regarding the NRDT is provided in the Database Template Specifications SOP.

16.4.5.3. Voucher Specimens

The scientific value of a comprehensive set of voucher specimens for vertebrate and vascular plants in a park cannot be overestimated. These collections document the occurrence of organisms despite changes in taxonomy and provide the basis for evaluating and monitoring changes in species richness and genetic diversity. At the same time, we must remember that science activities in parks should serve to preserve species, not adversely affect them. These aspects, along with the cost of curation and storage requirements, were considered in developing the voucher policy of the SECN.

Investigators will be required to collect adequate vouchers to document species occurrence, but extensive vouchering to document the full range of phenotypic variation will not be undertaken. Investigators must have a valid park collecting permit to collect specimens. Permission to take duplicate vouchers will be considered on a case by case basis by individual park resource managers. Voucher preparation will be the responsibility of the project scientist (whether they are NPS staff, contractors, or cooperators) and must follow the guidelines outlined in 36 CFR and the Museum Handbook (see Reference Documents above). NPS Management Policies; RM77, Natural Resource Reference Manual #77; and Director's Order 24 for Museum Collections will be followed while the best possible storage and curatorial arrangements are made to provide for long term maintenance and access.

Regardless of their final repository, all vouchers taken on NPS lands are the property of the National Park Service and are assumed to be on long-term loan to the housing institution/collection outside of the Service. It is mandatory that cooperators catalog all collected specimens in the Automated National Catalog System (ANCS+) and that all specimens are labeled with the appropriate NPS property label. Please refer to the Voucher Specifications SOP for additional information.

16.4.5.4. Final Submission of Products/Deliverables

All deliverables mentioned above (reports, GIS layers, and databases (including voucher information)) must be submitted on CD-ROM to the Southeast Coast Network - Project or Data Manager as well as to all parks involved. All CD's submitted must contain a "Readme" file providing the file name of each file on that CD and a brief description of each of those files. If possible, all deliverables should be provided on the same CD, including any Powerpoint presentations created for the project, digital photos taken during the project, etc. All files must be labeled clearly and be referred to in the accompanying "Readme" file.

16.5. Responsibilities

Project Scientist: The project scientist is responsible for the final submission of all products/deliverables – including the content and formatting described above. This document and other SOP's provide basic information regarding each specific topic, but may not answer all of your questions. If you have any questions or concerns, please contact the Project or Data Manager as soon as possible to resolve any potential issues. Any deviation from the products described above must be approved of by the Project Manager, in advance.

Data or Project Manager: The NPS data or project manager will provide potential cooperator(s) with the Product Specifications SOP along with the original request for proposals. Supply all relevant SOP's along with the contract or cooperative agreement for their reference during the study. Meet with the cooperator(s) prior to beginning any fieldwork to ensure that all participants are aware of and understand the project requirements and be responsive to questions from cooperators, as they arise.

Park Natural Resource or Curatorial Staff: Park natural resource or curatorial staff will work with the cooperator during the research permitting process and will provide the cooperator with voucher specimen – collection and accession numbers for all museum and/or herbarium specimens collected.

17. Appendix D: Gulf Coast Network Photo Management Strategy

17.1. Introduction

The Gulf Coast Network (GULN) Inventory and Monitoring (I&M) Program is tasked with organizing and maintaining a variety of research information, including digital images collected as data and images taken to document protocols and procedures. Concurrently, there is a growing need to make all NPS digital photos useful and available service-wide. This document describes a management strategy to organize, store, name and retrieve photographs in electronic format. The intent is to centralize processing and storage, reduce wasted disk space and to make photos available to all GULN employees.

17.1.1. Assumptions and Definitions

- This document refers only to photographs in an electronic format.
- In this document 'digital photo' refers to a photo in electronic format, regardless of how it was acquired, by scanning or from a digital camera.
- 'Photo processing' is used in this document to refer to the sum total of all steps necessary to go from a photo on your camera's storage card to the photo library.
- 'Documented' photos refer to photos that have completed metadata in either the parks photo database or a project database.

Photos are taken and used by NPS personnel for a variety of purposes. NPS photo users vary tremendously in their knowledge and access to imaging software and hardware. While it is impossible to anticipate every use/collection scenario, most photos should fit into one of the general "photo type" categories defined below:

- Library Photos. These photos are final products that have been edited, documented, reviewed and added to the network photo library. Metadata for these photos is created and stored using ThumbsPlus7 software. These photos can be used for multiple purposes by a variety of staff. They are public domain.

- Working photos. Photos in this category are "works in progress". Working photos should be documented and moved into the parks photo library or they should be deleted. Metadata for working photos does not exist but is in progress. Working photos are stored in centrally located employee specific folders adjacent to the park photo library.

- Data Photos. Data photos are photos collected as data. Examples include: 1) site specific photos documenting a shoreline classification and, 2) site location photos taken as part of a land bird monitoring project. Data photos are collected as part of a well-defined data collection protocol. Metadata for these photos is stored in a project level database, ThumbsPlus7 database, and housed within the project folder structure. Representative, unique and instructive data photos should be added to the parks digital photo library.

- Special Collections Photos. Occasionally there are special collection photographs, such as from a famous photographer or from an expedition or an archives collection.

- Aerial Photos (orthographic photos, not oblique snapshots out of an airplane door or window). These are usually produced professionally by a contractor, where very clear specifications are made. Digital aerial photos standards are discussed but are beyond the scope of this document to be described in detail.

- Oblique Aerial Photos. Oblique snapshots taken out of an airplane door or window. Examples include documenting damages to parks caused by hurricanes, fires, or other large scale disasters warranting the use of an airplane.

17.1.2. Digital Photo Cycle

The "digital photo cycle" typically runs through the following steps. Photos are acquired, stored, viewed, renamed, edited, documented, stored in a photo library, archived and, in some cases, deleted. Some of the issues and questions which crop up along the cycle include:

- Acquisition
    - Digital cameras – What is the best quality and size?
    - Scanning images – What parameters to use?
- Processing
    - Download and Storage – Where? For how long? Naming standards?
    - Viewing – With which software? Or as contact sheets?
    - Renaming – Manually or batch? Naming standards?
    - Editing – Which software? Should you write over the master or make a copy?
    - Documenting – Which attributes should be documented? How is the documentation 'attached' to the photo?
- Long-term Storage
    - On-line storage – Where? Managed by whom? For how long?
    - Off-line storage and archiving– Where and when? When to delete?

This document will attempt to address these questions and provide logical standards and guidelines to facilitate good photo management. This document discusses imaging standards and how digital photos will be stored in a repository, documented and cataloged. Figure 1 summarizes the general photo management process.

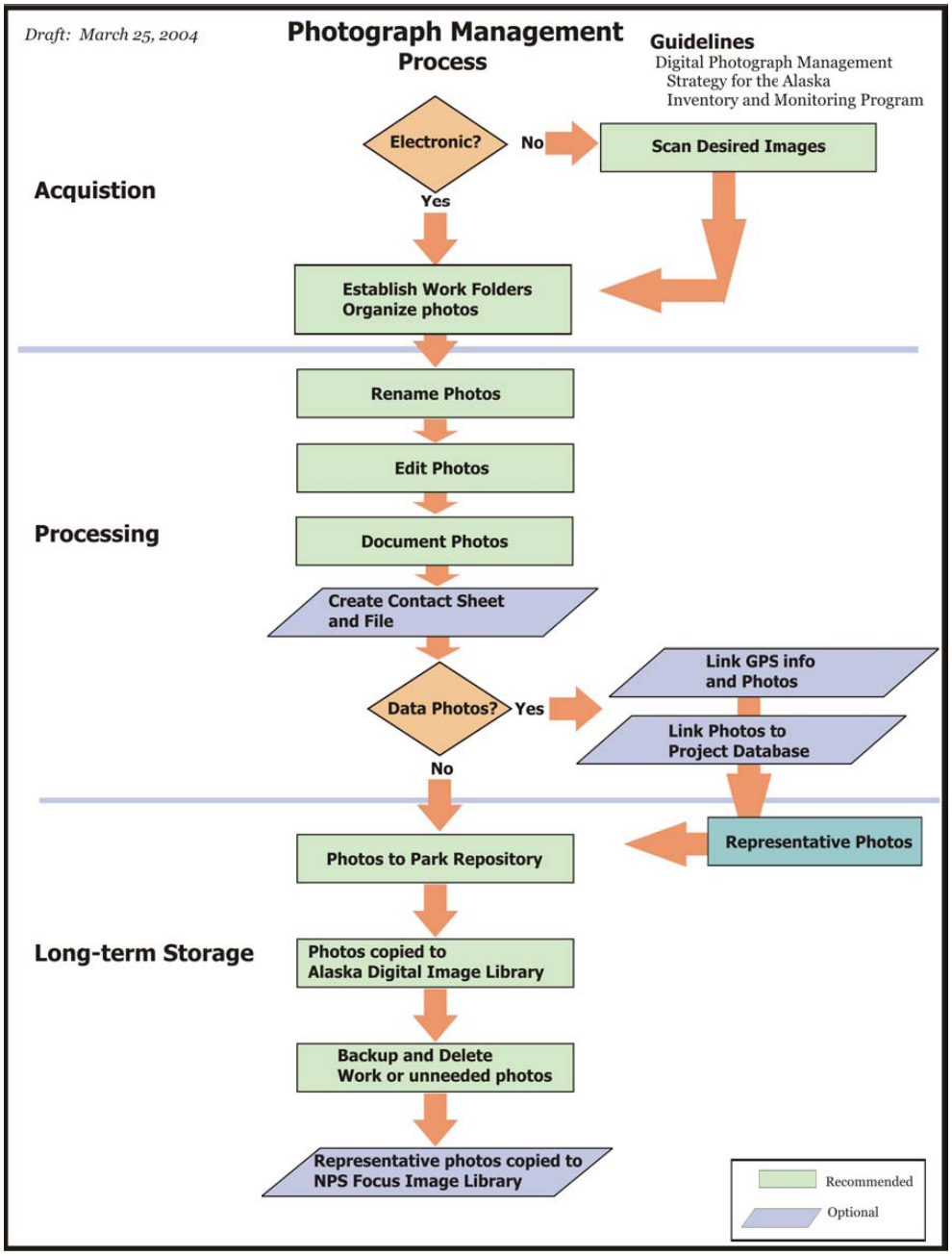


Figure 17.1. General photo management process.

17.2. Acquisition of Photos

17.2.1. Digital Camera Specifications

Digital cameras should be set at a resolution appropriate for the highest level use to which a photo may be used.

Publication quality photos should be taken at a minimum of 5 megapixels. If the camera will allow, the resolution should be set at 1760 x 1168 or higher. The quality should be set for “super fine” or “high”.

Because the destination of a photo is unknown at the time it is taken, it is recommended all photos be taken with this resolution or for lower resolution cameras, the highest resolution possible. It is best to decrease the resolution for web use or thumbnails in the office using software.

Most digital cameras can digitally imprint the date and time onto the photo image. In general, it is recommended that this feature not be use. Date and time data are embedded in jpg and tiff file headers by most digital cameras. If the image is being cataloged and documented it has value – imprinting the image reduces the image quality and hence the image value. The exception may be for projects collecting massive numbers of photos – imprinted dates and time could facilitate data linking.

Digital camera internal clocks are subject to minor cumulative errors. Accurate camera date and time is important to maintain, it is recorded as the file creation date for each photo and for jpg formats it is embedded in the photo itself. It is best to frequently sync a camera time to an active GPS or to the network time. Beware of time zone changes where appropriate.

Different brands of digital cameras name photos differently. Typically the camera will have several file naming options. These may include:

- Sequential numbering which resets each time a memory card is formatted or a new card is put in.
- Sequential numbering which loops from 0001 to 9999.
- Numbering based on date-photo sequence.

As a general rule, the most useful setting is a date-photo sequence, then the 0001-9999 loop and the least useful the card based sequence. Projects must carefully review the naming options of the project cameras for the most useful convention. Projects must also ensure the same convention is used by all cameras collecting data photos.

17.2.2. Scanning Specifications

17.2.2.1. Print Material Types

The resolution is selected based on the size of the original. The smaller the photo, slide, or other material, the higher the resolution should be used to acquire a detailed scan. Below are the **minimum** recommended scan resolutions for different formats.

Higher resolution scans will yield larger and better quality images.

- **35mm color slide or negative**
  (scanner should be set up to scan transparent materials - both hardware and software)
  - Choose source size of approx. 1.3 x .85 inches (software should auto-detect this exact dimensions)
  - Choose target size same as original
  - Choose resolution 2400 dpi, 24 bit color (do NOT use 32 or 48 bit color)
  - Save as uncompressed TIF file
  - scan to yield file size approximately: yields file size approximately 3120 x 2040 pixels, file size 15-20MB
- **3 1/2" x 5" color photograph**

    - Choose source size 3.5 inch x 5.0 inch (software should auto-detect the exact size)
    - Choose target size the same as the original
    - For color photo, choose resolution 600-700 dpi, 24 bit color (do NOT use 32 or 48 bit color)
    - For black & white photo, choose resolution 600-700 dpi, 8 bit grayscale (do NOT use 16 bit grayscale)
    - Save as uncompressed TIF file
    - scan to yield file size approximately: yields file size approximately 3000 x 2100 pixels, file size 15-20MB
- **4" x 6" color OR black & white photograph**
    - Choose source size 4.0 inch x 6.0 inch (software should auto-detect the exact size)
    - Choose target size the same as the original
    - For color photo, choose resolution 600 dpi, 24 bit color (do NOT use 32 or 48 bit color)
    - For black & white photo, choose resolution 600 dpi, 8 bit grayscale (do NOT use 16bit grayscale)
    - Save as uncompressed TIF file

  scan to yield file size approximately: yields file size approximately 3600 x 2400 pixels, file size 15-20MB
- **5" x 7" color OR black & white photograph**
    - Choose source size 5.0 inch x 7.0 inch (software should auto-detect the exact size
    - Choose target size the same as the original
    - For color photo, choose resolution 600 (450 if controls allow it) dpi, 24 bit color (do NOT use 32 or 48 bit color)
    - For black & white photo, choose resolution 600 (450 if controls allow it) dpi, 8 bit grayscale (do NOT use 16bit grayscale)
    - Save as uncompressed TIF file
    - scan to yield file size approximately: yields file size approximately 4200 x 3000 pixels, file size 20-25MB for color

- **8" x 10" color OR black & white photograph**
    - Choose source size 8.0 inch x 10.0 inch (software should auto detect the exact size)
    - Choose target size the same as the original
    - For color photo, choose resolution 300 dpi, 24 bit color (do NOT use 32 or 48 bit color)
    - For black & white photo, choose resolution 300 dpi, 8 bit grayscale (do NOT use 16bit grayscale)
    - Save as uncompressed TIF file
    - scan to yield file size approximately: yields file size approximately 4200 x 3000 pixels, file size 20-25MB for color
- **8 1/2" x 11" typewritten/printed paper**
    - Scan at resolution 300 dpi or 400 dpi if the text has very small print
    - Save as uncompressed TIF file
- **9" x 9" aerial photograph**

- If possible scan from diapositive transparencies rather than prints
- Scan at resolution 1200 dpi or as high as possible/feasible
- Scan with sharpness set to extreme
- Save as uncompressed TIF file

17.3. Processing Photos

Effectively dealing with hundreds of photos requires consistent downloading, naming, editing and documentation. This section describes the general process for managing photos and will provide additional information for each type of photo as defined in the Introduction.

After the images have been acquired, either by digital camera or scanning, the general processing of photos are as follows:

1. Establish a file organization for photos
2. Rename the photos
3. View, delete and edit the photos
4. Document the photos
5. Prepare photos for on-line long-term storage or off-line archiving.

More "how to" details on organization and processing are specified in the Appendices.

Special note for Data Photos: Data photos are photos collected as part of a documented data collection protocol. The project's data processing protocol should contain a detailed section on processing data photos. This protocol should include information, such as:

- File folder structure conducive to data entry and linking.
- Photo naming standard conducive to data entry and linking.
- Field collection method for uniquely linking each photo to its related field data sheet/record.
- Consistent photo database that facilitates linking of photos to the projects data.
- Step-by-step procedural documentation.

17.3.1. Organization

17.3.1.1. Original and Edited Photos

Raw, unaltered photos should be carefully preserved. The names of folders containing raw photos should clearly indicate that the folder contents are unedited originals. Raw photos folders may include multiple photos of the same subject, blurry pictures, or other less desirable photos. This original set of photos should be preserved as is. Copies of raw photos should be saved in clearly named folders for review and editing. For example:

/Subject_A
/Originals – raw, downloaded photos, including poor photos
/Edited –photos which have been processed: renamed and edited.

After downloading or scanning photos to the /Originals directory, this directory should be set to "read-only" permissions to prevent inadvertent edits.

17.3.1.2. Working Photos

It is unlikely that a photo will go directly from the scanner or camera exit port to the library photo folder or to the project data repository where it will live forever. Photos need to be processed before they are ready to be used in a database or posted to the photo library.

Photos being processed, or working photos, should be stored either in a folder within a specific project directory on the data server or within a user's photo library on their computer, depending on the individual project. If the photos pertain to a specific monitoring project, then they should be stored in the appropriate directory with in the project folder on the data server created by the data manager.

Example working photo directory:

\Working_Photos
    \User_A
        \Photo_SetA_Name
            \Originals
            \Edited
        \Photo_SetB_Name
            \Originals
            \Edited

17.3.1.3. Data Photos

3.1.3.1 Project Data Photos
Photos taken as part of a project's data collection protocol are project data that need to be organized, documented and preserved in conjunction with all other project data. Project data photos should be organized and contained within the project folders. Detailed project protocols should define how and where photos are downloaded, edited and rolled up into final folder locations.

It is recommended that a specific method of organizing data photos within project folders to maintain some consistency from project to project. For example:

\Project_A
    \Data
        \Photos
            \Originals
            \Miscellaneous
            \DataPhotos

It is recommended photos **not** be stored or embedded within a MS Access database. A photo linking tool between the photographs and the database should be used.

17.3.1.4. Miscellaneous Project Photos

Incidental or opportunistic photos taken by project personnel that are not data photos can be managed as miscellaneous photos. Miscellaneous photos taken as part of a project should be stored in the project miscellaneous photos directory. This allows the photos to stay with the project, but does not confuse photos that are data related. These photos may be further processed to become "Library" photos. Photos of interest to a greater audience should be copied to the network Digital Image Library.

17.3.1.5. Library Photos

Library photos are public photos readily available to be used for a variety of purposes and are not project specific (i.e. park photos used in a PowerPoint presentation). High quality photos are encouraged where possible, but may also include lower resolution photos. These photos have been reviewed, processed, documented and ready to be stored in a digital photo repository.

Library photos are organized by logical theme keywords related to the parks. The generic organization structure template is lengthy and subject to updates (Figure 1). It is encouraged the folder names remain the same for all park folders, however, in order to provide some consistency in search functions since folder names automatically become key words in Thumbs Plus.

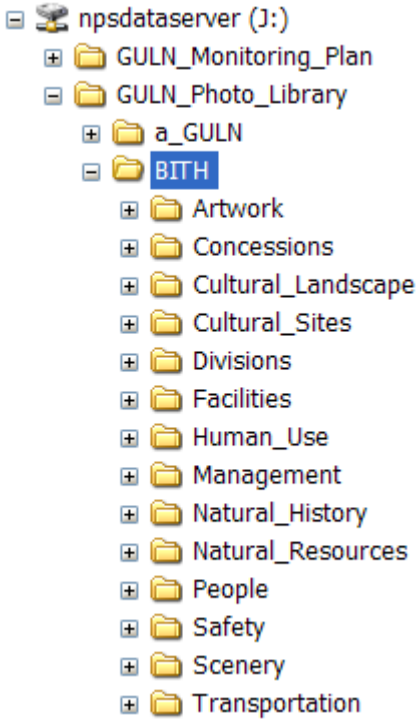


Figure 17.2. Sample of GULN Photo Library directory structure.

17.3.2. Image Naming Standards

17.3.2.1. GULN Photo Library Naming Standards

In order to maintain consistency in naming conventions throughout the photo library, it is recommended that the following naming strategy be adopted when adding images to the GULN photo library.

The image file name should consist of two parts:
1. A brief description of the image
2. A sequential number

Examples:
Black_Bird_025.jpg
Jackson_Monument_001.jpg

Note: Park Code is inherent in the directory structure and the date is attached to the picture's EXIF information by the digital camara which can be read by ThumbsPlus 7 software. If they are not, they should be included in the photo name.

17.3.2.2. Data Photo Naming Standards

These naming guidelines are designed for projects which collect hundreds if not thousands of photos.

File names should assist in the linking of the projects data and the photograph. Project needs may be driven by site, time, specimen, or method. Projects with a limited number of photos (<50) may elect to be descriptive with file names. Projects with larger number of images (>50) may use a sequential image naming standard. Provided here are various examples. In all cases, images should follow these guidelines:

- Document the naming standard used. This should be done in the projects Data Processing Protocol.
- No spaces in the file name; generally less than 20 characters
- Park code and year should either be included or conclusive by the directory structure.
- Use underscores to separate components

Instructions for Assigning Image File Names
Park code and year are included in the file name. The image file name should consist of four parts:

1. The park code (e.g. GUIS, JELA, BITH, PAIS)
2. The date of the image written as mmddyy
3. Project Code or description
4. Photo number ID

Examples:
BITH_012506_Lichens_1.jpg
VICK_061206_VPlants_001.jpg
JELA_2001_VPlants_002.jpg
GUIS_113005_SiteA_001.jpg

17.3.2.3. Voucher and Archive Photos

In many situations, vouchers can be obtained through an electronic format such as digital picture of a rare specimen or an audio recording of a bat species. Each photo or audio recording will be logged as a voucher specimen, will be assigned an accession and catalog number, and then entered into ANCS+.

In some cases, NPS Archives may scan images and store them electronically. They generally consist of legacy images that were already within the archives. These are usually images that were not previously generated electronically by the project. These photos are tied to a particular accession and catalog number, and hence, the image name reflects these numbers.

Images from more recent projects are provided to NPS Archives, usually on a media such as a CD or DVD. In this case, it is the CD itself that is being archived with an accession and catalog number, and not the individual images.

Instructions for Assigning Image File Names
The image file name should consist of three parts:

1. The park code (e.g. GUIS, JELA, BITH, PAIS)
2. Scientific name (for vouchers only)
3. 5-digit Accession Number, assigned by authorized staff
4. 5-digit Catalog Number, assigned by authorized staff

Example:
JELAB_Micropterus_Salmoides_00301_00001.jpg
JELAB_00305_00002.jpg

17.3.2.4. Aerial Photos Naming Standards

Instructions for Assigning Image File Names
The image file name should consist of five parts:

1. Park Code
2. Year
3. Roll
4. Flight line
5. Frame
6. Scanning level

Note: Most aerial photos are stored as TIFF or MrSID.

Example:
PAIS_970122035-12ext.tif
Where year = 1997, roll = 01, flight line = 22, frame = 035, and scanning level = 1200 dpi extreme resolution.

17.3.2.5. Oblique Aerial Photos Naming Conventions

Instructions for Assigning Image File Names
The image file name should consist of five parts:
1. Park Code
2. Word describing general location and/or event
3. GPS coordinates in UTM NAD83 (if available)
4. 2 digit year (if GPS coordinates are not available then include entire data, ex 082805= August 28, 2005)

Example:
JELA_Katrina_Chal05_235684_6584565
Where location= Chalmette Battlefield, Event= Hurricane Katrina, Year= 2005, UTM East= 235684, UTM North= 6584565

17.3.3. Image Viewing and Editing

Software to view and edit photographs are plentiful and are subject to individual preference. Because of this, no specific software is being recommended at this time.

To ensure consistent results and procedures, multi-year projects collecting photos as data should select a standard program to view and to edit photos. Standard software is also a fundamental project requirement to ensure a useful photo processing protocol. Standard software also facilitates communications between project personnel and minimizes confusion.

17.3.3.1. Viewing

Thumbnails give a quick icon view of many photos so photos may be quickly selected visually. It is recommended that 1) a quick viewing software or database be used to look at thumbnails, captions, descriptions, dates and keywords, and where appropriate, 2) a contact sheet of these thumbnails be printed for quick reference.

17.3.3.2. Editing

At a minimum, photos should be edited as follows:

- Poor quality photos should be deleted, except where the subject is highly unique.
- Medium quality photos should be assessed against existing photos of the same subject in the park photo library. If the photos duplicate the subject with no enhancement of quality or perspective, the photo may not be worth saving and should be deleted.
- Photos should be rotated to portrait or landscape.
- Photos should be rotated to make the horizon level.
- Photos of people should have 'red eye' removed.
- Photos should be cropped to remove edge areas that grossly distract from the subject.

Large groups of photos acquired under sub-optimal exposure or lighting can be batch processed to enhance contrast or brightness. Batch processing can also be used to resize groups of photos for use on the web.

17.3.4. Image Documentation and Cataloging

Images should be documented and cataloged. Documentation, or metadata, provides the minimum information a user will need to appropriately use the photo. Cataloging provides a collective means for searching, finding, and retrieving photos.

17.3.4.1. Software

A variety of software is available to facilitate this process. NPS does not yet have a standard software package to document, catalog and effectively use photos at the park level. The GULN has decided to use ThumbsPlus 7 (TP7) software to manage and document digital image files. Refer to Appendix A for a SOP of how to add photos to the GULN library and create metadata using TP7.

17.3.4.2. Minimum Metadata Attributes

All photos should be documented with these minimal metadata attributes:

- File name (usually inherent)
- File location (usually inherent)
- Storage Location of the original/hardcopy/negative
- Description
- Photographer, Contributors, or Archive Institution (all that apply)
- Collection Name, if applicable (historic and archives)
- Publisher
- Date Taken
- Format/Size/Source
- Type of Media of original

- Subject
- Keywords
- Park Code
- Place
- Any necessary credits
- Distribution Restrictions, such as copyright or sensitivity

Digital photos capture some metadata and will hold this information in an EXIF file that stays with the photograph until the photograph is modified. These include:

- File name
- Aperture
- Date Digitized
- Date Taken
- Exposure Bias
- Exposure Time
- F-Number
- Flash (yes, no)
- Focal Length
- ISO Speed
- Light source
- Metering Mode
- Shutter Speed

Additional Attributes may include (IPTC standards):

- Special instructions for how the photo was taken
- Subject distance
- Caption writer
- City, State, Country
- Equipment make
- Equipment model
- Object name

If the image is processed, for example cropped or modified in Photoshop, the following attributes should be documented:

- Height
- Width
- Artist
- Artist Copyright
- Creating Software
- Date
- Model of printer intended

Special Note for Data Photo Metadata: Projects should store data photo metadata attributes in the appropriate relational tables of the project database. Project databases should contain at least the minimum metadata requirements listed here.

17.4. Photograph Property and Use

All photos collected with the National Park Service funds and staff time are property of NPS. Contractors using photographs as part of their project should provide copies, preferably high resolution digital copies, to the NPS project manager.

When using a photograph, provide credit to the photographer. This is usually written on the right side of the photograph or at the bottom in a smaller font size (san serif) than the text in the document.

17.5. Photos of People and Rights to Privacy

When taking a photo of a person, the subject's right to privacy may come into play. Photos with the following criteria should seek a model waiver form:

- The person is recognizable, AND
- The person is not a government employee (on-duty), AND
- The photo will be used for profit to the photographer.
  OR
- Any photos of a minor, where the minor is recognizable.
- Photos with the following criteria do not need a model waiver form:
- Photos are public domain (unless protected by law)
- Photos are not for profit
- If photos are used for profit, such as in a magazine, it is not NPS who profits

In the case of NPS, it is rarely the case where a model waiver form is needed. Typically these photos are used for education and editorial purposes, where photos of subjects are acceptable.

17.6. Acknowledgements

I would like to thank the Southwest and Southeast Alaska networks for writing much of the material within this document.

17.7. References

Southwest Alaska Network and Southeast Alaska Network. 2004. Digital photograph management strategy for Alaska Inventory and Monitoring Program. Anchorage, Ak.: National Park Service. 61pp.

Evans, K. 2003. National Park Service focus image creation and management guide. Draft Rev. January 23. http://focus.nps.gov/doc/techdoc/ImagingGuide.html. Last accessed January 7, 2004.

Duboff, L. D. 1989. The photographer's business and legal handbook. New York: Images Press.

18. Appendix E: File Naming Conventions

Version date: March 13, 2006
Last modified: December 12, 2006

Author(s) of Sheet: Dorothy Mortenson, National Park Service, Inventory and Monitoring Program, Southwest Alaska Network; modified by Whitney Granger, Inventory and Monitoring Program, Gulf Coast Network
Website: http://www1.nature.nps.gov/im/units/swan/index.cfm?theme=info_guidelines

The purpose of this appendix is to provide a quick reference in file naming standards.

18.1. General information

Within any file created, be sure the following is documented:

Purpose
Who wrote the document
Date
Last modified
Control version by using date.

Example: The header for this "Cheat Sheet" provides an example.

18.2. Folder and file naming standards

Keep names short if possible but make them meaningful and intuitive. Avoid spaces, unusual characters (like % or & or /), or reserved words (like DATE) in both folder and file names. This is REQUIRED, WITH NO EXCEPTIONS.

18.2.1. Report File Names: Required format

Report files should be named in the following manner:
AuthorLastNameFirstInitial_YEAR_ParkCode_BriefTitle.pdf

AuthorLastName = The first author's last name or the agency office (i.e., SERO)
FirstInitial = First initial of the first author.
YEAR = Four digit year of the publication date.
ParkCode = Four letter park code if relevant and not included in report title.
BriefTitle = Provide a descriptive, but brief title. Concatenate with first letter of each word capitalized and no spaces.

For example: Schultz_2003_ InventoryOfFishOnJeanLafitteNationalPark.doc

All reports must be added to the GULN Reference Manager database and NRBib. Refer to Appendix B.

18.2.2. Data

Tabular data, such as MS Access databases, MS Excel spreadsheets, or others, should be named

in the following manner:
ParkCode_Year_ShortDescrip_Stage_versiondate.mdb or .xls or .txt

ParkCode = Four letter park code
YEAR = Four digit year of the publication date.
ShortDescrip = Provide a descriptive, but brief title. Concatenate with capitol letters and no spaces.
Stage = Stage of the database.

For example:
InDesign – The database is in design and no real data is included
RAW – Contains only the raw data; no QA/QC procedures are done
QAQC – Data has been in some state of QA/QC
Final – Data has completed all stages, but there still may be some versioning, which is controlled by date. Select the latest data for the most current database.

18.2.3. Image File Names

Please refer to the GULN photo management strategy (Appendix D).

18.2.4. Exceptions

There is no exception for the Report File Naming. On occasion, however, a different file naming strategy may be needed for data files or images. For example, it may be important to keep track of the author of data. In this case, you may find it more intuitive to use the Report Naming Standard. All names should indicate at a minimum which park, date, and subject. It should be easy to tell what the most current dataset is.

19. Standard Operating Procedures for Addressing Document Versions and Updates

Gulf Coast Network
Version 1.0
December 2007

Revision History Log:

Prev. Version #	Revision Date	Author	Changes Made	Reason for Change	New Version #

The final GULN monitoring plan submitted to WASO in September 2007 and was published in the Natural Resource Report Series. As the monitoring program develops, we will continue to make changes and updates to the monitoring plan and its' protocols. This SOP documents how the network will handle version changes and updates of all official monitoring reports and protocols developed as part of the monitoring program.

19.1. Updating the GULN Monitoring Plan and Appendixes (including Data Management Plan).

The monitoring plan has a revisions history log on page 3, which needs to be updated any time there is a meaningful change made to the plan or one of the appendixes. Spelling, grammar, and wording, do not constitute a meaningful change. If a change is made to the monitoring plan, be sure to document the section the change was made in using the "Changes Made" row of the history log. The same goes for changes to the appendixes; be sure to note which appendix was revised in the monitoring plan history log. Furthermore, each appendix has a revision history log that needs to be updated to reflect document changes. Specifics should be given as to what section was changed and why.

In addition to documenting changes in the revision log, the plan and appendices must be replaced with the new version on the data server and the GULN internet and intranet websites. The final monitoring plan can be found on the GULN Data Server at J:\GULN_Monitoring_Plan\Final_Monitoring_Plan\GULN_Monitoring_Plan_Final.

On the internet and intranet, the plan can be downloaded from the reports page (INTERNET http://science.nature.nps.gov/im/units/guln/networkhome/reports.cfm) (INTRANET http://www1.nrintra.nps.gov/im/units/guln/networkhome/reports.cfm) and the monitoring page (INTERNET http://science.nature.nps.gov/im/units/guln/networkhome/vitalsignsmonitoring.cfm) (INTRANET http://www1.nrintra.nps.gov/im/units/guln/networkhome/vitalsignsmonitoring.cfm). All links should be updated with the new version of the Plan and the appendixes.

19.2. Updating GULN Protocol Development Statements (PDS) and Vital Signs Monitoring Protocols.

Each GULN PDS and monitoring protocol contains their own History Revision Log. When changes are made to these documents, the Revision Log must be updated and the version number on the document changed. There is no need to make any changes to the monitoring plan revision log because PDSs and protocols are independent documents that reference the monitoring plan. The only exception is the Freshwater Water Quality monitoring protocol. This is the only protocol that is embedded in the appendixes of the monitoring plan. When updating the WQ protocol and SOPs, follow the directions stated in the previous section.

PDSs and protocols also reside on the network internet and intranet. Each vital sign has a web page and a link to the PDS and protocol. Be sure to update both links on the pages.

The Department of the Interior protects and manages the nation's natural resources and cultural heritage; provides scientific and other information about those resources; and honors its special responsibilities to American Indians, Alaska Natives, and affiliated Island Communities.

NPS D-XXX, Month Year

Natural Resource Program Center
1201 Oakridge Drive, Suite 150
Fort Collins, CO 80525

www.nature.nps.gov