National Park Service
U.S. Department of the Interior

Natural Resource Program Center

Inventory and Monitoring Program

Data Management Guidelines for Inventory and Monitoring Networks

Natural Resource Report NPS/NRPC/NRR—2008/035

Note – If this plan will be used as a template an error occurs that I cannot correct (a Word wizard may have more luck). The fourth physical page of this document is invisible and hiding. It does not show up in Word, but when printed to PDF is does show up as a blank page that should be deleted. Once this occurs the page order will be correct.

To do this in Acrobat navigate to the blank page, go to the Document tab > Delete pages, and delete page 4.

Also, blank pages should have all footer info – including page number – deleted.

In Acrobat Go to Tools > Advanced Editing > Touchup Object Tool. On each blank page select footer info and delete.

This can be done in Word but it is much less painful in Acrobat (for some of us anyway).

For a NRR Series Blank Template go to:

http://www.nature.nps.gov/publications/NRPM/index.cfm

Delete this text box!

Thanks,

J. Bennett

ON THE COVER
Mesa Arch, Canyonlands National Park. NPS Photo

Inventory and Monitoring Program

Data Management Guidelines for Inventory and Monitoring Networks

Natural Resource Report NPS/NRPC/NRR—2008/035

National Park Service
Natural Resource Program Center
Office of Inventory, Monitoring, and Evaluation
1201 Oakridge Drive, Suite 150
Fort Collins, Colorado 80525

May, 2008

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; "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 (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:

National Park Service. 2008. Data management guidelines for inventory and monitoring networks. Natural Resource Report NPS/NRPC/NRR—2008/035. National Park Service, Fort Collins, Colorado.

NPS D-1968, May 2008

Contents

Tables

Figures

Appendices

Change History

Version numbers will be incremented by a whole number (e.g., Version 1.3 to Version 2.0) when a change is made that significantly affects requirements or procedures. Version numbers will be incremented by decimals (e.g., Version 1.6 to Version 1.7) when there are minor modifications that do not affect requirements or procedures included in the plan.

The following revisions have occurred to this plan since May 30, 2008

Version #	Date	Revised by	Changes	Justification
1.1	8 July 08	J. Bennett	Formatting	To Follow NRR formatting guidelines more closely

Acknowledgements

This document is truly a product of many efforts. The community of Inventory and Monitoring Program data managers is an unprecedented assemblage of talent that is contributing much to the advancement of data and information management within the National Park Service. This group has fostered a collaborative environment with the free sharing of information and ideas for the benefit of all.

Persons who have developed and contributed material presented in this plan and who deserve specific credit include: Debbie Angell, Margaret Beer, John Boetsch, Kristina Callahan, Rose Cook, Vel Decker, Gordon Dicus, Fred Dieffenbach, Patrick Flaherty, Brent Frakes, Ulf Gafvert, Whitney Granger, Mark Hart, Pat Lineback, Scott Miller, Sean Mohren, Dorothy Mortenson, Nathan Piekielek, Dave Press, Tom Richie, Gareth Rowell, Geoff Sanders, Dennis Skidds, Heidi Sosinski, Sara Stevens, Leona Svancara, Nicole Tancreto, Helen Thomas, Bob Truitt, Doug Wilder, Mike Williams, Brian Witcher, and Christina Wright.

In addition, national-level data management staff have provided the vision and created the tools that have been essential to our work. In particular, our thanks go to Peter Budde, Chris Dietrich, Kathy Dratch, Joe Gregson, Willene Hendon, Simon Kingston, Alison Loar, Danelle Malget, Lisa Nelson, Wendy Schumacher, and Mark Wotawa for their support, hard work, and dedication.

We thank Gary Williams, who has administered the I&M Program with a careful and steady hand over the years.

Finally, we thank Steve Fancy for his ongoing commitment to data management in the National Park Service, and for the insight and guidance he has provided to all of us.

CHAPTER 1. Introduction

1.1 Background

Information is the common currency among the activities and staff involved in natural resource management in the National Park Service (NPS). The central mission of the National Park Service's Inventory and Monitoring (I&M) Program is to acquire, manage, analyze, and distribute scientific information on the status and trends of specific park natural resources (National Park Service 1998). Intended users of this information include park managers, planners, cooperators, researchers, and the general public.

Under the program, more than 270 park units with significant natural resources have been organized into 32 I&M networks to facilitate baseline inventories and to conduct long-term resource monitoring (Figure 1-1). Each network links parks that share similar geographic and natural resource characteristics to improve efficiency and reduce costs. Parks within a network share funding and professional staff to avoid duplication of efforts.

Figure 1-1. Map showing the names and boundaries of the 32 I & M Program networks.

A cornerstone of the I&M Program is the strong emphasis placed on data management. Data management refers to the framework by which data are acquired, maintained, and made available. A robust data management system is particularly important for the success of long-term programs where the lifespan of a data set will extend across the careers of many scientists. All I&M networks are expected to invest at least one-third of their available resources in data management, analysis, and reporting activities.

The goal of each network's data management program is to maintain, in perpetuity, the ecological data and related analyses that result from the network's resource inventory and monitoring work. The purpose of this document is to describe the overarching policies relevant to network data management, to establish objectives common to all networks, and to provide national guidance that will ensure the highest standards for data acquired and managed by all networks. These standards include:

Accuracy: The quality of the data collected and managed by the I&M Program is paramount. Analyses performed to detect ecological trends or patterns require data with minimal error and bias. Inconsistent or poor-quality data can limit the detectability of subtle changes in ecosystem patterns and processes, lead to incorrect interpretations and conclusions, and could greatly compromise the credibility and success of the I&M Program. To ensure that networks produce and maintain data of the highest possible quality, procedures are established to identify and minimize errors at each stage of the data life cycle.

Security: Digital and hard-copy data must be maintained in environments that protect against loss, either due to electronic failure or to poor storage conditions. Networks must have in place proper storage and backup procedures and disaster recovery plans, as well as established records management process. In addition, collaboration with the NPS Museum Management Program enlists the expertise of museum curators and archivists to ensure that related project materials such as field notes, data forms, specimens, photographs, and reports are properly cataloged, stored, and managed in archival conditions.

Longevity: Countless data sets have become unusable over time either because the format is outdated, or because metadata is insufficient to determine the data's collection methods, scope and intent, quality assurance procedures, or format. While proper storage conditions, backups, and migration of data sets to current platforms and software standards are basic components of data longevity, comprehensive data documentation is equally important. Networks must ensure that data sets are consistently documented, and in formats that conform to current federal standards.

Usability: One of the most important responsibilities of the Inventory and Monitoring Program is to ensure that data collected, developed, or assembled by staff and cooperators are made available for decision-making, research, and education. Providing well-documented data in appropriate formats and in a timely manner to park managers is especially important to the success of the program. Networks must ensure that:

- data can be easily found and obtained
- data are subjected to full quality control before release
- data are accompanied by complete metadata
- data are provided in formats that are most useful to end users
- sensitive data are identified and protected from unauthorized access and distribution.

1.2 Objectives of this Document

This concept of a national-level data management plan was developed as the result of over 25 individual network plans being written between 2003 and 2006. Network data managers realized that significant portions of their plans were common to all networks; in particular, sections

describing the overarching objectives and policies. By establishing a national-level plan containing these common elements, network data managers could concentrate instead on establishing, documenting, and updating their specific network-level guidelines and standard operating procedures (SOPs).

This document, which establishes national-level data management guidance, is intended as a reference that guides and complements the specific procedures that are developed by each network. This plan outlines the I&M Program's approach to establishing and maintaining systems that serve the data management needs of the networks, their member parks, and the national I&M Program. It is intended to be a guide for current and future network staff, to ensure the continuity and documentation of data management methods and procedures, to establish best practices, and to provide a baseline of standards common to all networks.

Network data management plans are intended to broadly describe how the national plan will be implemented at the network level, provide general orientation on a network's data management program (e.g., infrastructure, staffing), and direct the reader to more detailed guidance documents and SOPs (Figure 1-2).

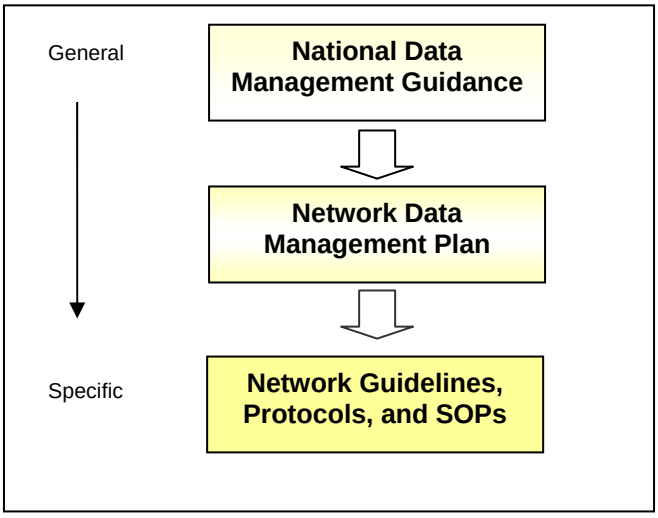

Figure 1-2. Relationship of the national-level data management guidelines to network-level data management documents

Network data management guidelines typically contain explanations of procedures that apply to multiple situations; for example, file or field naming, backup and restore procedures, or website management. Protocols and SOPs provide an additional level of detail, with step-by-step procedures for situations where there is little or no tolerance for variation. Together these documents and procedures ensure that processes that are critical for I&M network continuity and data integrity are maintained, even in the event of staff or program changes.

1.3 Organization

This document offers a guide for I&M network programs and provides the foundation for implementing and maintaining data management systems and best practices. Each of the chapters contains the following sections:

- *Objectives*: Specific statements describing the results to be achieved by the procedures contained within the chapter
- *Laws and Policies*: Requirements and sideboards established through laws, policies, and mandates at different levels of governance, ranging from broad federal law, to Department of Interior (DOI) and NPS policies, to policies developed specifically for the Inventory and Monitoring Program.
- *General Standards and Guidelines*: Generally-accepted I&M network standards and guidelines applicable to specific chapters.
- *References and Resources*: Citations of or links to specific laws, policies, or documents pertinent to chapter content.
- *Related Guidance Documents and/or SOPs*: A listing of network-level guidance or procedural documents that may be needed to accomplish the stated chapter objectives.

1.4 Key Terms and Definitions

Since the principal purpose of this document is to provide standards and guidelines for the development, management, and dissemination of data, it is important to apply consistent terminology. A glossary of terms and acronyms used both within this report and by I&M networks can be found in Appendix A; however, the following definitions of general terms provide the basis for key concepts in this and following chapters:

Data are distinct measurements or observations of a variable, usually formatted in a special way: they include symbols or representations of facts or ideas that can be communicated, interpreted, or processed by manual or automated means.

Raw data are data in their original form, i.e., data that have not been altered, summarized, or grouped into broader categories. Raw data can exist in many forms: as hand-written information on field data forms and in notebooks, as unaltered photographs, sound and video recordings, data logger files, remote sensing imagery, and Global Positioning System (GPS) files. They may be typed or uploaded directly into a computer upon acquisition, or transcribed later from data sheets.

Derived data are raw data that have been processed, or converted to another form using some automated or manual process. Raw natural resources data are often processed and packaged for summation, statistical analysis, and graphical display, or the production of maps and other information products.

Information is created from data as a result of processing, manipulating, synthesizing, or organizing data in a way that provides interpretation or meaning.

Tabular data are usually organized into logical tables of records and attributes, arranged in a matrix of rows and columns. Tabular data can be displayed, manipulated, and stored as simple text files or in applications software (e.g., spreadsheets, relational databases).

Spatial data are any data that reference geographic coordinates. Tabular data that contain spatial references are also considered spatial. For example, a table of wildlife observations might include the x and y coordinates of the location of observation, along with other information, such as species identifier, observer name, and date of observation.

GIS (geographic information systems) data contain information about the location and shape of, and relationships among, features on the surface of the earth and are usually stored as geographic coordinates and topology. Topology is used to compare the geographic locations of features relative to one another (e.g., roads connected to a highway, two vegetation polygons adjacent to one another). GIS data can be categorized as either vector or raster and require specific software applications for processing and manipulation.

Legacy data are existing data that may be at risk of becoming obsolete due to software or metadata limitations, technological changes, or the loss of institutional knowledge.

Programmatic data are data developed specifically for a project with well-defined objectives and specific requirements describing data use and application. As an example, satellite imagery may be acquired for a geographic area to provide an analytical data framework for monitoring landscape change over a period of time.

Non-programmatic data are data developed by other programs or organizations, but that are of value in meeting network objectives. For example, weather data acquired by the National Weather Service, or taxonomic information managed by the US Department of Agriculture.

Sensitive data are data that through loss, unauthorized access, or modification, could be used in such as way as to adversely affect valuable resources, the national interest, the conduct of federal programs, or individual privacy. Examples of sensitive natural resources data might include the locations of rare flora or fauna species, caves, or cultural sites.

Metadata are information about data. A complete set of metadata describes all aspects of the subject data, including the who, what, where, when, how, and why of a data set, along with any processing that has occurred. Metadata are considered an essential component of any good data set. The term also applies to the synoptic information for a document, publication, sample, or other physical material.

1.5 Types of data covered by this document

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 resulting 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. To maintain this context, this document encompasses a range of products that are coordinated or managed by networks or under their auspices. These products fall into six general categories: raw data, derived data, information, documentation, reports, and administrative records (Table 1-1).

Table 1-1. Categories of data products and project deliverables

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

1.6 Data Sources and Priorities

There are multiple sources of data related to natural resources in network parks. The types of work that may generate these data include:

- natural resource inventories
- vital signs monitoring
- protocol development pilot studies
- special-focus studies performed by internal staff, contractors' or cooperators' external research projects, or studies performed by other agencies on park or adjacent lands
- resource impact evaluations related to park planning and compliance
- resource management and restoration work.

Because the I&M Program focuses on natural resource inventories and long-term monitoring, a network's first data management priority is the data and information that result from these efforts. However, the standards, procedures, and approaches to data management developed by networks can also applied to other natural resource data sources.

Considering the volume of data that has been produced in park units, priorities must be set for network data management efforts. As a general rule these are:

1. Produce and curate high-quality, well-documented data originating with the Inventory and Monitoring Program

2. Assist with data management for current projects, legacy data, and data originating outside the Inventory and Monitoring Program that complement program objectives

3. Help ensure good data management practices for park-based natural resource-related projects.

1.7 Maintenance of this Document

This document will be updated as needed to accommodate the inevitable changes that will occur in laws, policies, technology, or general data management practices. Recommendations for changes can be forwarded to the national Inventory and Monitoring Program data manager by any interested party or user of this document (e.g., park resource managers, project leaders, technicians, superintendents, external users). Simple changes can be made immediately in the document, while substantive changes will be made during version updates and in consultation with network data managers.

Document versions will be designated by using version numbers in parentheses that are appended to the heading of the modified section. For example, the first version of a section in Chapter 4 would have the heading "4.1." A revision to this section would have the heading "Chapter 4. (v2)." The Change History log at the front of this document contains an update schedule and key that identifies and summarizes changes. Previous versions will be archived in their entirety for reference.

A working version of this document will be available on the Natural Resource Program Center SharePoint site: (http://nrpcsharepoint). Finalized versions will be available on the Inventory and Monitoring Program Internet site (http://science.nature.nps.gov/im/datamgmt).

1.8 References and Resources

National Park Service. 1998. National Parks Omnibus Management Act of 1998. Public Law 105-391. (http://www.nps.gov/legal/laws/105laws.htm)

National Park Service. 1999. Natural Resource Challenge: the National Park Service's action plan for preserving natural resources. (http://www.nature.nps.gov/challenge/)

National Park Service. Natural Resource Program Center SharePoint. (http://nrpcsharepoint). Accessed 16 January 2008.

National Park Service. 1999. Natural Resources Inventory & Monitoring. Director's Order #70, NPS-75. (http://www.nature.nps.gov/nps75/nps75.pdf)

National Park Service. Natural Resources Inventory & Monitoring Program, data management. (http://science.nature.nps.gov/im/datamgmt). Accessed 16 January 2008.

1.9 Related Network-level SOPs

- Development and review process for network guidance documents
- Template for network SOPs

CHAPTER 2. Infrastructure and Systems Architecture

Modern information management infrastructure and systems architecture represent the foundation of network information management systems. *Infrastructure* refers to the system of computers, servers, and related hardware that is functionally or directly linked through computer networking services. *Systems architecture* refers to the applications, database systems, repositories, and software tools that make up the framework of a network's information management enterprise.

2.1 Objectives

The objectives of proper management, configuration, and maintenance of I&M network infrastructure and systems architecture are:

- Ensure staff has appropriate access to electronic files that are secure and protected from accidental or malicious loss
- Simplify software installation and management, and expedite access and retrieval of data using automation tools
- Make available collaborative technologies and tools that enhance networking and communications at both an intra- and interagency level
- Develop and implement efficient means for storing, managing, searching, and disseminating electronic files.

2.2 Laws and Policies

The Department of Interior (DOI) and National Park Service (NPS) have established guidelines and standards for the acquisition, configuration, and deployment of hardware and software in I&M network offices. Specifically:

- All DOI agencies are required to standardize on Microsoft Office products for desktop application and use of spreadsheets, documents, databases, and presentation software (DOI, Chief Information Officer, November 30, 2005). (http://www.doi.gov/ocio/erm/software.html)
- All DOI agencies are required to standardize on ESRI products when using GIS software (DOI, Chief Information Officer, November 30, 2005). (http://www.myinterior.doi.net/ocio/cto/ctoc/Version3.1/ProductsByStandard.pdf)
- All DOI agencies, including the NPS, are required to limit administrative rights to servers and computers to those staff that are required for operational use. All users with administrative rights must have an approved request, with justification, on file with the agency's Office of Chief Information Office (OCIO) (http://classicinside.nps.gov/documents/OCIO%202005-0004.doc).
- Local and wide area networks (LAN/WAN) should conform to Department of Interior and National Institute of Standards and Technology (NIST, SP800-53) security guidelines (http://csrc.nist.gov/publications/drafts/800-53A/draft-SP800-53A-fpd-sz.pdf).

In general, I&M networks will conform to any federal, department, agency or bureau data infrastructure and systems architecture policies and standards. Further, I&M staff will conform to national I&M Program standards and mandates to ensure long-term program integration and information sharing.

2.3 General Standards and Guidelines

I&M networks situated within an NPS office typically rely on park, regional, and national information technology (IT) personnel and resources to maintain their computer infrastructure and systems architecture. Local park IT specialists are primarily responsible for development and management of local infrastructure with oversight and support from regional and national IT programs. Development and management of network architecture is broadly distributed among local, regional and national IT staffs and direct participation by other local and national programs including GIS and I&M staffs.

I&M networks that are hosted by a university or other government agency or organization must collaborate closely with the host in addressing the network's infrastructure needs. Typically this means that an I&M network's local area network (LAN) or each individual computer is connected to a larger LAN or wide area network (WAN) that is maintained by the university or other agency/organization. An agreement is recommended that designates network staff as a point of contact and that staff computers are to be maintained as part of the larger network. The larger host network infrastructure or IT Plan should be referenced in network SOPs so that responsibilities for servers, backup and archiving, Internet connections, and other fundamental services are clearly articulated and understood.

Core infrastructure and systems architecture capabilities for I&M networks should include:

- A network directory structure with a decentralized security system allowing appropriate individuals the ability to manage and maintain file security permissions
- Data structures and data models that optimize flexibility and scalability and integrate with other national applications.
- Software tool(s) that simplify and streamline staff access to spatial data and metadata.
- Installation of software, patches, updates, and tool extensions that is automated and delivered through software installation and management portals
- Reliable high-speed connectivity to the Internet
- Active Directory accounts for all NPS staff and key cooperators to enable access to internal NPS data resources.

2.4 Infrastructure

The foundation of an information management program is a reliable and secure network of computers, servers and an associated backup schema. Other than personal computers, I&M networks may or may not have separate computer infrastructure including file or application servers.

There are multiple components associated with the NPS infrastructure including park, network, regional, and national level systems (Figure 2-2). Each level can be the repository of different components of the information system that are maintained and secured by the

respective IT, resource specialists and/or other responsible staff. This includes hosting and managing electronic files being created, managed, and disseminated by staff and cooperators.

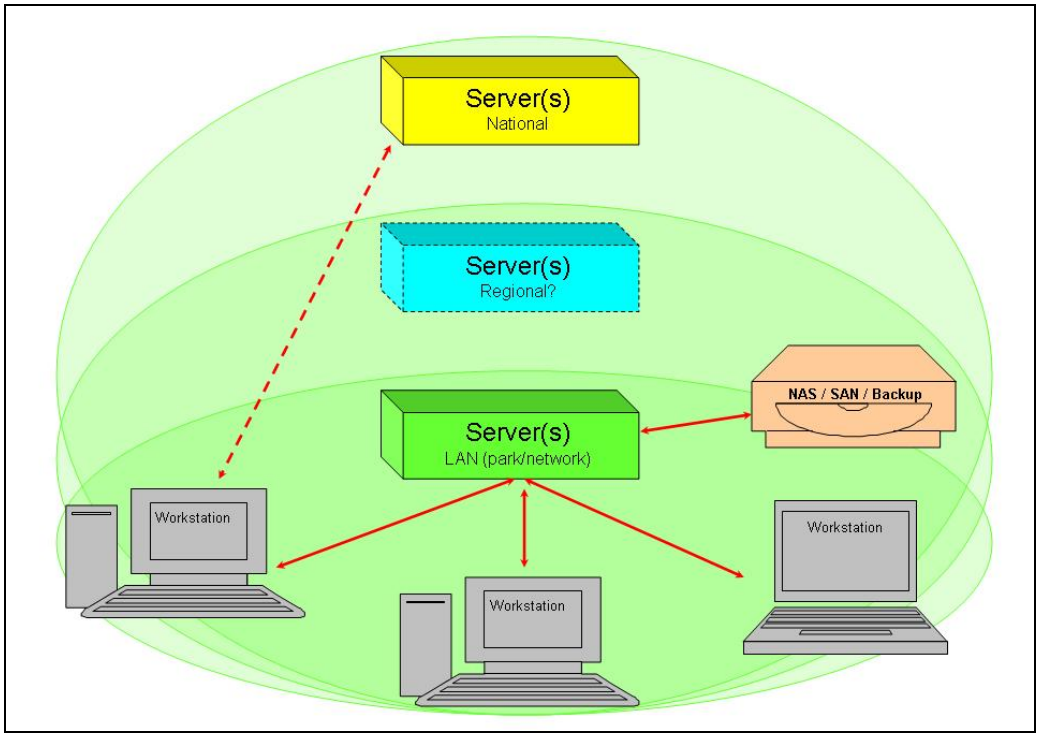

Figure 2-1. Generalized infrastructure components. Dotted arrow depicts ability to log in remotely.

2.4.1 Servers

Computer servers may be broken into five primary functional types (one physical server may perform more than one of the functions listed below):

- *Work File Servers* – Typically team servers or user servers, where anyone can store working files. These servers work similarly to a PC's C: drive, but are located on a server for ease of backups and file sharing.
- *Data File Servers* - Read-only repository of data files. Files stored here do not typically change often and need a consistent location, such as in the case of GIS data, imagery, or final reports. Write access is under strict control, whereas all NPS employees have read access. A contact person monitors what is stored in this location to ensure proper archiving and documentation. These servers are designed for large storage capacity.
- *Database Servers* – Uses special database software, such as SQL Server, Oracle, or ArcSDE. Write access is under strict control. Databases on these servers are for long-term data storage, and require quality database design, documentation, and administration.
- *Application Servers* –Provide the applications, or "front end," to accessing the data. These are closely tied to the data file and database servers, but should be located on a different physical server. Files on these servers change frequently and backups need

to support these changes. Application servers are often Internet servers as well. These servers are designed for speed.

- *Internet Servers* – Provide access to applications and files via the Internet and subject to security restrictions. Special software may be used on Internet servers, such as ColdFusion for web site pages and database applications, and ArcIMS for Internet interactive mapping.

It is not recommended that I&M networks establish public-facing application or Internet servers, which may require an extensive documentation and approval process through the NPS Chief Information Officer. Instead, these functions should be provided by the national-level I&M servers and infrastructure based in Fort Collins, Colorado.

I&M networks will generally use or access one or more of the following servers:

National-level Servers

National servers provide both secure (NPS only) and public workspaces that are accessible via the Internet. They are available to I&M networks for:

- Master applications, e.g., integrated client-server versions of NatureBib, NPSpecies, Research Permit and Reporting System
- Centralized repositories, e.g., NPS Data Store, Monitoring Protocol Database
- Public access sites, e.g., web sites for monitoring networks
- Enterprise-level tools such as SQL Server and SharePoint.

Regional-level Servers (optional)

Depending upon the location of the network offices, regional servers may be used for specific functions.

I&M network-level Servers

A network may have its own server(s) or be integrated into other units such as a park or cooperative institution.

Local- or park-level Servers

In the majority of instances an I&M network is integrated into an existing park unit server(s). Local area networks provide shared workspace for staff, and also provide each staff member with dedicated workspace that is part of a regular backup routine.

2.4.2 Workstations

I&M network staff manage files on their personal computers, and, in most instances, place files that need to be shared on local server(s). Staff are responsible for regularly copying their workstation files to workspace to a network or external drive for backup.

2.4.3 Local Administrative Rights

The DOI and agency offices of the Chief Information Office are responsible for the guidance and oversight of administrative rights on computers and servers. Park, I&M network, and regional IT staff administer the guidance and issue rights as they apply on case-by-case bases; however, a principle of least privileges is to be applied as the best practice (OCIO 2005-0004). (http://classicinside.nps.gov/documents/OCIO%202005-0004.doc)

2.4.4 Network Security

Local and wide area networks should conform to Department of Interior and National Institute of Standards and Technology (NIST, SP800-53) security guidelines. All sensitive electronic files should be placed in protected folders with limited read and write access. Electronic file and directory permissions administration should be partially decentralized with file and folder administrative rights controlled by limited I&M network. A method for tracking of file permissions is recommended. I&M network or park staff needing read or write access to the directory structure should be enabled through Microsoft Active Directory.

There are rare circumstances where encryption of sensitive electronic files and/or folders may be necessary. As an example, files on removable media or portable devices may need additional protection from unauthorized access or use. NPS encryption policy and procedures are under development; data managers should work with a local IT specialist if this capability is required.

2.5 Systems Architecture

Integrated inventory and monitoring of natural resources is complex, multidisciplinary, and requires a broad suite of tools, cooperators, and standards for long-term success.

2.5.1 Digital Data Formats

Databases

Well thought-out database design standards are necessary to promote compatibility among data sets, scalability, and long-term stability. As much as possible, I&M network standards for database design will mirror the NPS Natural Resource Database Template (NRDT). The NRDT is a flexible, relational database core designed for storing inventory, monitoring, and research data (including raw data collected during field studies). The template was designed to be used in the development of standalone databases; however, the basic structure can be implemented in enterprise-level systems (e.g., SQL Server) as well. NRDT also interacts with GIS software (e.g., ArcView or ArcGIS). See Chapter Five for additional information on NRDT.

Tabular data

Tabular data can exist in any number of applications (e.g., word processing, spreadsheets, databases) and formats (e.g., dat, txt, xml, dbf, xls, csv). Standardizing on the Microsoft Office Suite is required. When and where possible it is also strongly advised to document any attribute and table information. Additional guidance on this is provided in Chapter Eight.

Spatial data

I&M network and park unit staff use many different raster and vector data formats for storing and managing spatial data. Vector data formats commonly include Arc/Info coverages, shapefiles, and personal geodatabases. Raster data formats commonly include MRSID, TIFF, GEOTIFF, and Grid structures. ESRI (Environmental Systems Research Institute) software products have historically demonstrated excellent backward compatibility with older data structures; however it is recommended that data be migrated to current versions when possible.

The coordinate system standard for the National Park Service is Universal Transverse Mercator with North American Datum 1983 (UTM NAD 83). Generally, existing spatial data should be migrated to this data projection and it should be used for any new spatial data development.

With the release of ArcGIS 9.2, the *file geodatabase* is available. This data format has numerous advantages including: a) providing a relatively simple and scalable solution for GIS users; b) working across operating systems; c) scaling to handle very large data sets; d) an efficient data structure optimized for performance and storage; and e) increased flexibility for multi-user editing and use of the same spatial data layer.

Image data
Photographic information has become an important component of resource management, and can potentially consume considerable amounts of electronic storage resources. Photographic information in the forms of digital images, scanned photographs, and satellite imagery holds tremendous value and utility; however, the careful, long-term stewardship of these data along with appropriate documentation are both required to avert information loss.

There are a number of photo or image management applications and software programs available both commercially off-the-shelf or easily designed and developed in a relational database. It is recommended I&M networks develop and adhere to a well-documented and disciplined image management schema in order to preserve these assets long-term.

Audio and video media data
Audio and video media can represent actual data values (e.g., hydrophone or Anabat recordings) and also have potential for various reporting and outreach purposes. The enabling technology for networked audio and video is easily available to those with a personal computer, and I&M networks will be increasingly taking advantage of this capability. SOPs that define the best formats and associated software, and address storage capacity issues, will be required as these media are more broadly used.

2.5.2 Database Applications Development

Rather than developing a single, integrated database system, most I&M network database design is based upon modular, standalone project databases that share design standards and links to centralized data tables. Individual project databases are developed, maintained, and archived separately. There are several advantages to this strategy:

- Data sets are modular, allowing greater flexibility in accommodating the needs of each project area
- Individual project databases and protocols can be developed at different rates without a significant cost to data integration
- One project database can be modified without affecting the functionality of other project databases.

Databases are recommended to be in Microsoft Access format, SQL Server, or other established industry-standard relational database management system (RDBMS). Databases

produced under contract or cooperative agreement should also follow these recommendations.

Also recommended is the development of an integrated data dictionary that will allow databases across projects and across disciplines to share common attributes whenever possible. Likely lists of standard attributes and values to be developed include: park unit codes, place names, watershed codes, species identifiers, and vegetation or land type classifications. To the extent possible, an I&M network can work with the Washington Support Office (WASO) and other networks to develop and use shared look-up tables for data consistency and inter-operability at a higher level across programs.

2.5.3 Geographic Information Systems

ESRI GIS applications and products are regularly used by NPS, and GIS is an essential component of an I&M network's data management framework.

For small park units where there are few, if any, resource staff, I&M network staff are encouraged to provide some level of spatial data management support. This may include assistance with documentation, archiving, data acquisition, and data dissemination as requested. In the case of network park units that are larger and have either dedicated GIS staff or collateral duty personnel, I&M networks should work in collaboration with these identified staff in support of spatial data relative to their park units.

While most GIS programs at the park or I&M network level are run independently from one another, there are some common infrastructure solutions that are applicable across jurisdictional boundaries:

- Automated GIS Software Installations and Updates: Because of the increasing numbers of GIS software users, it is most efficient to implement ESRI software installation protocols that streamline installation and updates, including silent updates for existing ArcGIS software users. Initial software installations should be completed through a flexible intranet portal, but future updates conducted through silent and bulk updates.
- NPS Theme Manager: This ArcGIS extension, developed by the Alaska Regional Office, can be deployed and managed to deliver spatial data and metadata to staff. Theme Manager provides a front end for spatial data and makes it easier for staff to locate and access spatial data. The installation and management can be semi-automated, with updates rolled out to users each time they start a new ArcGIS project.

2.5.4 Web Sites

Web sites are a fundamental tool for providing information about the I&M Program and I&M networks. I&M web sites are hosted from a server at the NPS Natural Resource Program Center (NRPC) in Fort Collins. I&M networks are required to follow a common design standard for both their Internet and intranet sites that align with NPS web design guidelines.

I&M network web sites are designed to provide information on inventories, vital signs monitoring, and data management. They also serve products such as the results of I&M vital signs monitoring (e.g., resource briefs, progress reports, trend reports), monitoring plans, data

management plans, other documents and publications, and other items such as newsletters, calendars, and outreach materials. While standardized guidelines and formats provide a common look, feel, and navigation sequence for users, latitude is provided to accommodate individual network requirements or preferences.

The structure and content of web sites must comply with National Park Service standards, FOIA requirements for accessibility of public information, and Section 508 of the Rehabilitation Act which ensures access to electronic information by persons with disabilities. Web development should adhere to the guidance posted on the national I&M intranet site, and to policies and guidance on natural resources web publishing provided in NPS Director's Order #70 (National Park Service 2001).

Sample Network page models for both Internet (publicly-facing) and intranet (NPS only) pages, and general web assistance are provided by WASO I&M support staff, along with security coding and support. The most current web guidance is posted on the I&M Data Management intranet (http://www1.nrintra.nps.gov/im/datamgmt/webdev/).

Internet Sites
The Internet web sites are designed to disseminate material that is relevant to the NPS mission of stewardship, preservation, education and protection of public resources. All network Internet sites and pages are hosted on servers located at WASO in Fort Collins.

In order for I&M Internet sites to be relevant to the widest possible audience, high-level pages (i.e., home page and secondary pages) should provide general and summary content that provides background and context for the I&M Program and network. Links then provide paths to detailed information that may have a more limited or specialized readership.

Network web designers should take into consideration the file size of images, maps, or documents that are posted, and adjust their content to avoid excessive download times.

Intranet Sites
The NPS intranet web (http://www1.nrintra.nps.gov/im/) sites are available only to those with access via either an NPS computer (authenticating by the user's Active Directory login) or via a virtual private network (VPN) login. Intranet sites are intended to house information pertinent to NPS staff and cooperators, including data that may be preliminary, sensitive, or otherwise not appropriate for public distribution. All network intranet sites and pages are hosted on servers located at WASO in Fort Collins.

Extranet Sites
NPS security policies currently limit the availability of extranet capabilities (i.e., the ability for authenticated external cooperators to access internal NPS networks). NRPC continues to investigate possible solutions to these limitations and realizes the wide-reaching benefits of extranets.

Personal Communications
The use of email and telephones to conduct personal communication between individuals at separate locations is ubiquitous. Instant messaging is near real-time and offers additional flexibility for communicating with one or more employees.

Collaboration Technologies
The I&M Program is a widely-dispersed program requiring frequent and effective communication among staff and cooperators. Several NPS enterprise-wide tools are proving to be particularly useful:

IBM Lotus Sametime Meeting

Sametime allows presentations and content on a user's computer to be shared with others within NPS. In combination with conference calling, Sametime can be an effective alternative to meetings requiring travel. *Lotus Notes Sametime Meeting* is currently available to all NPS employees with a NPS computer network login profile.
http://np013denver.nps.gov

Microsoft SharePoint

SharePoint is a web-based collaboration tool that allows multiple users to create, view or contribute to documents, surveys, wikis, discussion forums, and other types of projects that require the participation of many. Unlike web sites, all users with appropriate permissions, can easily post or contribute content.

SharePoint requires setup and management by an IT network administrator, and instances are available through NRPC (http://nrpcsharepoint) and many regional NPS offices. In 2008, NRPC established a SharePoint site for I&M networks, which they may use as needed (http://imnetsharepoint).

2.6 National Applications

The basic tenet of I&M information management is to provide integrated natural resource databases and information systems that provide park staff with relevant, high-quality data for management decisions, resource protection, and interpretation. I&M information needs are broadly separated into two categories:

- *Detailed data and information needed for onsite resource management and protection*. The information used to guide natural resource management decisions must be specific to inform and be 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 National Park Service for use by NPS and DOI managers and central office personnel to answer requests from Congress and for budget, program, and project planning.

In order to provide both categories of information, the NPS Natural Resource Program Center (NRPC) and the I&M Program actively develop and maintain a national-level information management framework and suite of applications that use a combination of desktop applications and Internet-based systems.

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. Other programs such as cultural resources and maintenance have their own distributed applications that often intersect and also may have broader program application and use.

The principal applications and data repositories developed for the I&M Program and service-wide programs, and referenced throughout this document, are summarized below.

2.6.1 NatureBib

NatureBib is the master database for bibliographic references that pertain to park units. It also contains citation data from related databases such as NPSpecies, Dataset Catalog, and the NPS Data Store. NatureBib currently focuses on park natural resource-related references; however, it may eventually be linked to references on cultural resources and other park operations. NatureBib comprises both a master, web-based application and interface as well as a version in MS Access that can be run locally on a user's desktop.
http://www.nature.nps.gov/nrbib

2.6.2 NPSpecies

NPSpecies is the master database for information on species within park units. The database contains data on species that are known or suspected to occur in a park unit, and the associated physical or written evidence for the occurrence of the species (i.e., references, vouchers, or observations). Taxonomy and nomenclature are based on ITIS, the Integrated Taxonomic Information System (http://www.itis.gov/).

The current version of NPSpecies for each park or network can be downloaded from the master web site into a local, MS Access version. The Internet-based version is accessed via password-protected log-ins 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.
http://science.nature.nps.gov/im/apps/npspp/index.htm

2.6.3 Dataset Catalog

Dataset Catalog is an MS Access-based application that provides parks, networks, and cooperators a means to document data set holdings. Although not designed as a comprehensive metadata tool, the Dataset Catalog is used for creating abbreviated metadata about a variety of natural resource data sets. The Dataset Catalog helps parks and networks begin to meet Executive Order 12906 mandating federal agencies to document all data collected after January 1995.
http://science.nature.nps.gov/im/apps/datacat/index.htm

2.6.4 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; however, 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.
http://www.nature.nps.gov/water/infodata.cfm

2.6.5 Natural Resource Database Template

The Natural Resource Database Template (NRDT) is a set of relational database tables that parks and networks can use to develop applications for capturing natural resource inventory and monitoring data. NRDT is a core database structure that can be modified and built upon by users depending on the components and specific protocols required. Accompanying NRDT is the Front-end Application Builder that can be used to build user interfaces, giving developers more time for protocol-specific form and report development.

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 web site.
http://science.nature.nps.gov/im/apps/template/index.cfm

2.6.6 NPS Data Store

The NPS Data Store is a key component of the data dissemination strategy employed by the NPS servicewide. The web-based search tool links data set metadata to a data server. The interface allows customized public or protected searches of natural resource data sets, 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.
http://science.nature.nps.gov/nrdata

2.6.7 Vital Signs Internet Map Services (VSIMS)

VSIMS provides a consistent way to share and analyze vital signs monitoring data in a spatial context. The application allows users to spatially select areas of interest, and to query, browse, and download vital signs monitoring data and spatial layers. VSIMS is currently under development and is available on the intranet only. Planned capabilities include

additional spatial data layers, monitoring data upload and management tools, and additional security functions in a publicly-accessible version.
http://science.nature.nps.gov/nrgis/applications/webmapping/default.aspx

2.6.8 Natural Resource Monitoring Protocol Clearinghouse

This is a web-based clearinghouse of protocols that have been developed by the I&M networks or other established protocols used in national park units. The database provides a summary and download of protocols.
http://science.nature.nps.gov/im/monitor/protocoldb.cfm

2.6.9 Natural Resources Monitoring Partnerships (NRMP)

NRMP is a web-based clearinghouse of monitoring projects and protocols for a diversity of state, provincial, federal, university and organizations across Canada and the U.S. The site is hosted by the National Biological Information Infrastructure (NBII), a broad, collaborative program hosted by the USGS Biological Informatics Office. A number of tools are available to search natural resource protocols or to locate a monitoring program. Information from the Protocol Clearinghouse (above) is periodically provided to NRMP. For further information and to search the biological information database see their web site
http://www.nbii.gov/portal/community/Communities/Toolkit/Inventory_&_Monitoring/Natural_Resources_Monitoring_Partnership

2.6.10 NPS Automated National Catalog System (ANCS+)

ANCS+ is the official curatorial cataloging system of the NPS and is developed under the guidance of the NPS Cultural Resources Program. ANCS+ is run as a stand-alone application at the park level, with annual data transfers to the national level. I&M networks do not have direct access to ANCS+, which is managed by park curators; however, networks do play an important role in contributing data to this system.
http://www.nps.gov/history/museum/

2.6.11 Research Permit and Reporting System (RPRS)

Permits are required for anyone, including I&M networks, conducting on-the-ground research in a park unit. RPRS is a web-based application used by researchers, park staff, and the public for completing and tracking the steps of requesting and issuing research permits, and reporting on research accomplishments via the IAR - Investigator's Annual Reports. IARs are also a valuable source of information on previous research conducted in parks.
https://science1.nature.nps.gov/research/ac/ResearchIndex

2.7 Service-oriented Architecture and the Next Generation of National Applications

In 2006, prompted by the comments and feedback from the users of many of the national applications described above, and also based on the recommendations of an outside consultant, the NRPC Director issued a policy on NRPC Natural Resource IT Applications that established the basis for transitioning these applications to service-oriented architecture (SOA). End-users of these applications clearly indicated the need for simplified access and streamlined workflow in order to effectively catalog and manage the processes and products of natural resource-related efforts. SOA, which is DOI best-practice for application

development, provides an alternative to the existing "silo" architecture, and provides the foundation for a modular application design and extensive information sharing.

2.7.1 Integration of Research Management Applications (IRMA)

The Integration of Resource Management Applications (IRMA) project was begun in October, 2006 in order to implement the 2006 NRPC policy. IRMA is the umbrella project that encompasses the planning and implementation of the SOA framework, and the transition of existing stand-alone applications into an integrated system of data services. Microsoft's .NET has been selected as the data architecture and SQL Server is the adopted database platform. Software development is being implemented exclusively in the C# language. In addition, a blade server infrastructure has been established that will allow for failover, redundancy, more integrated management, and system health monitoring. A new storage area network was also acquired to meet expected significant storage requirements.
http://nrpcsharepoint/irma/default.aspx
http://www1.nrintra.nps.gov/nrpc_soa/

Transition Plans
NPSpecies is the first large application to be converted to SOA. This application is being redesigned for greater simplicity, efficiency, and speed. It is expected to accommodate any number of taxonomic classification systems, and will include a flexible service for managing species observation data at a park unit or sub-unit level. The transition is expected to be complete in late 2008. The NPS Data Store and NatureBib applications are scheduled for conversion after NPSpecies, and will be combined into a single service. This conversion is expected to be completed in early 2009. Additional applications will follow, based upon user requirements and resources available.

Implications for I&M Networks
I&M networks will continue using the existing applications until the phased release of SOA services. Data migration to the new services will be automated to the extent possible; however, I&M network staff will need to be directly engaged in the conversion process.

2.8 References and Resources

Dataset Catalog. Desktop application. (http://science.nature.nps.gov/im/apps/datacat/index.htm). Accessed 16 January 2008.

IBM Corporation. IBM Lotus Sametime. (http://np013denver.nps.gov/). Accessed 16 January 2008.

Integrated Taxonomic Information System. Secure online version. (http://www.itis.gov/). Accessed 16 January 2008.

National Biological Information Infrastructure (NBII). Natural Resources Monitoring Partnership. (http://www.nbii.gov/portal/community/Communities/Toolkit/Inventory_&_Monitoring/Natural_Resources_Monitoring_Partnership/). Accessed 16 January 2008.

National Institute of Standards and Technology. Draft guidance for assessing the security controls in federal information systems. NIST Special Publication 800-53A

(http://csrc.nist.gov/publications/drafts/800-53A/draft-SP800-53A-fpd-sz.pdf). Accessed 23 January 2008.

National Park Service Automated National Catalog System. Secure online version. (http://www.nps.gov/history/museum/). Accessed 16 January 2008.

National Park Service. Under development. Natural resource protection. Director's Order #77. (http://data2.itc.nps.gov/npspolicy/DOrders.cfm)

National Park Service. Geographical Information Systems Program, Vital Signs Internet Mapping Service. (http://science.nature.nps.gov/nrgis/applications/webmapping/default.aspx). Accessed 16 January 2008.

National Park Service. Integration of Resource Management Applications (IRMA). (http://nrpcsharepoint/irma/default.aspx). Accessed 16 January 2008.

National Park Service. Integration of Resource Management Applications (IRMA). (http://www1.nrintra.nps.gov/nrpc_soa/). Accessed 16 January 2008. [Intranet site]

National Park Service. 2001. Internet and intranet publishing. Director's Order #70. (http://www.nps.gov/policy/DOrders/DOrder70.html)

National Park Service, Natural Resource Program Center. 2006. Policy on NRPC natural resource IT applications. (http://www1.nrintra.nps.gov/nrpc_soa/docs/IT_Policy_26Oct2006.doc) [Establishes policy and guidance for the Integration of Resource Management Applications (IRMA) project]

National Park Service. Natural Resource Program Center SharePoint. (http://nrpcsharepoint). Accessed 16 January 2008.

National Park Service. Natural Resource Program Center SharePoint, I&M networks. (http://imnetsharepoint). Accessed 16 January 2008.

National Park Service. Natural Resources Inventory & Monitoring Program. (http://science.nature.nps.gov/im/index.cfm). Accessed 16 January 2008.

National Park Service. Natural Resources Inventory & Monitoring Program, I&M network data management: web authoring. (http://www1.nrintra.nps.gov/im/datamgmt/webdev/index.cfm). Accessed 16 January 2008. [Intranet site providing web authoring information for I&M networks]

National Park Service. Natural Resources Inventory & Monitoring Program, Protocol Database. (http://science.nature.nps.gov/im/monitor/protocoldb.cfm). Accessed 16 January 2008.

National Park Service. Natural Resources Inventory & Monitoring Program's Natural Resource Database Template (NRDT). (http://science.nature.nps.gov/im/apps/template/index.cfm). Accessed 16 January 2008.

National Park Service. [No date] Information technology management. Director's Order #11A. (http://www.nps.gov/policy/DOrders/DO11A-14day.html)

National Park Service. [No date] Access your national parks: a guide to accessibility programs. (http://www.nps.gov/access/).
[web authoring information to ensure access to electronic information by persons with disabilities]

National Park Service. 2005. Office of the Chief Information Officer (OCIO). OCIO policy number: OCIO 2005-0004. (http://classicinside.nps.gov/documents/OCIO%202005-0004.doc). Accessed 23 January 2008.

National Park Service. Research Permit and Reporting System. (https://science1.nature.nps.gov/research/ac/ResearchIndex). Accessed 16 January 2008.

National Park Service. Water Resources Division, information and data. (http://www.nature.nps.gov/water/infodata.cfm). Accessed 16 January 2008.

National Park Service. 2006. Web publishing. Director's Order #11C (draft). (http://www.nps.gov/policy/DOrders/DO-11C.pdf)
[Under review. Will replace and supersede DO#70 when approved.]

National Park Service Data Store. Secure online version. (http://science.nature.nps.gov/nrdata/). Accessed 16 January 2008.

National Park Service, Office of the Chief Information Officer. IT policies, standards, procedures, and guidance. (http://inside.nps.gov/waso/waso.cfm?prg=308&lv=3)
[List of IT-related policies and procedures]

National Park Service Species Database. Secure online version. (http://science.nature.nps.gov/im/apps/npspp/index.cfm). Accessed 16 January 2008.

Natural Resource Bibliography. Secure online version. (https://science1.nature.nps.gov/naturebib/). Accessed 16 January 2008.

U.S. Department of Interior. 2005. Guidelines and standards for desktop applications.(http://www.doi.gov/ocio/erm/software.html)

U.S. Department of Interior. 2005-04 (NPS) Local Administration Policy (305.1.1) (http://inside.nps.gov/waso/custommenu.cfm?lv=3&prg=308&id=2494)

2.9 Related Network-level SOPs

- Management of network Internet and intranet web sites
- Infrastructure organization and maintenance
- File and directory naming conventions
- Backup and restore routines for I&M network data
- Image management
- Managing audio and video data resources
- Management, documentation, and dissemination of spatial data
- SOPs related to network use of NatureBib, NPSpecies, the NPS Data Store, and other national applications that may need network-specific documentation

CHAPTER 3. Project Management and the Data Life Cycle

Data management within an I&M network must coordinate the broad phases of short- and long-term inventory or monitoring projects, as well as the detailed steps of data acquisition and processing. Establishing a structure, sequence, and steps for these phases is needed to ensure that all tasks are completed, responsibilities are clear, resources are available, and overall workflows are synchronized.

3.1 Objectives

The objectives of establishing and maintaining proper project management and a data management life cycle are:

- Ensure that a sequence of steps is applied to all network projects, from project inception through to project conclusion
- Ensure necessary and appropriate participation of staff and cooperators at all stages of project development
- Ensure that all data resulting from projects are consistently processed, reviewed, and finalized.

3.2 Laws and Policies

The use of standard project management practices is becoming integrated into National Park Service information system development. In 2003, the DOI Chief Information Officer issued a directive establishing standard project management certification requirements for persons managing major and non-major IT investments (U.S. Department of Interior 2003). This directive does not currently apply to information systems developed or managed at the I&M network level; however, it indicates a recognition that effective project management techniques are an essential component to project success. Project management certification can be acquired through various sources; NPS recommends that certification be provided by the Project Management Institute (PMI).

3.3 General Standards and Guidelines

There are two main types of projects handled by I&M networks:

- Short-term projects: These include individual park research projects, inventories, or pilot work done in preparation for long-term monitoring or research.

- Long-term projects: These include vital-signs monitoring projects central to the I&M program, and multi-year research projects and monitoring performed by other park programs, agencies and cooperators. Long-term projects often require a higher level of documentation, peer review, and program support.

Both short-term and long-term projects share many work flow characteristics, and both generate data products that must be managed and made available. Long-term projects, however, have an increased need to adhere to and maintain standards and will require the ability to compare data over an extended period of time (decades for long-term monitoring).

Data management is a vital component in the development of high-quality scientific data and must be considered in every stage of project development. Primary responsibility for these activities rests with different individuals according to the different phases of a project (see Chapter Four).

3.4 Stages of Project Development

Project planning can be broken down into a series of stages, with the number of stages dependent upon the methodology used. The Inventory and Monitoring program has traditionally used a five-stage method, which is similar to other institutions such as PMI.

The I&M program's five project stages are: planning and approval; design and testing; implementation; product integration; and evaluation and closure (Figure 3-1). PMI project stages are: initiation; planning; executing; controlling; and closeout. Steps taken by both methodologies correspond closely, and networks can manage their projects using either method. Stages are described below, with the PMI names given in parentheses.

3.4.1 Planning and Approval (Initiation)

Most decisions regarding project scope and objectives are made at this stage. Funding sources, permits and compliance are also addressed at this time. Primary responsibility rests with project leaders and program administrators; however, data managers must be informed in order to anticipate data management needs and timelines. All contracts, agreements, and permits should include standard language that describes the data formats, specifications, metadata requirements, and timelines for project deliverables.

3.4.2 Design and Testing (Planning)

Details regarding how data will be acquired, processed, documented, analyzed, and reported are developed during this stage, as is data discovery. The project leader is responsible for developing and testing project methodology, or for modifying existing methods to meet project objectives. Project leaders and data managers must work together during this phase in order to build and reinforce good data management throughout the project. Data integrity and quality can be assured most easily by beginning collaborative development as soon as possible after project approval.

Key data management details to address in this phase include developing documentation of project databases (e.g., entity relationship diagrams, data dictionaries, business rules, user interface) and formal metadata. Devoting adequate attention to this aspect of a project is the single most important step in assuring the quality, integrity and usability of the resulting data. Although post-hoc changes will inevitably be required, careful design can minimize database rework.

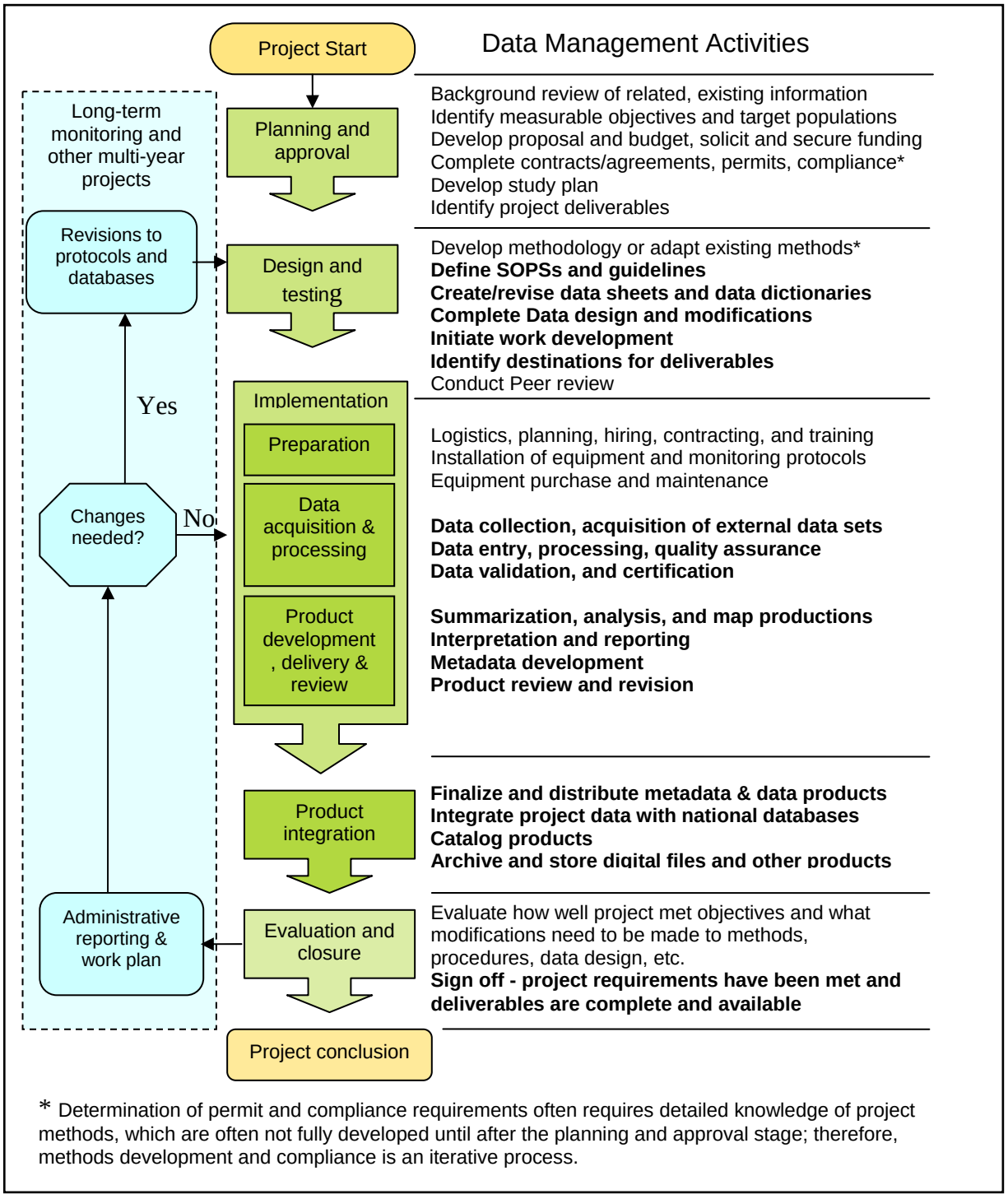

Figure 3-1. Data management steps during the five stages of project development. Key data management activities are in bold type.

3.4.3 Implementation (Executing and Controlling)

During this phase data are acquired, processed, error-checked and further documented, and products such as reports, maps, GIS themes, and other products are developed. The project leader oversees all aspects of implementation, including logistics planning, contracting, training, equipment procurement, data acquisition, documentation, report preparation and final delivery. Data management staff function primarily as facilitators, providing training and support for database applications, GIS, GPS and other data processing applications; facilitation of data summarization, validation and analysis; and assistance with the technical aspects of documentation and product development. The specific roles of data management staff will depend primarily on the technical capabilities of the project staff. As much as possible, these roles should be determined in advance of implementation. Toward the conclusion of this phase, project staff members work to develop and finalize the deliverables that were identified in the project planning documents (e.g., protocol, study plan, contracts, agreements, permits).

3.4.4 Product Integration and Distribution (Executing and Controlling)

Data are certified by the project leader and merged from any working databases to a final database. Metadata are finalized, and products are posted in national repositories or otherwise distributed in the desired formats to the intended audience. This stage may also include upload to I&M national-level databases (e.g., cataloging in NatureBib, updating NPSpecies, posting metadata and associated files to the NPS Data Store). These updates allow the information from the project to be searchable and available to others service-wide. Certain projects may also have additional integration needs, such as when working collaboratively with other agencies.

In general, all raw and derived data products, metadata, reports and other documentation should be delivered to the data steward assigned to the project. Administrative records should be delivered to appropriate park and I&M network staff as specified. All project deliverables should be distributed according to specifications, which should be stipulated in all protocols, contracts, agreements, and permits. Products that do not meet program requirements should be returned for revision.

3.4.5 Evaluation and Closure (Closing)

Records are updated to reflect the status of the project and its associated deliverables in a network project tracking application. 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 (non-cyclic) projects, this phase represents the completion of the project. After products are catalogued and made available, program administrators, project leaders, and data managers should work together to assess how well the project met its objectives, and to determine 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 methods, standard operating procedures, and other procedures are maintained in a tracking table associated with each document. Major revisions may require additional peer review.

3.5 The Data Life Cycle

Data take on different forms during various phases of a project, and are maintained in different places as they are acquired, processed, documented, analyzed, reported, and distributed. This is referred to as the "data life cycle" (Figure 3-2), which is characterized by the series of events listed as follows:

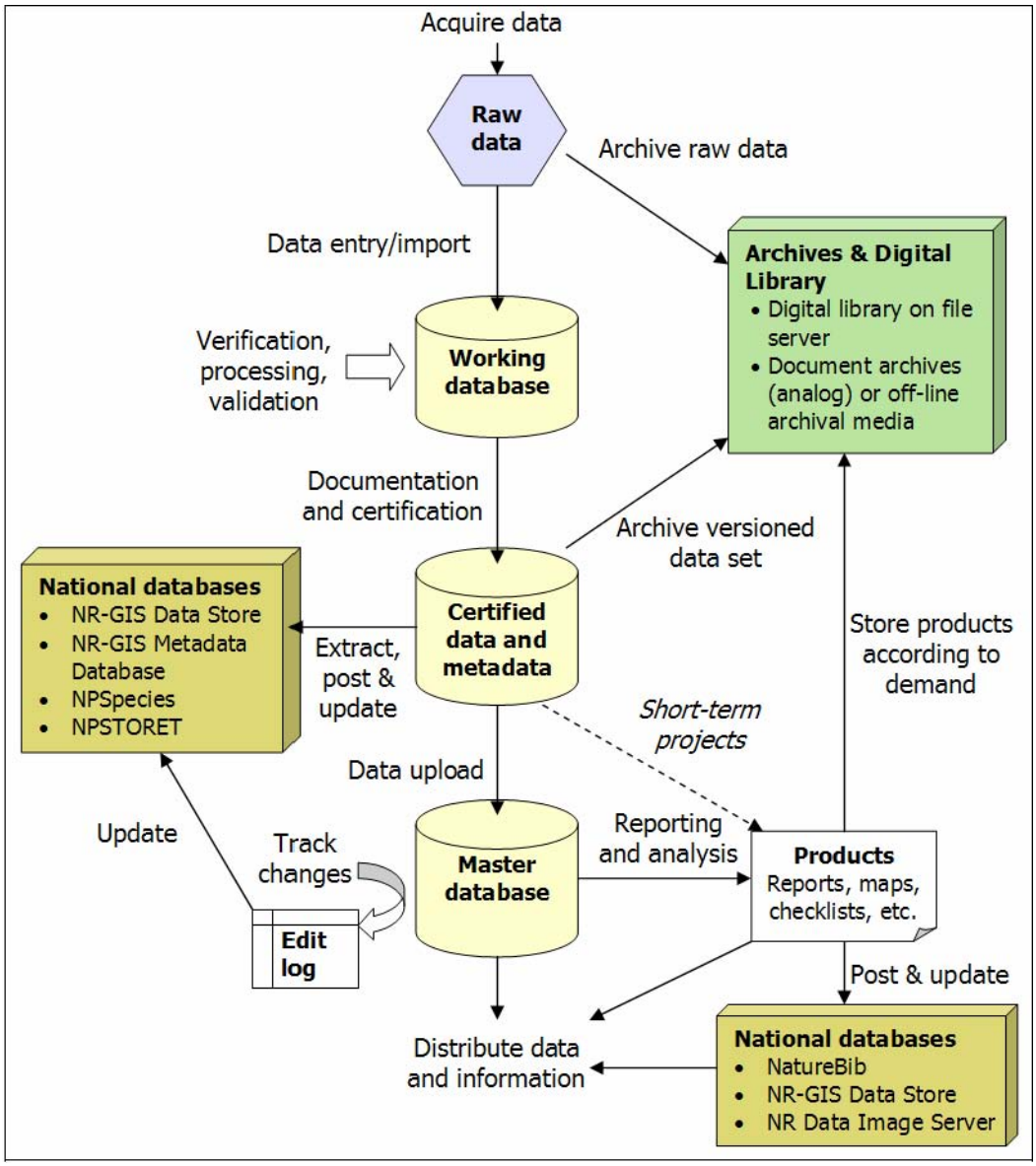

Figure 3-2. The data life cycle.

1. *Acquire data* – Data are acquired in digital or analog form. Digital data can be recorded on handheld computers and PDAs, tablets, data loggers, or laptop computers. Analog data are entered on field data sheets. Acquisition can include both programmatic and non-programmatic data.

2. *Archive raw data* – Copies of all raw data files are archived intact. Digital files are copied to the digital library; hard copy forms are either scanned and placed in the digital library

or are copied and placed in the archives. Note: 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.

3. *Enter/import data* – hand-recorded data are keyed in and digital data files are uploaded to the working database.

4. *Verify, process, and validate* – Verify that raw data have been accurately transcribed. Data are processed to remove missing values and other flaws; and data are validated through visual inspection and queries to capture missing data, out-of-range values, logical errors, and violations of pre-defined business rules (see step 2, above).

5. *Documentation and certification* – Develop or update project metadata and certify the data set. Certification is a confirmation by the project leader that the data have passed all quality assurance requirements and are complete and documented. After certification, all data and metadata are ready to be posted and delivered.

6. *Upload data* – Certified data are uploaded from the working database to the master project database. This step may not be required 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.

7. *Archive versioned data set* – A project data certification form is completed and filed with the data set. Copies of certified data and metadata should be stored in the ntwork's 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, XML).

8. *Disseminate data and update national databases*– Certified data and metadata, and digital image products are posted to national repositories (i.e.,NPS Data Store) to make them more broadly available to others. National databases, including NPSpecies, NPSTORET, and ANCS+ are updated with data obtained from certified data sets.

9. *Reporting and analysis* – Certified data are used to generate information 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. *Distribute information products* – Information products such as reports, maps, and checklists are disseminated to the public through the national and network web sites and cataloged in NatureBib.

11. *Share data and information* – Data, metadata, reports and other information products can be shared in a variety of ways – by FTP or mailing in response to specific requests, or by providing direct access to project records to park staff and cooperators.

12. *Track changes* **–** All subsequent changes to certified data are documented in an edit log, which accompanies project data and metadata upon distribution. Significant edits will trigger reposting of the data and products to national databases and repositories.

13. *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.

14. *Catalog project products* – Catalog products and all information associated with a project, including results of analyses and paths of dissemination. Project tracking databases can be useful tools for this purpose.

For long-term projects, this sequence of events occurs in an iterative fashion, repeating at the end of each field season or other logical data collection and reporting period. Conversely, this 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 needed, and the extent to which product development and delivery is repeated.

3.6 References and Resources

NPS Information Systems Project Management Community (http://inside.nps.gov/waso/custommenu.cfm?lv=3&prg=737&id=3488)

Project Management Institute (http://www.pmi.org/Pages/default.aspx)

U.S. Department of the Interior. 2003. Project management certification requirements for managing information technology investments. Office of the Secretary. (http://inside.nps.gov/waso/custommenu.cfm?lv=3&prg=308&id=2141)

3.7 Related Network-level SOPs

- Checklist for initiating a new project
- I&M network project management
- Project specifications and deliverables
- Guidelines for protocol development
- Data processing and data lifecycle workflow
- Procedures for certifying project data
- Guidelines for project and budget tracking

CHAPTER 4. Data Management Roles and Responsibilities

Data management is about individuals and organizations as much as it is about information technology, database practices, and applications. All staff participating in an I&M network project have a role in data stewardship, and project data sets and products reflect all who have contributed. In order to meet the data management goals and standards developed by the National Park Service and the I&M Program, all staff must understand their associated roles and responsibilities.

4.1 Objectives

The objectives of delineating data management roles and responsibilities are:

- clearly define roles associated with functions
- establish data ownership throughout all phases of a project
- instill data accountability
- ensure that adequate, agreed-upon data quality and metadata metrics are maintained on a continuous basis.

4.2 Laws and Policies

There are no laws or NPS policies governing the establishment of roles and responsibilities with respect to data management. Documenting data-related roles and responsibilities is, however, a best practice: it notifies all persons involved with a project or I&M network of their role in ensuring that data collected, managed, and disseminated are of the highest quality.

4.3 General Standards and Guidelines

Individuals from different disciplines and backgrounds must collaborate to ensure that data are collected using appropriate methods, and that resulting data sets, reports, maps, models, and other information products are credible, representative, and accessible for current and future needs.

The principal roles assumed for the production, analysis, management, and end use of data will vary based on the I&M network and the nature of the project. For a small staff, many activities might be accomplished by a single individual. For large programs, a much higher degree of specialization is typically required. Table 4-1 summarizes general types of data management activities and the roles typically associated with them. Although each position is associated with only one category according to overriding responsibilities, many positions contribute to multiple categories.

Table 4-1. Categories of data stewardship.

Stewardship Activity	Description of Activities	Principal Roles
Production	Creating 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 Crew Leader • Project Crew Member • 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.	• Ecologists • Resource Specialists
Management	Preparing and executing policies, procedures, and activities that keep data and information resources organized, available, useful, compliant, and safe.	• Project Leader • Data Manager • GIS Manager • IT Specialist • Database Manager • National-level Data Managers
End Use	Obtaining and applying available information to develop knowledge that contributes to understanding and managing park resources.	• Park Managers • Superintendents • Network Coordinators • Others

4.4 Specific Data Management Roles and Responsibilities

4.4.1 Data Managers

The I&M network data manager works with national I&M Program data management staff and regional resource information management personnel to maintain a high level of involvement in service-wide and regional databases and data management policy. The network data manager works locally with network personnel, park staff, and cooperators to promote and develop workable standards and procedures that result in the integration and availability of data sets.

Key contacts for the network data manager include park and/or regional GIS and data managers, and the project leaders for each monitoring or inventory project. Consistent and productive communication among these staff leads to common understanding and better synchronization of network and park data management activities. Networks can benefit greatly by a close working relationship with park staff, including personal visits, joint meetings and training sessions, and the meetings and work of the network's Technical Steering Committee and/or Board of Directors. Involvement and input from park scientists and resource information management staff is essential.

Data managers will:

- Develop and maintain the infrastructure for metadata creation, project documentation and project data management
- Create and maintain project databases in accordance with best practices and current program standards
- Provide training in the theory and practice of data management tailored to the needs of project personnel
- Develop ways to improve the accessibility and usability of digital data
- 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.

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

- Define the scope of the project data and create data structures 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 business rule constraints that can be built into the database structure to facilitate quality control, such as required fields, data domains, pick-lists and conditional validation rules
- Create user interfaces that streamline 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
- Ensure that project documentation is complete, complies with metadata requirements, and enhances the interpretability and longevity of the project data
- Ensure regular archiving of project materials
- Inform project staff of changes and advances in data management practices.

The amount of collateral duties a data manager may have to perform varies from network to network depending on the physical location, budget, and overall resources available to a network. While some of these other duties cannot be avoided it should be noted that data managers are not responsible for:

- Park GIS and database solutions outside of the scope of the I&M Program. Data managers are encouraged to assist parks when skills and expertise are required, but should not be solely responsible for implementing projects at the park level
- Park IT network administration and support, including hardware and software issues
- General data entry or keying data from field forms
- I&M Program administrative support

4.4.2 GIS Specialists

GIS specialists manage spatial data themes associated with I&M network inventory and monitoring projects, as well as other spatial data related to natural resource inventories and vital signs monitoring. GIS specialists maintain standards for geographic data and are responsible for sharing, integrating, and disseminating GIS data as needed.

The GIS specialists will work with project leaders to:

- Determine the GIS data and analysis needs for the project
- Develop procedures for field collection of spatial data including the use of GPS and other spatial data collection techniques
- Display, analyze, and create maps from spatial data to meet project objectives
- Properly document data in compliance with spatial metadata standards

GIS specialists will work directly with data managers to:

- Design databases and other applications for the network
- Create relationships between GIS and non-spatial data
- Create database and GIS applications to facilitate the integration and analysis of both spatial and non-spatial data
- Establish and implement procedures to protect sensitive spatial data according to project needs
- Develop and maintain an infrastructure for metadata creation and maintenance
- Ensure that project metadata are created and comply with national and agency standards

4.4.3 Project Leaders

Since the data management aspects of every inventory or monitoring project normally require the expertise and involvement of several persons over a period of months or years, it makes sense that one person is charged with keeping track of the objectives, requirements, and progress for each project. This project leader is usually a resource management specialist or scientist with training, experience, or authoritative knowledge in the field of science related to the project.

The project leader is responsible for data quality during all phases of the project, including collecting, entering, handling, reviewing, summarizing, and reporting data. Developing project documentation and metadata are crucial elements of the project leader's role. The overlap between project and data management responsibilities is illustrated in Figure 4-1.

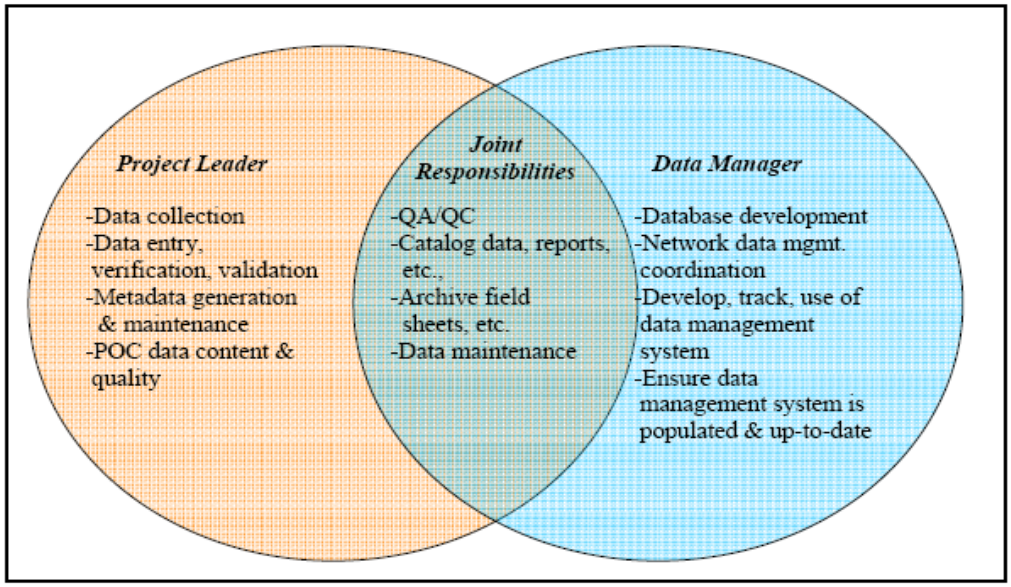

Figure 4-1. The overlap between data management responsibilities of the project leader and data manager.

Specifically, a project leader 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
- 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 documentation for each field data collection period
- 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 analyses of data, and their public availability
- acting as the main point of contact concerning data content

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

- develop quality assurance and quality control procedures specific to project operations
- identify training needs for staff related to data management philosophy, database software use, and quality control procedures
- coordinate changes to the field data forms and the user interface for the project database
- document and maintain master data
- identify sensitive information that requires special consideration prior to distribution
- manage the archival process to ensure regular archiving 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 these transformations
- identify and prioritize legacy data for conversion and convert priority data sets to a modern format
- increase the interpretability and accessibility of existing natural resource information

4.4.4 Shared Responsibilities

Keeping track of data from the time of acquisition until archiving is the shared responsibility of everyone involved with data–whether producer, analyst, manager, or end user. Proper data stewardship is a principle of mutual accountability rather than one job for one individual.

Data stewardship requires that persons involved in a projects activities learn and understand the expectations for continuous information management. This is equally important for I&M network staff, park employees, and contractors or cooperators. All project participants should receive training, briefings, materials, and additional regular communication about data stewardship from supervisors, project leaders, and data managers. Thus, everyone

understands how their efforts relate to park and network management objectives, National Park Service and Department of Interior policies, and other federal government requirements.

I&M project leaders must recognize and understand information management issues and requirements. In particular, well-trained field crews, and an ongoing plan for improving their technical skills (e.g., GPS, data quality assurance/quality control) will significantly improve all aspects of data processing.

Careful documentation of data sets, data sources, and the methodology by which the data were collected or acquired is a key responsibility shared by all project participants. 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. Careful documentation provides the end user with consistent data of known quality and that documentation requires the coordinated effort of all who are involved with a project.

4.5 References and Resources

4.6 Related Network-level SOPs

- Network data management roles and responsibilities
- Project roles and responsibilities

CHAPTER 5. Databases

The quantity and complexity of data acquired as part of I&M network inventory and vital signs monitoring projects requires a well-planned and consistent approach to database design, development, and management. Databases are the cornerstone of an I&M network, and are the primary repositories of data from long-term studies that will need to be accessible and applicable for decades to come. In addition, network databases must accommodate a wide diversity of data types, enforce data integrity, support a variety of queries and output, and help substantiate resource management decisions.

5.1 Objectives

The objectives of effective database design, development, and management within I&M networks are:

- Ensure the long-term stability, security, and accessibility of data
- Encourage common practices, standards, and sharing of database development efforts among networks
- Encourage data exchange and multidisciplinary applications of data
- Ensure databases are designed and built using NPS- and industry-accepted best practices

5.2 Laws and Policies

There are no NPS-wide standards on database design or management. The I&M Program has established the Natural Resource Database Template as the preferred structure for databases designed for vital signs monitoring data.

5.3 General Standards and Guidelines

The I&M Program has developed a series of recommendations related to database design that include:

- Natural Resource Database Template (NRDT) data structure and standards (http://science.nature.nps.gov/im/apps/template/index.cfm)
- Guidelines for long-term monitoring protocols (Oakley et al. 2003)

In addition, I&M networks continue to develop and refine database-related standards based on practical experience. Many of these standards are posted on the I&M Data Management intranet site (http://www1.nrintra.nps.gov/im/datamgmt/)

5.3.1 Natural Resource Database Template

The Natural Resource Database Template (NRDT) has been developed by the Inventory and Monitoring Program as its database standard (National Park Service 2007). NRDT is a core database structure that can be modified to accommodate specific data needs and sampling protocols. NRDT:

- Provides both a data interchange standard and a standard MS Access database core that allows flexibility in application design
- Serves as a starting point for application development that can be extended as necessary to accommodate any inventory or monitoring field sampling protocol

- Standardizes location and observation data to facilitate the integration of data sets
- Acts as a design platform for developing database applications in MS Access allowing users to enter, edit, display, summarize, and generate reports for inventory or monitoring data sets
- Integrates with other I&M data management systems and data standards including Geographic Information System (GIS) tools and data, and the NPS Metadata Profile
- Is accompanied by an MS Access Front-end Application Builder that can be used to create a user-interface. This includes many built-in utilities such as table linking, backups, compaction, and lookup table management.

Monitoring protocols that include a NRDT-based application should be posted on the web-based monitoring protocol clearinghouse. Users can download a written monitoring protocol from the clearinghouse, and contact the network data managear to obtain the corresponding database objects such as tables, queries, forms and reports. The fact that a similar NRDT structure is used by many simplifies the sharing and exchange of work.

5.3.2 Project Database Standards

In addition to the NRDT structure described above, project database standards are necessary for ensuring compatibility among I&M data sets, which provides the groundwork for data sets to be aggregated or summarized. Well-thought-out standards also encourage sound database design. The following considerations should be addressed when designing project databases:

- Standardized database object naming: objects such as tables, fields, queries, forms, and reports should be named according to conventions adopted in an SOP
- Links to central databases: links to data tables that reside in central databases prevent storing redundant information in multiple project databases. These tables typically contain information that is complex to maintain and is applicable to many projects. Examples include: master version tables for vital signs projects; network-wide species and vegetation data; data user logs; and data edit logs.
- Standardized look-up or enumeration tables: many data tables are used repeatedly in project databases. Standardized versions of these tables should be stored at a central location so they can be easily copied or linked to project databases.
- Project-specific fields and tables: the remainder of database objects can be considered project-specific, although there will typically be some overlap among projects. I&M networks are encouraged to work with other networks performing similar projects in order to identify any standards or commonalities that might be shared.

Levels of Database Standards

Links to central databases and the use of common lookup and/or enumeration tables represent the first level of data standards because they are consistent among data sets. A second level of standards is established by the core template fields and tables, which are standardized where possible (however, project-specific objectives could lead to variation among projects). The third level of standards is applied most flexibly to accommodate the range of needs and possibilities for each project, yet always with compatibility and integrity

in mind. Figure 5-1 illustrates these levels, with central data providing the foundation upon which levels two and three are built.

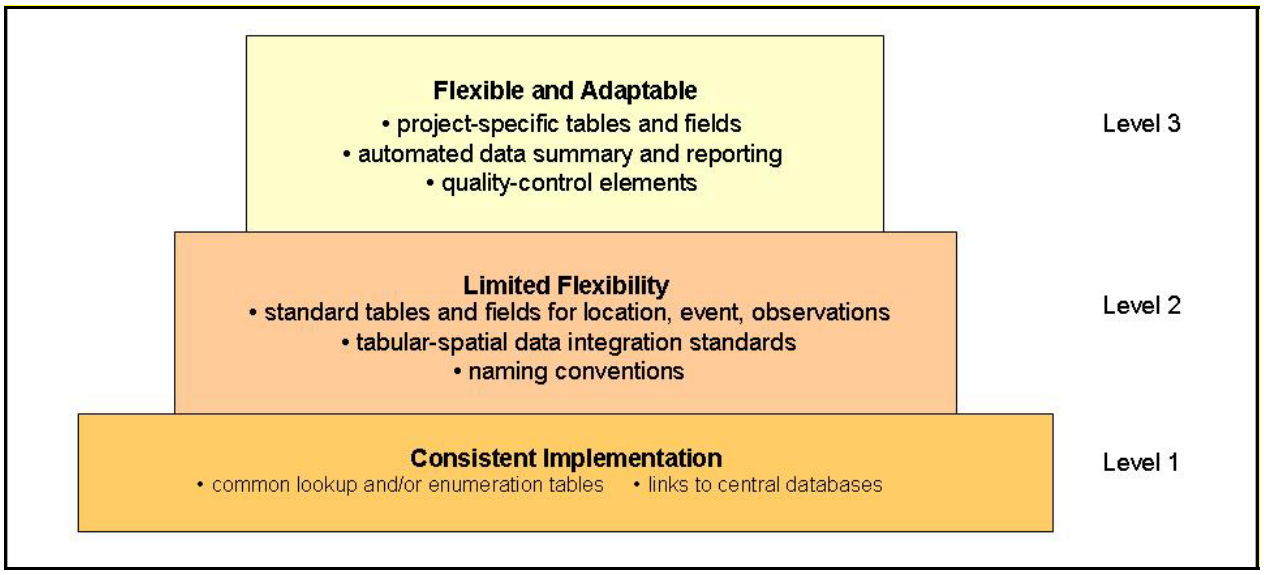

Figure 5-1. Different levels of data standards and their corresponding degree of variability.

Examples of standard attributes and values to be developed include: park unit codes, place names, watershed codes, species identifiers, and vegetation or land type classifications. To the extent possible, networks should collaborate among themselves to develop and use shared look-up tables for data consistency and potential multi-network data sharing.

5.4 Database Planning and Workflow

Successful databases are the result of the thoughtful use of technology to successfully meet user requirements. These requirements should be defined precisely before beginning any application development. Database planning and development should be an iterative process among data management staff and users, with interim designs and products available for review, comment, and modification as needed.

5.4.1 Application Development Process

Figure 5-2 presents a sequence of stages in the application development process.

Products developed early in the project life cycle (e.g., study plans, sampling protocols and standard operating procedures) are essential background for database developers. Stages one through three are accomplished primarily by the project principal investigators. The project is proposed, approved,and protocols and standard operating procedures are developed. These materials will be the foundation for database requirements in later stages.

Stage Four consists of an initial data management consultation where background materials that will define the application are assembled and organized. Monitoring protocols are critical since they contain much of the intellectual work that define what information will be collected, who will collect it, and what it will be used for. At the initial consultation the investigator should be prepared to provide:

- Project Narrative and draft SOPs
- List of stakeholders, end users, and their requirements
- Field forms (if using)
- Data collection flow diagram
- Examples (even if hand drawn) of how the information will be analyzed and reported
- List of software or applications that need to be compatible, such as statistical software or export/import formats for other agencies
- Sample data for testing
- Estimation of number of records per year.

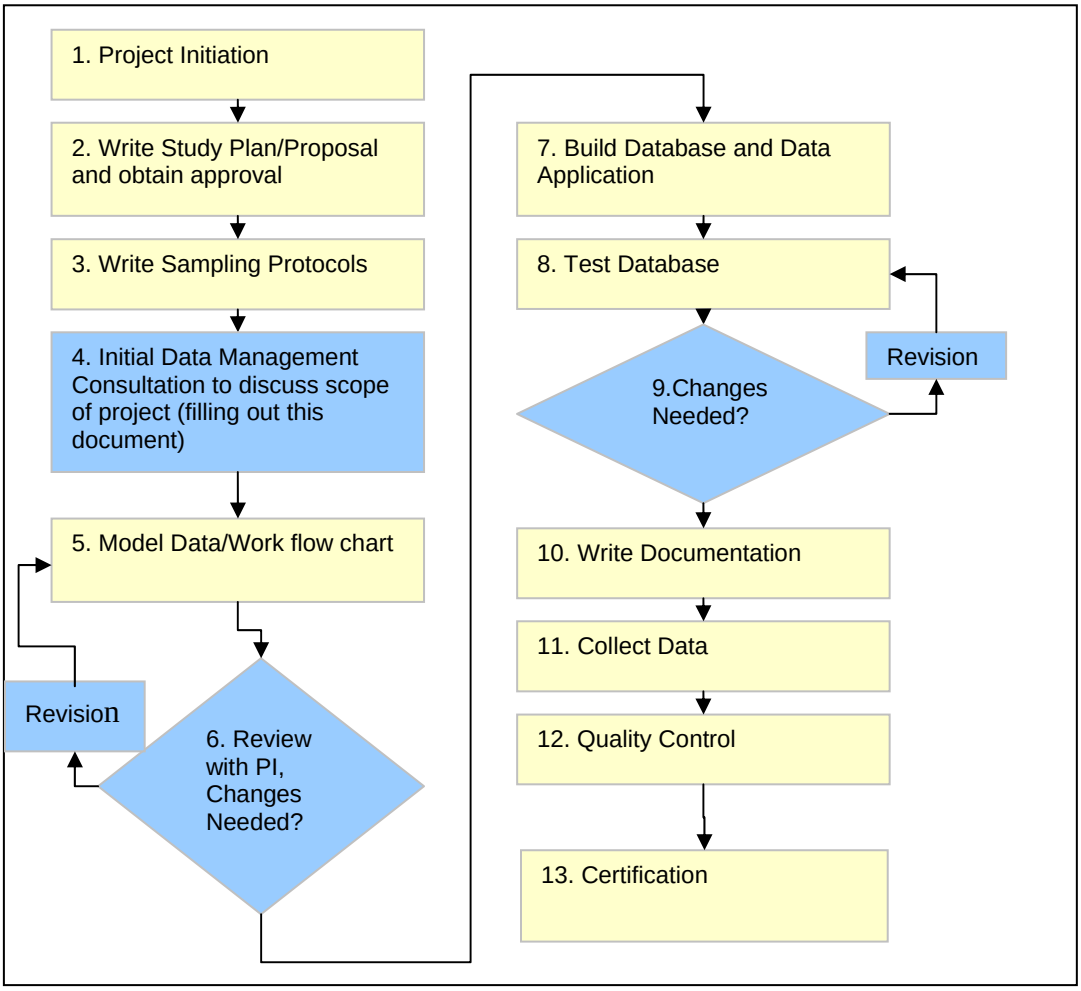

Figure 5-2. Monitoring database application development process. Stages shown in blue represent tasks that require specific input from scientists to guide the database development process.

A database planning checklist is helpful at this stage. Such a checklist should include:

- A list of stakeholders and their requirements
- A work and data collection flow chart
- Specifications for

    - The life span of the project
    - The location where data will be collected
    - Time of data collection
    - Manner of data collection
    - Field computing requirements and risks
    - Database versioning requirements
    - Data type (tabular, spatial, etc.)
- Preliminary data model (what entities are of interest; what are their attributes?)
- Integration with other software, hardware
- Integration with other stakeholder agencies or organizations
- GPS/GIS requirements
- Milestones for tracking project progress

Stages five and six involve modeling the project's data. This must be an iterative process among data managers, database developers and project leaders. In these stages, the model is developed, reviewed and revised until it meets all requirements. It is least expensive and most efficient to resolve any issues with the data application at this stage, rather than after the database is in development.

Database and application development can proceed once the data model is mature. Stages seven through nine represent another iterative cycle where stakeholders must participate in the application's tuning and refinement. Documentation (Stage 10) should be fairly well defined from the data model and supporting documentation. Instructions on the use of the database should be included in the project's data management protocol as well as a help file and user guide. This step is particularly important for monitoring databases where it is assumed that many persons unfamiliar with the database will be using it during each sampling season.

Stages 11 through 13 represent database implementation, at which point it undergoes an established series of quality assurance and certification steps.

While databases will reach a certain level of stability, changes and updates will be ongoing as field methods are revised, and technology and platforms evolve.

5.5 Data Modeling

The majority of the work involved in building databases occurs long before using any database software. Successful database planning takes the form of a thorough user requirements analysis, followed by data modeling.

Understanding user requirements is the first planning step. Databases must be designed to meet user needs, ranging from data acquisition through data entry, reporting, and long-term analysis. Data modeling is the methodology that identifies the path to meet user requirements. Data models may be as simple as a written document or drawing, or may be complex and constructed with the aid of software engineering tools.

5.5.1 Model Development

Four elements are required to develop the data models:

Project staff need to participate in the model development. Discussions may begin among a few, with later 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.

Protocols for the vital sign will provide the greatest substance to the models. The protocols provide the goals, objectives, sampling designs,methods, standards, analyses, and reporting requirements. Development of these documents provides an opportunity for project leaders and data managers to identify suitable data model requirements. The outcome should be a clear, precise description of the framework for a functional monitoring database.

Reference materials such as field forms, drawings, mock-up reports, and classification schemes to be used will play a significant role in the data models.

Frequent interactions among the project leader, 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 modeling should be iterative and interactive. The following broad questions are a good starting point:

- What are the database objectives?
- How will the database assist in meeting those objectives?
- Who are the stakeholders in the database? Who has a vested interest in its success?
- Who will use the database and what tasks do those individuals need the database to accomplish?
- What information will the database hold?
- What are the most atomic bits of information the database will hold and what are their characteristics?
- Will the database need to interact with other databases and applications? What accommodations will be needed?

5.5.2 Conceptual Data Models

Conceptual data models are constructed to graphically portray the processes specifically related to the implementation phase of a project – especially those that involve data acquisition, processing, quality assurance/quality control, and data reduction. 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, or kept as alternatives to LDMs, with less complex projects.

Conceptual data models should contain the following:

- 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.
- A flow diagram of procedures, what information is needed and when, and what information is being collected or produced and when
- Descriptions or mock-up illustrations of how the data should be presented.

5.5.3 Logical Data Models

A logical data model is an abstract representation of a set of data entities and their relationship, usually including their key attributes. 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 four main components:

Data entities are 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

Entity attributes are 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%)

Logical relationships illustrate 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

Structural hierarchies demonstrate the structure and order of relationships among data entities, which can be determined once the logical relationships are known:

- Site locations
- Sampling event
- Species observations
- Water samples, temperature, pH, dissolved oxygen
- 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

5.5.4 Physical Data Models

The physical data model is used to design the actual database, depicting data tables, fields and definitions, and relationships between tables. 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.

5.6 Other Database Considerations

5.6.1 Data Normalization

Normalization is the process of designing database tables to ensure that data elements in one table are not repeated in other tables. This ensures that data are stored in one location only, enforces data integrity, and guards against logical inconsistency in the data. There are six levels of normalization (normal forms); however, design beyond the third normal form would likely not be required.

5.6.2 The Triple Constraint (Scope, Time, Resources)

Building an effective database is a complex process that requires a considerable commitment of resources. Database development often needs to balance competing demands and trade-offs for quality, cost (resources), and time. In the first illustration in Figure 5-3, increasing or decreasing one point on the triangle will affect the other two points. Another perspective on this is in the second illustration: attaining two points on the triangle usually means sacrificing the remaining point. The concept of the triple constraint can be a helpful illustration when working with project staff, and can help clarify priorities and limitations.

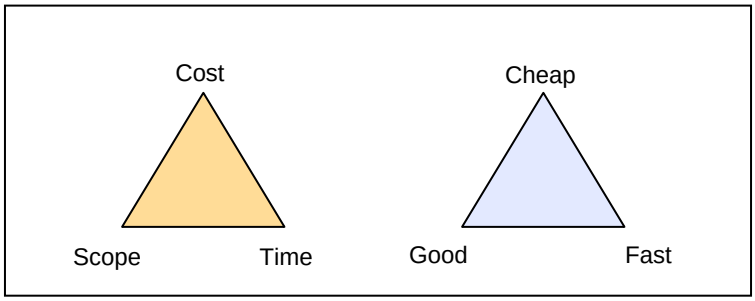

Figure 5-3. The triple constraint triangles

5.6.3 Build, Buy or Adapt

For some data management functions within an I&M network there may be options available such as building a new database, buying commercial off-the-shelf software (COTS), or adapting a database or system that has been built elsewhere. For most vital-signs monitoring projects, data managers will either be using an NRDT core as a starting point, or will be adapting a similar database built by another I&M network. Data managers should investigate options for COTS for specific functions such as photo management, database modeling, or project management.

5.6.4 Database Software

All databases have a software mechanism for storing structured data and a language for querying and manipulating the underlying information. While MS Access has been the primary database application software used by I&M networks, it is proving to be limiting for networks needing a larger central data repository or with many distributed concurrent users. Client-server or n-tier database architectures are better options to securely store and serve data, can more easily handle concurrent use, transactions, triggers, and database permissions and constraints. Microsoft SQL Server Express Edition is an option that some networks are choosing as an alternative to MS Access. Other options would be Microsoft SQL Server Workgroup or Standard editions. Most NRPC applications, including components of the IRMA project, are moving towards SQL Server as the primary database platform (as opposed to the competing Oracle products), which will increase the ability of NRPC staff to provide technical assistance to networks.

5.7 References and Resources

National Park Service. Natural Resources Inventory & Monitoring Program, data management intranet. (http://www1.nrintra.nps.gov/im/datamgmt/). Accessed 16 January 2008.

National Park Service. 2007. Natural Resources Inventory & Monitoring Program, natural resource database template [documentation and MS Access database]. (http://science.nature.nps.gov/im/apps/template/index.cfm)

Oakley, K., L. Thomas, S. Fancy. 2003. Guidelines for long-term monitoring protocols. Wildlife Society Bulletin 31(4): 1000-1003.

5.8 Related Network-specific SOPs

- Checklist for initiating a new project
- Database application specifications
- Database object naming conventions
- Database documentation and data dictionary guidance

CHAPTER 6. Data Acquisition, Processing, and Reporting

Interdisciplinary or multi-scale natural resource programs and projects increasingly rely on data gathered from multiple sources. This chapter describes the general steps involved with acquiring, processing, and reporting data to meet standards established by the NPS I&M Program, although these steps could apply to other NPS park units or programs. Data acquisition also encompasses physical objects (e.g., photographs, voucher specimens), which are often collected as part of resource management, inventory and monitoring, and other research projects.

6.1 Objectives

Acquiring, processing, and reporting on natural resource-related data are core responsibilities of I&M networks. To effectively meet these responsibilities networks must:

- Identify, evaluate, and acquire, through a variety of sources, data needed by networks to properly carry out inventory or vital signs monitoring projects within parks
- Determine the most reliable and efficient means or tools for collecting data
- Establish routines and procedures for data processing
- Identify and meet the information needs of a broad range of end users that includes scientists, park managers, park interpreters, and the general public.

6.2 Laws and Policies

Current laws and policies encourage federal programs to make use of existing spatial data before investing in the acquisition of new data sets. Specifically, Executive Order 12906, Section 3(d) states that each agency must adopt internal procedures to ensure the agency accesses the National Geospatial Data Clearinghouse before it expends federal funds to collect or produce new spatial data, to determine whether the information has already been collected by others, or whether cooperative efforts to obtain the data are possible.

As part of any new project, the networks will use the National Geospatial Data Clearinghouse (geodata.gov) and NPS Data Store (science.nature.nps.gov/nrdata) for searching existing natural resources data set. The networks will also use other Internet resources, agency contacts, and other means to find data that may benefit its programs.

6.3 General Standards and Guidelines

General standards and guidelines for the acquisition and processing of programmatic and non-programmatic data (as defined in Chapter One) are provided below. Quality assurance and control procedures, which are essential for ensuring the production of superior data and information products, and which are components of all stages of the data life cycle, are discussed in greater detail in Chapter Seven.

6.4 Programmatic Data

Project leaders and data managers are responsible for ensuring data collection, data entry, verification, validation, storage, backup, and archiving are consistent with I&M network standards. In addition to general standard operating procedures that define network-wide requirements, protocol-specific SOPs may be developed that detail procedures or

methodologies. Project leaders and cooperators on I&M projects should always receive a copy of their network's relevant project specifications and protocols.

6.4.1 Data Collection

A wide range of tools is available for field data collection. All methods involve some trade-offs in terms of expense, efficiency, and susceptibility to data entry/transcription errors. Project protocols should provide detailed specifications on how the following tools are used with individual projects. Justification for the choice of methods should also be provided.

Field Forms are the most common method of recording field data. Field forms are inexpensive but require neat, legible handwriting. Manual field forms pose greater opportunities for error during the collection/data entry process compared with other methods: data noted on forms must be keyed later into project databases, sometimes by persons other than the original data recorders. This method tends to require more data entry time and more rigorous QA/QC.

Field Computers increase data collection and data entry efficiency. Data can be directly downloaded to office desktops, eliminating the data entry step. QA/QC checks can be implemented in the database, further reducing data entry error and processing time. The drawback to field computers is usually greater expense, weight, battery life, complexity, and the risk of data loss if backups are not performed consistently. Two types of computers are frequently used in the field:

Personal Data Assistants (PDAs)**:** the small size and relative low cost of these devices make them attractive options for collecting field data. PDAs can be weatherized fairly easily and inexpensively. However, they have limited internal memory and data must be continually backed-up to larger storage devices. Most require additional processing/programming to transfer or create the structure of the master database in the field units. Working with relational database structures can be difficult to impossible if the database is complex, and uploading field data to a larger device requires specialized skills or software.

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, heavier, more expensive, and harder to weatherize than the PDAs, but more powerful as well. They are the best choice for field projects that are very data-intensive or rely on complex data structures. MS Access databases, as well as spatial data layers, can be directly transferred from the desktop to field units without additional programming. Data from the field units can in turn, be uploaded to the desktop with minimal effort.

In some cases, the best choice is a combination of field forms and computers, for example, where large amounts of notes or comments need to be recorded in the field.

Automated Data Loggers are mainly used to collect ambient information such as weather data or water quality information. These units must be properly calibrated and maintained, which requires proper training of field crews and SOPs that outline these procedures. Two types of automated data loggers are available:

Permanently deployed devices provide continuous, real-time, or frequent data and are useful during conditions unsafe for field staff (e.g., severe storms, flood events). Data still must be retrieved; however, this can be accomplished remotely via satellite, landline, or cell phone. Providing sufficient and reliable power may be a challenge. Automated data-loggers require proximity to AC power, installation of solar power, or regular site visits to change batteries. Wilderness regulations may restrict the type of and location where a device is deployed.

Portable hand-held devices are deployed for sampling only during site visits.

GPS Units are generally used during field work to collect location (geographic coordinates, altitude) information. Two types of GPS units are most frequently used:

Recreation Grade Units include both self-contained units (e.g., Garmin Map76CSx) and companion units for PDAs (e.g., Holux). These work well for collecting general position information but are not recommended for high-accuracy (sub-meter) location information.

Survey Grade Units such as many Trimble GPS receivers are used for collecting accurate sub-meter location information. Increased accuracy comes at a greater cost.

Tape Recorders such as 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.

Photographs provide an excellent visual record of field visits and are useful for capturing 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. Slides and photographs should be stored and archived according to guidelines, and must have essential metadata associated with each image in order for the information to have long-term value.

Remotely Sensed Data includes satellite imagery and aerial photography. Remote sensing 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 include:

- Accuracy and resolution needed
- Frequency of measurement
- Costs
- Licensing for public use
- Ortho-rectification standards

Each remote sensing product is unique. It is imperative that users fully understand the product being used and that products be accompanied by well-documented metadata. Any projects involving remote sensing should include consultation with a professional remote-sensing specialist. Project managers should consider the trade-offs between accuracy and cost among different imagery sources.

A network's GIS library can include satellite imagery coverage for most of the parks, along with base cartographic layers, digital orthophoto quadrangles, and natural resource theme layers.

Any additional remote sensing data acquired will likely require some spatial or spectral processing depending on how they are received. All such data sets should be received in a geo-referenced format although some may need to be transformed to the network standard (UTM zone, NAD 83, meters). The extent of any additional processing will depend on how the data are to be used and what type of information is to be extracted from them. Specific requirements and procedures should be outlined in project protocols or SOPs.

6.4.2 Data Processing

Ideally, each project will have a database developed prior to the collection of any data in the field. I&M network databases are developed in conjunction with project protocols and are generally based on the Natural Resources Database Template (Chapter Five). Built-in procedures for data validation and QA/QC should be part of the database development and user interface. Data processing should encompass the following steps, and should occur as soon as possible after data are collected:

- Field crews enter all data into an approved project database, under the supervision of the project manager.
- Field crews periodically forward project data files to the project leader and/or data manager (refer to individual protocols for specific requirements).
- All data undergo QA/QC procedures (see Chapter Seven for more specifics on data verification and validation).
- The data manager maintains the master copy of the database and updates it with certified data files received from the project leader.
- National databases are updated as per the procedures outlined in Chapter Ten.

Spatial Data

See network-specific standard for data collection with GPS and spatial data standards for complete details on methods for data collection and processing of spatial data.

Photographs

Photos taken as part of a project's data collection protocol constitute data and need to be organized, documented and preserved in conjunction with all other project data. The processing, documentation, and storage of photographs, including digital photos, are detailed in network-specific SOPs. In general, the level of processing required will depend on the purpose of the photograph. Photos are regularly used by many park programs such as resource management, maintenance, fire management, and research.

Editing of digital photos may be done to improve orientation or correct for lighting conditions (e.g., rotation to best orientation and removal of 'red eye'), but should never include alterations that change the original content of the photo. Photos may be cropped to remove edge areas that grossly distract from the subject. Poor quality photos can be deleted or destroyed, except where the subject is unique. Photos of medium quality should be assessed against existing photos of the same subject; if they duplicate a subject with no enhancement of quality or perspective, they may be deleted or destroyed. Working photos should be stored either in a workspace within a specific project or within a user's photo library. Digital aerial photos and imagery should be geo-referenced. All digital data photos should have geographic coordinates associated with them.

Remote Sensing Data

The processing of remotely-sensed data is often project specific. Therefore, project plans, protocols, and SOPs should document these procedures. Some steps will be common to all data sets, including:

- Images are geometrically registered using nearest-neighbor resampling methods and co-registered to UTM NAD83. Registration accuracy is assessed
- Images are radiometrically corrected and converted to exo-atmospheric reflectance
- Atmospheric corrections are made, if applicable
- All iterations (raw, intermediate corrections, and final) are maintained until project completion when raw and final products are archived.

Vouchers

The acquisition and processing of biological specimen vouchers will be guided by the policies outlined in network-specific SOPs, and under the guidance of the museum manager, curator, or permits manager for the park in which the specimens are collected.

Project leaders will provide parks with material and data, in a prescribed MS Excel format or comma-delimited ASCII files (.csv) for automated uploading into ANCS+ and input to NPSpecies. Data provided to non-NPS curators will be in MS Excel format as well. Specimens are owned by the parks in which they were collected and are curated by park staff unless agreements have been made with outside institutions. See Chapter Eleven for more details.

6.5 Non-programmatic Data

A large amount of the data used to manage the natural resources of I&M network parks is collected by entities outside of the network (e.g., universities and other parks, agencies, and NPS programs). Locating and cataloging these kinds of data is referred to as data mining. Data collected and products produced by such efforts provide a great deal of information about park natural resources and are therefore relevant to many of its programs. These data can be classified as follows:

- *Current or ongoing* – These are data that are currently being acquired or were recently acquired, are recognized to be of value, and are targeted for incorporation into existing or planned network projects. Their acquisition and use should follow very specific guidelines established in project or network protocols. Sources can be internal to NPS or

from external sources. These data might be used to complement program data collection, to fill in gaps of missing data, or as a basis for comparison.

- *Legacy* - These data are found and compiled through the data mining process. They may include vertebrate and vascular plant species data, other important natural resource inventory data, specimen or voucher data, bibliographic data, and existing monitoring data sets.

Non-programmatic data can be obtained from academia, private organizations and non-profit groups, as well as local, state, and federal government agencies. Within the NPS, park, regional and national programs are all potential sources of information. Examples include:

- *NPS natural resource inventory programs* – National-level programs collect data, maintain databases, assure data quality, and perform the trend analyses relevant to a suite of 12 natural resource topics, including vegetation, geology, soils, and air quality.
- *EPMT* - Exotic Plant Management Teams (EPMT) collect and maintain data regarding the presence of exotic species in many parks, and develop and document the methods used to treat these species. Data are stored in the Alien Plant Control and Monitoring Database (APCAM) and maintained by the EPMT data manager.
- *Fire Program* **–** Data on fire occurrence within the networks are maintained both at individual park and regional levels. National databases such as Fire-Pro, SACS and Fire Program Analysis (FPA) have been, and will be, used to maintain information regarding fire incidence and the resources dedicated to fire management.
- *GIS* – The networks are supported by regional GIS specialists to help ensure regional GIS data are available and accurate. Much of these data are also available through the NPS GIS Clearinghouse.
- *Biological Resource Management* **-** The regional Threatened and Endangered (T&E) Species Specialist provides management support and coordinates reporting of T&E species populations.
- *Research Permit and Reporting System* – search tools enable users to search and view summaries of park-based research.
- *ANCS+* - Documents, photographs, specimens, and other research products are frequently archived at the park or regional level.

Information collected during the data mining process should be maintained either electronically or as hard copy format depending on its original form, and should be cataloged and documented as fully as possible (Chapter Eight). Data should be cataloged to the appropriate repository as summarized in Table 6.3 and discussed in Chapter Ten.

Table 6-1. Summary of possible data sources for different types of information and repositories where they are maintained

Type of Data	Possible Source	Repository
Bibliographic / Literature	• Online literature databases (e.g., First Search or Biosis) • Library catalogs (e.g., academic or research institutions) • Park or regional archives through ANCS+ • Network parks: resource libraries, offices, file cabinets	• NatureBib • Reference cabinets for hard copy materials • Digital archive for electronic materials
Geographic Data	• Regional centralized GIS data • Federal and state geographic data clearinghouses • Local, state, and federal government offices • Regional and park GIS specialists	• NPS Data Store • Digital archive
Biologic / Natural Resource Data	• Voucher collections (museums, parks, universities) • Network parks: local or network computers	• NPSpecies • Dataset Catalog • Digital archive

Data mining is an integral part of project development, but efforts should not be limited solely to project development needs. This should be an ongoing process requiring regular data searches and visits to network parks to ensure that parks maintain and catalog as much material relevant to managing their natural resources as possible. Encouraging data sharing among parks will assist in this process and may alleviate the need for regular searches of park records.

6.5.1 Data Collection

The collection of data from non-programmatic sources should follow program specifications as outlined by project protocols or SOPs. Procedures should be standardized as much as possible and include:

- Contacting data stewards and informing them of program needs
- Establishing Memoranda of Understanding (MOUs) if needed, to address data ownership, dissemination, sensitivity, citation, or other use issues
- Developing a contingency plan in case the data source is no longer available
- Determining whether data can be consistently exported/imported and establishing a schedule
- Determining how the data will be stored and integrated into the program
- Determining how errors will be addressed
- Determining if documentation is adequate and if not, completing where necessary

- Identifying any interest in the exchange of program data and information with the outside program.

Agency or organizational stewards of these data often have the expertise to conduct proper quality assurance and the capability to function as a repository and clearinghouse for validated data. In some cases, portions of external databases may be incorporated into network databases and thereby made more accessible to staff.

6.5.2 Data Processing

Much of the data identified during the mining process is likely to be the legacy type. As time and resources permit, legacy data meeting quality and relevance standards should be converted to file formats compatible with current software standards. Hardcopy references and other materials containing legacy data can be scanned and saved as .pdf files, and uploaded to NatureBib or the NPS Data Store where they become available to many. All legacy data sets should be reviewed and cataloged as follows:

- Enter species-related data into NPSpecies (this is especially important for park-based biological inventories).
- Enter all natural resource reports and publications related to network parks into NatureBib. Hard copies should be stored in the appropriate park or network collections and electronic copies uploaded to NatureBib.
- All GIS data should be stored in the proper projection (UTM NAD83) and accompanied by FGDC-compliant metadata.

6.6 Data Analysis and Reporting

Data summary and analysis is essential to providing relevant and useful information for natural resource managers and scientists. Data managers and stewards must provide valid data in formats that support scheduled and ad hoc display, query, analysis, summary, and reporting. Routine and scheduled data summary, analysis and reporting requirements and procedures should be identified in project protocols.

The development of data products should be guided by project objectives, protocols and data management SOPs. Products defined in advance, such as routine summaries and output formats can be automated through the use of queries and reports stored in a project's database. Project leaders should work with data managers to create, and test these queries. Queries can be used to view raw or summarized data, or to output either type as custom-formatted reports or as files that can be imported by other analytical tools (e.g., statistics software). Other queries can be built to facilitate data exploration and unscheduled analysis. Specific needs should be determined by individual projects.

Most likely, data will be exported from project databases for most statistical analyses beyond means, standard deviations, and other descriptive statistics. Networks will typically use third-party statistical software (e.g., R, SAS, SPSS) for generating frequency distribution plots, tests for normality analysis of variance, time series analysis, and others.

6.7 Changes to Procedures

Changes to established vital signs data collection or processing procedures are discouraged unless there are acceptable, valid reasons for altering the methodologies. Ideally, problems are identified during the design and testing stages of the project and changes implemented prior to the collection of any field data. Protocols should attempt to identify any foreseeable issues that might occur as well as contingencies to address them. Once procedures are in production, changes are usually costly both in terms of data management resources required and in terms of potential erosion of existing data quality.

Inevitably, unforeseen problems may occur that require procedure or protocol revision after data collection has begun. Significant changes to protocols must be approved by the principal investigator, project leader and data manager. The key official should evaluate the proposed changes and determine if additional peer review is required before accepting them. All changes should be carefully documented within project SOPs and any associated databases or metadata files.

Any changes that occur as a result of a change in a project's protocol should be documented in the formal documentation developed for the database (Chapter Five).

6.8 References and Resources

Executive Order 12906. 1993. Coordinating geographic data acquisition and access: the National Spatial Data Infrastructure. (http://www.archives.gov/federal-register/executive-orders/pdf/12906.pdf) Accessed 10 February 2008.

Geodata.Gov. http://geodata.gov. [access to maps, data, search options, articles, and other tools]. Accessed 16 January 2008.

National Park Service Data Store. Secure online version. (http://science.nature.nps.gov/nrdata/). Accessed 16 January 2008.

6.9 Related Network-level SOPs

- Data mining in park units
- Digitizing documents
- Requesting park research permits
- Data collection with GPS and portable data recorders
- Developing GPS basemaps for use in GPS units
- Using GPS photolink to georeference photos
- Specifications for geospatial data deliverables
- NPSpecies data entry procedures
- NatureBib data entry procedures
- NPS Data Store data entry procedures
- Image management
- Management of audio or video data
- Collection and processing of specimens
- Use of automated data loggers (setup, calibration, maintenance, etc.)

CHAPTER 7. Quality Assurance and Quality Control

The success of the Inventory and Monitoring Program is dependent on the quality of the data it collects, manages, and disseminates. Analyses performed to detect ecological trends or patterns require data that are recorded properly and have acceptable precision, accuracy, 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 can greatly compromise the credibility of the program managing it.

Quality assurance (QA) can be defined 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. Quality control (QC) is a system of technical activities that measure the attributes and performance of a process, item, or service relative to defined standards (Palmer 2003). While QA procedures maintain quality throughout all stages of data development, QC procedures monitor or evaluate the resulting data products.

7.1 Objectives

The objectives of I&M network quality assurance and quality control are:

- Ensure that inventory and vital signs monitoring projects produce high-quality and credible data that can be confidently used by managers, researchers and the public
- Design, document, and implement standard quality assurance and quality control procedures that minimize or eliminate errors at every project stage.

7.2 Laws and Policies

Director's Order #11B, "Ensuring Quality of Information Disseminated by the National Park Service," was issued in order to promote data quality (Appendix 7A). It defines 'quality' as incorporating three key components: objectivity, utility, and integrity.

> *Objectivity* consists of: 1) presentation: whether disseminated information is being presented in an accurate, clear, complete, and unbiased manner within a proper context; and 2) substance: 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 information is not compromised through corruption or falsification.

Director's 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.

The I&M networks hold as policy that techniques and procedures chosen for a project will maximize data quality.

7.3 General Standards and Guidelines

To ensure that I&M networks produce and maintain 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). QA/QC procedures specific to any project must be detailed in a project's protocols and SOPs; however, some general concepts apply to all network projects.

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

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

Errors are more likely to be detected when data set expectations are clearly documented and what constitutes a 'significant' error is understood. The significance of an error can vary both among data sets and within a single data set. For example, a two-digit number with a misplaced decimal point (e.g., 99 vs. 9.9) may be a significant error while a six-digit number with an incorrect decimal value (e.g., 9999.99 vs. 9999.98), may not. However, one incorrect digit in a six-digit species Taxonomic Serial Number could indicate a totally different species.

QA/QC mechanisms are designed to prevent data contamination, which occurs when a process or event introduces either of two fundamental types of errors into a data set:

- Errors of *commission* include those caused by data entry or transcription, or by malfunctioning equipment. These 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.

The types of QA/QC needed for a project will vary based on data handling procedures (Figure 7-2).

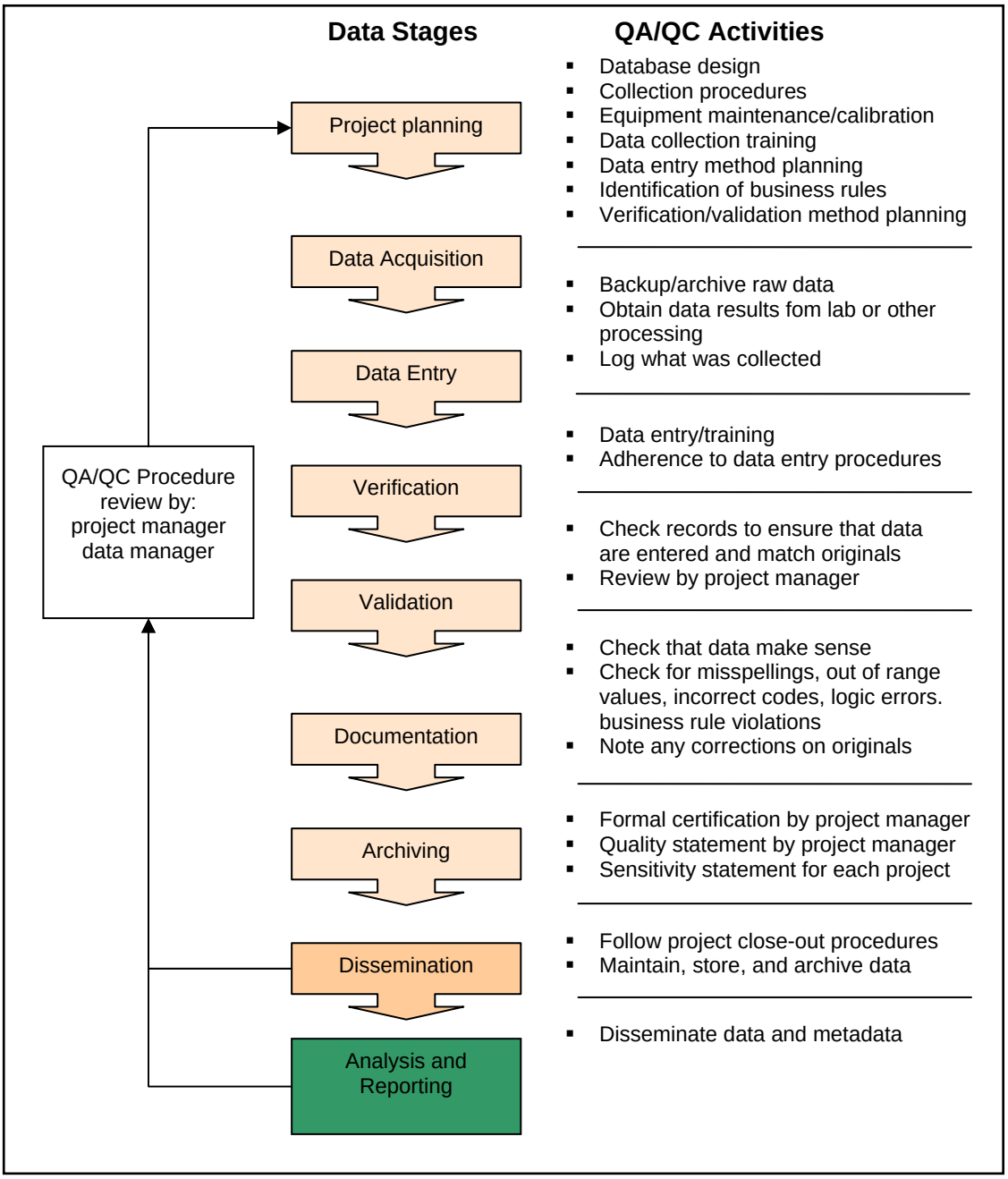

Figure 7-1. General course of data and associated quality assurance/quality control procedures.

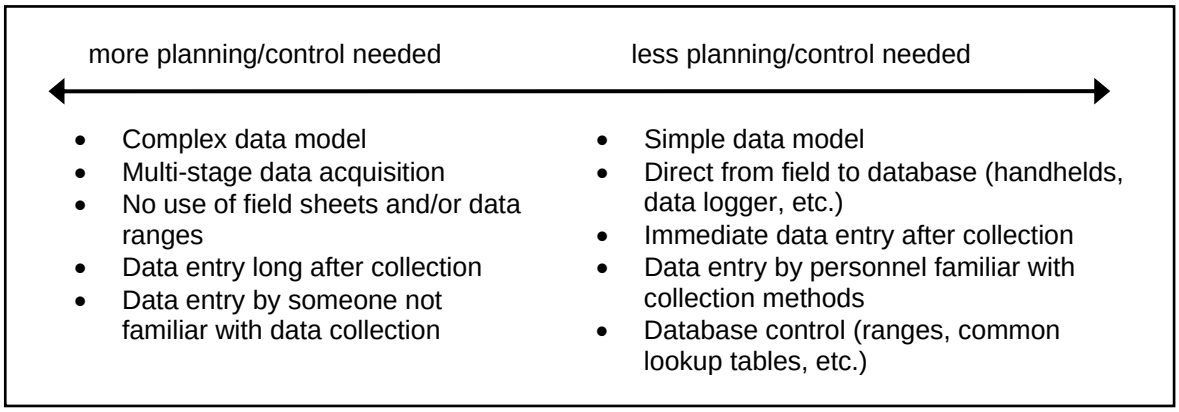

Figure 7-2. Some common information management elements influencing the amount of QA/QC needed.

7.4 Project Planning and Data Design

Quality assurance begins with choice of project methods and information to be gathered. Quality assurance is achieved by stipulating that:

- Common lookup tables are created for values recorded by multiple projects (e.g., weather variables, standard equipment, field personnel, etc.)
- Each vital sign protocol has SOPs that address core information management practices (e.g., field crew training, use of handheld computers, equipment maintenance, and data backup, entry, verification, and validation, etc.)
- Each project uses handheld computers for data collection if possible. If not, at a minimum, standardized field sheets are used
- Databases adhere to the standards outlined in this plan (Chapter Five)
- Database data entry forms, if necessary, resemble the field sheets
- Automated error checking features will be included in database applications
- Database application design will maximize the use of auto-fill, auto-correct, range limits, pick lists, and other constraints specific to projects
- Database applications will include a means to track the date a record is created or modified, the name of the person creating/modifying the record, and errors reported on the data after dissemination
- Database maintenance logs will be maintained for each network's database and housed in association with database files.

7.5 Data Collection

Chapter Six addressed general data acquisition and initial handling as well as changes to data collection protocols. Attention to detail during the data collection phase, however, is crucial to overall data quality. Unlike a typographical error that occurs during data entry, an incorrect observation in the field is not easily corrected.

The I&M Program adopts the following guidelines regarding data collection that affect data quality:

- Field crews will receive proper training

- Any project using field equipment will include a calibration and maintenance SOP that will also specify establishment of an equipment maintenance log
- Wherever possible and appropriate, data loggers or field-based computers will be used to collect data. When this is not possible, data will be recorded on pre-formatted, project-specific data sheets that reflect the overall design of the project and are designed to minimize the amount of writing necessary to effectively record observations
- When field sheets are necessary, or more appropriate, the format will be reflected in the computer data entry interface to help ensure all relevant information is recorded and subsequent data entry errors are minimized
- Completed field forms will be proofed for errors each day in the field. Edits will be made by drawing a single line through the information to be changed, adding any replacement information in clear writing next to the original entry, and initializing the changes
- Separate SOPs will describe the use of data loggers or field-based computers and will include direction for daily review and back-up.

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 on another computer where they can be further manipulated. The goal of data entry is to transcribe field observations into a computer database with 100% accuracy, although errors are unavoidable. Subsequent data verification is conducted to ensure that raw data matches entered data. Following verification, data validation may result in changes *to the entered data.* Data entry should be a separate operation from data validation and care must be taken to not impose validation (beyond that automatically imposed by programming rules in a database) during data entry.

The I&M Program adopts the following guidelines regarding data entry:

- To the extent possible, data entry will be automated. This may simply entail downloading data from field-based computers but may include the application of new technology to allow for machine-driven data entry (e.g., voice- or optical character-recognition software)
- Data will be entered as soon as reasonably possible after collection
- Data entry will be completed by someone familiar with data collection. The project leader (with assistance from the data manager if needed) must ensure that data entry staff is familiar with the database software, database structure, and any standard codes used by a network. At a minimum, data entry technicians should know how to open a data entry form, create a new record, edit an existing record, and exit the database properly. They must also learn how to correct mistakes made while typing
- Possibly, data could be entered by two qualified persons; one person to read the observations and the other to enter the data
- Data will be entered into pre-designed database forms that resemble field sheets and maximize error control.

7.7 Data Verification and Validation

Data quality is assessed by applying verification and validation procedures as part of the quality control process. *Data verification* checks that the digitized data match the source data, while *data validation* checks that the data make sense. Although data entry and verification can be handled by personnel who are less familiar with the data, validation requires in-depth knowledge about the data.

7.7.1 Data Verification

The I&M Program adopts the following guidelines regarding data verification:

- Project leaders are responsible for specifying in the project protocol one or more of the data verification methods available and ensuring proper execution. At the discretion of the project leader, additional verification methods may be applied
- Data verification is carried out by staff thoroughly familiar with data collection and entry
- All records (100%) should be verified against original source data
- A subset of randomly selected records (10%) should be reviewed after initial verification by the project leader. If errors are found, the entire data set should be verified again
- A record of the verification process for each data set, including number of iterations and results, will be prepared by the project leader as part of formal metadata generation (see Chapter 8 for more details)
- Spatial data collected as part of the project will be viewed in a GIS and visually inspected for accuracy (e.g., points located outside park boundaries, upland locations occurring in water).

Methods for Data Verification

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

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.

Visual review after data entry: Upon completion of data entry, all records are printed and compared with the original values from the hard copy. Errors are clearly marked and corrected in the database as soon after data entry as possible. Reliability increases if someone other than the person keying the data performs the review. Alternatively, two technicians (one reading from the original data and one checking the entered data) can perform this review.

Duplicate data entry: The data entry person completes all data entry as normal. Random records are then selected (e.g., every *nth* record) and entered into an empty replica of the permanent database, preferably by someone other than the person initially keying the data. A database query is then used to automatically compare the duplicate records from the two data sets and report on any mismatches. Disparities are manually reviewed and corrected if

necessary. This method adds the overhead of retyping the selected records, as well as the creation of a comparison query. However, it becomes increasingly successful as the value of *n* decreases. Professional data entry services frequently use this method.

Data verification can be improved by calculating summary statistics and identifying duplicate or omitted records. For example, the number of known constant elements, such as the number of sampling sites, plots per site, or dates per sample can be evaluated. Databases can also be built with controls to prevent duplicate records, although tests must be performed to identify missing records. The more checks that are devised to test for completeness, the greater will be one's confidence in the quality of the data set.

7.7.2 Data Validation

Validation is the process of reviewing computerized data for range and logic errors and may accompany data verification *only* if the operator has sufficient knowledge of the data and subject. More often, validation is a separate operation carried out *after* verification by a project specialist who can identify generic and specific errors in particular data types.

General step-by-step instructions are not possible for data validation because each data set has unique measurement ranges, sampling precision, and accuracy. Specific guidelines should be written into all project protocols and SOPS. Invalid data commonly consist of misspelled species names or site codes, wrong dates, or out-of-range errors in parameters with well defined limits (e.g., pH). 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 3 feet high). These types of erroneous data are called *logic errors* because they produce illogical (and incorrect) results. The discovery of logic errors has direct, positive consequences for data quality and provides important feedback to the methods and data forms used in the field. Histograms, line plots, and basic statistics can reveal possible logic and range errors.

The I&M Program adopts the following guidelines regarding data validation:

- Project protocols will address a process for data validation that includes at least one of the available methods
- Corrections or deletions as a result of data validation require notations in the original paper field records about how and why the data were changed, with the editor’s initials
- Modifications of the field data will be clear and concise while preserving the original data entries or notes (i.e., no erasing)
- Validation efforts will also include a check for the completeness of a data set since field sheets or other sources of data could be overlooked or inadvertently omitted.
- Use of automated routines and/or data summary and visualization (e.g., histograms, line plots, and basic statistics) will be maximized to identify possible logic and range errors
- Use of database programming will be maximized to control data entry. This will be achieved via the use of lookup tables and/or field-type design in a database (e.g., yes/no field types, input masks), and constraints such as valid ranges, uniqueness, referential integrity, and nullity.

Methods for Data Validation
The following general methods can be used as guidelines:

Data entry application programming: Certain components of data validation are built into data entry forms. This method is essentially part of the database design and is discussed earlier in this chapter. Not all fields, however, have appropriate ranges known in advance. Caution must be exercised when using lookup tables to constrain variable values. Values occurring outside the range set by a lookup table (established during database design) may not always be invalid. As part of data validation procedures, the project leader is responsible for correct use of lookup tables or other automated value range control.

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

Data quality assurance procedures should not try to eliminate outliers. Extreme values naturally occur in many ecological phenomena; eliminating these values simply because they are extreme is equivalent to pretending the phenomenon is 'well-behaved' when it is not. Eliminating data contamination is perhaps a better way to explain this quality assurance goal. When an outlier is detected (via GIS, database, graphic, and statistical tools for ad-hoc queries and displays), the possibility of contamination will be evaluated and noted.

Other exploratory data analyses: Palmer and Landis (2002) suggest calculations for assessments of precision, bias, representativeness, completeness, and comparability may be applicable and, for certain types of measurements, evaluation of detection limits 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 Review, Conformance, and Communication

The National Park Service requires QA/QC review and approval prior to communicating or disseminating data and information. Documentation of the QA/QC standards used in producing the information and that substantiate the quality of the information must be distributed with the related data and information. Mechanisms must also be in place for receiving and addressing comments or complaints pertaining to data quality (see also Chapter Ten).

As part of the close-out and evaluation stage of each network project, QA/QC procedures will be reviewed by the project leader and recommendations for change will be included in the annual report. Similarly, network data management staff will review and revise the QA/QC procedures included in this information management plan and/or SOPs as needed.

To ensure the highest quality, data stewards should conduct periodic audits to ensure compliance with the information management plan and vital signs protocols. Such quality checks promote a cyclic process of continuous feedback and improvement of the both the data and quality planning process. Audits may 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 final step in the QA/QC process is preparation of summary documentation that assesses overall data quality. The statement of data quality is composed by the project leader and incorporated into formal metadata for the data set. Metadata should also provide information on the specific QA/QC procedures applied and the results of review. Typically, data quality information will be conveyed as part of FGDC-compliant metadata and will be available via the NPS Data Store.

7.9 Roles and Responsibilities

Producing and maintaining high quality data is the responsibility of everyone involved with the handling of project data. 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. While Chapter Four discusses data management-related roles and responsibilities, selected QA/QC duties are emphasized below.

Project leaders need to:
- be aware of QA/QC procedures in protocols and convey their importance to technicians and field crews
- ensure compliance with the protocols
- plan for and ensure proper execution of data verification and validation
- review all final reports and information products.

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

The data manager is responsible for:
- developing network-wide SOPs to ensure data quality
- making project leaders, technicians, and others involved aware of the established procedures and enforcing adherence to them
- evaluating the quality of all data and information against NPS standards before accepting project data for dissemination outside the network
- performing periodic data audits and quality control checks to monitor and improve quality control operations.

7.10 References and Resources

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

National Park Service. 2002. Director's Order #11B: Ensuring quality of information disseminated by the National Park Service. (http://www.nps.gov/policy/DOrders/11B-final.htm)

Palmer, C. J. 2003. Approaches to quality assurance and informatin management for regional ecological monitoring programs. Pages 211-225 *in* Busch, D. E. and J. C. Trexler, editors. 2003, Monitoring Ecosystems: interdisciplinary approaches for evaluating ecoregional initiatives. Island Press, Washington, D.C.

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

7.11 Related Network-level SOPs

- QA/QC and field data collection standards
- QA/QC procedures for network vital signs data

CHAPTER 8. Data Documentation

Data documentation is perhaps the single 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 increases 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.

8.1 Objectives

The objectives of data documentation are:

- Ensure the longevity of data and their re-use for multiple purposes
- Ensure that data users understand the content, context, and limitations of data sets
- Facilitate discovery of data sets
- Facilitate interoperability of data sets and data exchange

8.2 Laws and Policies

Executive Order 12906 requires all federal agencies to fully document all spatial data collected or produced since 1995, either directly or indirectly, with metadata that meet specific standards developed by the Federal Geographic Data Committee (FGDC) as outlined by its *Content Standard for Digital Geospatial Metadata* (CSDGM) (Federal Geographic Data Committee 1994). Executive Order 12906 also directs agencies to plan for legacy data documentation and provide metadata and data to the public.

The NPS GIS Committee requires all new spatial data sets be described according to FGDC standards and the NPS Metadata Profile. The *NPS Metadata Profile* (http://science.nature.nps.gov/nrdata/docs/npsprofile.cfm) extends the original ten sections of the CSDGM to include the *FGDC Biological Data Profile, ESRI Metadata Profile (ESRI 2003), and the NPS Profile.* Profiles are simply formatted sections of information that are included in a formal metadata document. The NPS Metadata Profile is used by the Natural Resource and NPS Data Store for managing GIS and other data for both internal use and publication to the NPS GIS Clearinghouse within the NPS Focus Digital Library and Research Station.

The *FGDC Biological Data Profile* (FGDC Biological Data Working Group, and USGS Biological Resources Division 1999) contains elements (i.e., specific information sections) for describing biological data sets (both spatial and non-spatial). Metadata that contain these elements can be disseminated to the National Biological Information Infrastructure (NBII, http://metadata.nbii.gov/) Clearinghouse and the NPS Data Store.

The *ESRI Metadata Profile* (ESRI 2003) contains elements to better document aspects of ESRI data formats, particularly objects stored in geodatabases, and allow ArcCatalog to maintain metadata automatically.

The *NPS Metadata Profile* contained within the NPS Metadata Tools and Editor comprises information about the park units to which the data pertain.

8.3 General Standards and Guidelines

Documentation is essential if an organization is going to maintain the value of information it collects. As personnel change and time passes, information about data can be lost and, consequently, the data lose their significance. Complete metadata descriptions of the content and accuracy of a data set will help ensure the appropriate and long-term use of a data set. Such descriptions also may provide some protection for the producing organization if conflicts arise over the misuse of data.

From project development through final delivery of information products, I&M network data managers and project leaders must place a high priority on documenting the aims, quality, and meaning of data, and allow for the time this will take when scheduling. Detailed protocols provide an important source of information about the data produced by a project and therefore project leaders must track and document protocol versions. Metadata will adhere to strict standards, and their development will be guided by the practices described below. Crews will be trained to record decisions made in the field that affect data quality or meaning. Project leaders will start developing metadata at the onset of the project and set the tone for record-keeping throughout the course of the project.

8.3.1 Content Standards

Network metadata content standards will adhere to those set by FGDC and NPS policy as described above. The CSDGM consists of seven sections (Sections 1-7), all of which contain elements required for spatial data. The NPS Metadata Profile adds another section (Section 0). The Biological and ESRI Profiles are added as elements to Sections 1-7. A summary of the CSDGM and NPS Metadata Profile sections is provided below.

Section 0, NPS Information**:** purpose of the metadata, relevant park unit(s), and data steward.

Section 1, Identification Information**:** who produced the data set, when and why it was produced, and where it is from. Constraints on access (e.g., for sensitive data) and use are also recorded in this section. Also included are the geographic extent, bounding rectangles, taxonomy, and analytical tools used in processing (for biological data).

Section 2, Data Quality**:** the accuracy of attributes and geographic positions and the procedures used to ascertain accuracy. This section also documents the completeness and lineage of the data set. Lineage includes source(s) of the data and processing steps, and methodologies used (for biological data).

Section 3, Spatial Data Organization**:** methods of spatial reference. Mandatory for spatial data.

Section 4, Spatial Reference Information**:** coordinate system definitions. Mandatory for spatial data.

Section 5, Entity and Attribute Information**:** attribute names, definitions, codes and their meanings and other information essential to a basic understanding of the data.

Section 6, Distribution Information**:** methods and contacts used for obtaining data. Also documents information critical for using biological data formatted in ASCII.

Section 7, Metadata Reference Information**:** includes who created the metadata, when it was created, the profile used, and the frequency of update.

Sections 8, 9, 10: repeated sections used throughout the above-cited sections 1-7.

Metadata that are *fully compliant* with FGDC and NPS standards have entries in Section 0 and all element fields in Sections 1-7 where the *Optionality* field contains the term 'mandatory' or 'mandatory if applicable.' A mandatory element must be populated for every data set. A mandatory if applicable element must be populated if the data set exhibits the characteristic being documented by the metadata element. For example, the element defining the vertical coordinate system is mandatory if a data set contains elevation data.

Metadata that are *minimally compliant* with FGDC and NPS standards have entries in Section 0 and all 'mandatory' and 'mandatory if applicable' element fields in Sections 1, 6, and 7, and Section 2 for biological data. These include the fields used by the NPS Data Store. Data sets documented to this extent can be distributed via the Data Store's online upload utility.

8.3.2 Metadata Requirements for Different Data Types

Spatial Data will contain, at a minimum, all of Section 0 and the required elements of Sections 1-7.

Non-Spatial Data will include, at a minimum, all elements of Section 0, the required elements of Sections 1, 6, and 7, and Section 2 for biological data. The minimum requirements for non-spatial data therefore meet the requirements for minimum compliance with the FGDC and NPS standards. The Dataset Catalog application is a desktop tool for documenting a variety of non-spatial data, and allows the formatting and export of records to the NPS Data Store.

Relational Databases will be documented according to the standards outlined above. Complete documentation will also include entity relationship diagrams, business rules, and programming code. The NPS Metadata Profile currently does not support this type of documentation, so it should be stored separately from the formal metadata, in a folder with the database. Relational databases should also use internal documentation, including comprehensive table and field descriptions. The NRDT structure contains a field in which the name of a corresponding metadata document can be entered. NRPC has also developed a tool that extracts a data dictionary from MS Access databases, which is helpful in preparing or updating documentation.

Legacy Data will be documented to the extent possible according to the standards outlined above. Metadata records that accompany legacy spatial and non-spatial data sets are suitable for upload to the NPS Data Store if they include entries sufficient for minimum compliance with FGDC and NPS standards. Priority for legacy data set documentation is:

- Data sets needed for current project development
- Data sets used frequently by park staff or cooperators
- Historic data sets archived for possible future use

Data from Outside the Network are data that are generated and/or managed outside of I&M programs but used in analysis with I&M data or distributed in any manner by the NPS. These data require the same level of documentation produced for I&M-generated data, including but not limited to, data produced under contract with the NPS. Metadata will be requested from the originating entity by the network data manager or project lead.

Generally, external data will not be posted on an I&M network or park local or wide area networks (LAN/WAN) without accompanying metadata. This will include any metadata downloaded with the data, plus additional information regarding date of download and any alterations made to the data by NPS staff. Staff posting data will make a reasonable effort to make up for any deficiency in the original metadata, but should not create new metadata for data from well-known sources (e.g., USGS digital line graph). NPS staff will occasionally post 'value added' external data (e.g., a digital elevation model clipped to park boundaries and converted to the standard projection) to the NPS Data Store, and associated metadata will reflect the source data as well as processing prior to upload.

Sensitive Data. Metadata documentation (Section 1, Constraints on Access) provides one means of labeling sensitive data in order to ensure their protection and integrity over time. Sensitive data therefore will be documented according to the standards outlined above, although these data sets will not be uploaded to the NPS Data Store. Instead, they will be archived in secure locations on park and/or I&M network servers. These procedures will change as the NPS Data Store develops the procedures and acquires the necessary hardware to securely manage sensitive data.

Digital Photos. A national metadata standard for digital photos is currently under development by NPS. In the meantime, digital photo metatdata should include, at a minimum:

- Author – the name of the photographer, with agency affiliation
- Title – who or what is in the image (suitable for a brief caption)
- Location – description of where the image was taken
- Date – when the image was taken
- Park code corresponding to image location
- Access constraints – who may view the image (e.g., public, NPS only)
- Copyright information – restrictions on using the image
- Contact information – who to contact for further information

Electronic Documents. Most software programs provide a means of storing metadata within individual electronic files. In the case of Microsoft Word documents this is via the document properties item accessible from the file menu. For JPEG and TIFF files, the Exchangeable Image File Format (EXIF) data are embedded within the image file itself. Portable Document Format (PDF) documents also use the EXIF schema for file documentation. Although EXIF and Microsoft document properties are based on the Dublin Core basic element set and do not include all the elements of the CSDGM, use of these internal file documentation capabilities is encouraged.

8.4 Metadata Tools

ArcCatalog (from ESRI) is a multifunction application for managing spatial data and for editing FGDC compliant metadata. ArcCatalog uses stylesheets to display XML metadata in a format for easy viewing and editing. Because ArcCatalog metadata are linked to the data sets they describe, certain inherent properties of data sets, such as bounds, coordinate system, feature count, and attribute names, can be automatically populated and maintained in the metadata.

NPS Metadata Tools and Editor (http://science.nature.nps.gov/nrdata/tools/) is a custom software application for authoring and editing NPS metadata. It extends the basic functionality of ArcCatalog for managing spatial metadata and provides a stand-alone tool for creating and manipulating non-spatial metadata outside of ArcCatalog. The Metadata Editor formats metadata according to the NPS Metadata Profile. Like ArcCatalog, editing is done with stylesheets. A variety of stylesheets are offered, each tailored to a specific type of metadata and displaying only the elements or sections pertinent to that type. NPS-specific stylesheets are based on the NPS Metadata Profile extension and contain all elements in the FGDC, ESRI, and Biological metadata standards and profiles plus NPS-specific elements. Metadata can be parsed using the USGS MetaParser program, which comes bundled with the Metadata Tools. Output format is XML, suitable for upload to the NPS Data Store. Metadata Tools includes utilities for searching, cataloging, parsing, and spell-checking metadata records.

NPS Database Metadata Extractor is an add-in for MS Access 2000-2003. The extractor automatically harvests entity (table) and attribute (field) metadata from MS Access databases, including domains. It further allows the user to edit and review the harvested metadata and make batch edits and to export metadata to a FGDC-compliant XML file. Exported XML can be used in the Metadata Tools & Editor either by opening it to start a new metadata record or by updating with template to fill section 5 of an existing record. This tool will eventually become part of the NPS Metadata Tools & Editor.

Metadata Parser is a program developed by the USGS that comes bundled with both ArcCatalog and the NPS Metadata Tools and Editor. It is used to validate metadata records by checking the syntax against the CSDGM and to generate compliant output files for posting to clearinghouses. The parser generates a textual report indicating errors in the metadata, primarily in the structure, but also in the values of some of the scalar elements where values are restricted by the standard.

The "Metadata in Plain Language" questionnaire (Schweitzer 2006) is a user-friendly MS-Word document that can be used to elicit metadata needed to complete all FGDC

required elements for non-GIS data sets. It currently does not include questions specific to the Biological Data Profile nor the NPS Profile and therefore, does not alone, meet the content standards specified by the NPS Metadata Profile. Additional information must be obtained to comply with these specifications.

8.5 Work Flow Process

I&M networks should establish a standard operating procedure for metadata generation and maintenance. In general, metadata development begins with project design and planning (Chapter Three). As the project progresses, metadata documentation will be completed with information such as field season dates, protocol changes, and taxonomic information. Completion will occur after the data set is certified by the data manager or project leader, which is then followed by upload to the NPS Data Store. The relationship between local applications and tools, and national-level storage and delivery system is illustrated in Figure 8-1.

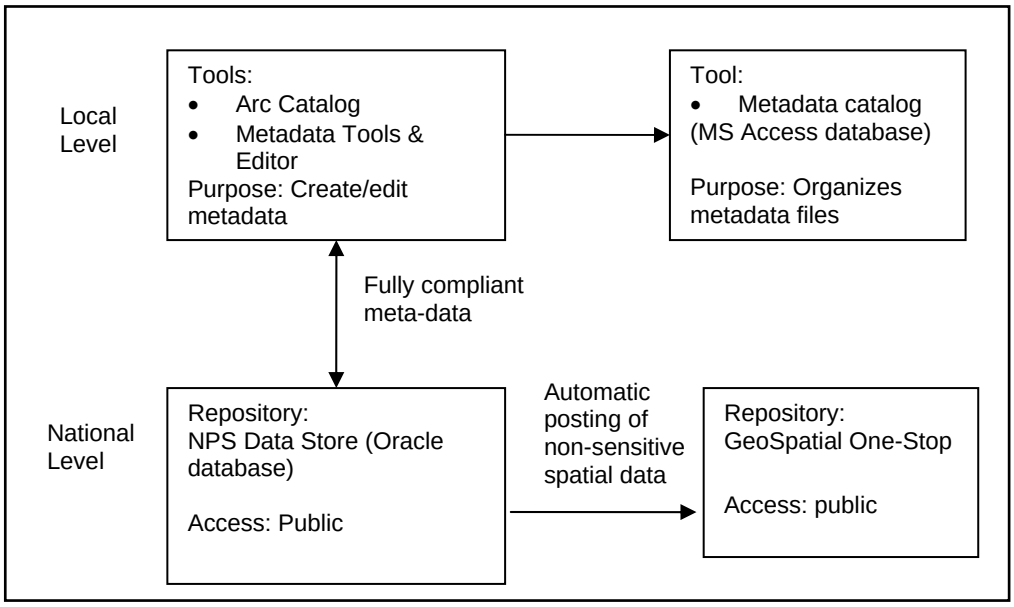

Figure 8-1. Local and national applications and tools for metadata

Once metadata are complete, the I&M network data manager or project leader saves one copy with the primary data set at the local level, and uploads an XML-formatted copy to the NPS Data Store.

Metadata can easily be searched with Metadata Catalog, one of the tools in the NPS Metadata Tools and Editor. Metadata Catalog creates a Microsoft Access database that stores the location, filename and other information about all metadata files in a selected directory.

8.6 References and Resources

Executive Order 12906: Coordinating Geographic Data Acquisition and Access, the National Spatial Data Infrastructure (http://www.archives.gov/federal-register/executive-orders/pdf/12906.pdf)

Federal Geographic Data Committee, 1994. Content standards for digital spatial metadata (June 8 draft). Federal Geographic Data Committee. Washington, D.C.

FGDC Biological Data Working Group, and USGS Biological Resources Division, 1999. Content Standard for Digital Geospatial Metadata - Biological Data Profile, FGDC-STD-001.1-1999 Federal Geographic Data Committee. Washington, D.C.

Michener, W. K. 2000. Medata. Pages 92-116 *in* Michener, W. K. and J. W. Brunt, editors, 2000, Ecological data: design, management and processing. Blackwell Science Ltd. Methods in Ecology Series, Malden, MA.

National Biological Information Infrastructure (NBII). NBII Metadata Activities. (http://metadata.nbii.gov/portal/server.pt). Accessed 16 January 2008.

National Park Service. Natural Resource Program Center. NPS Metadata Tools and Editor: a comprehensive editing tool for National Park Service metadata. (http://science.nature.nps.gov/nrdata/tools). Accessed 16 January 2008.

National Park Service. 2007. Digital photo metadata standard. (http://focus.inside.nps.gov/docs/DigitalPhotoMetadata/index.html). Accessed 16 January 2008.

National Park Service. NPS Metadata Profile. (http://science.nature.nps.gov/nrdata/docs/npsprofile.cfm). Accessed 16 January 2008.

Schweitzer, P. 2006. Metadata in plain language. U.S. Geological Survey, Reston, Virginia. (http://geology.usgs.gov/tools/metadata/tools/doc/ctc). Accessed 10 February 2008.

8.7 Related Network-level SOPs

- Metadata development guidelines
- Metadata interview form

CHAPTER 9. Data Ownership and Sharing

For the purposes of releasing data, the term *sharing* means providing data to an individual or entity upon request or as part of an agreement to do work that benefits the parks. The term *dissemination* includes releases to the public via publications or other means of delivery such as the NPS Data Store. Chapter 10 provides detailed standards and guidelines for data dissemination. Guidelines for data sharing, including sensitive data procedures, are described in this chapter.

9.1 Objectives

The objectives of data ownership and sharing guidelines are:

- To establish a clear understanding of the ownership of data and any associated responsibilities
- To protect sensitive data from unauthorized access and inappropriate use

9.2 Laws and Policies

The National Park Service defines conditions for the ownership and sharing of data, natural resource collections, and results based on research funded by the United States government in two principal documents - the Freedom of Information Act, 5 U.S.C. § 552, referred to as FOIA (Appendix 9A), and Office of Management and Budget (OMB) Circular A-110 (Office of Management and the Budget, 1999).

FOIA stipulates that the United States government, including the National Park Service, must provide access to data and information of interest to the public, regardless of whether or not they were created by the federal government. FOIA is intended to establish a right for any person to access federal agency records that are not protected from disclosure by exemption or exclusions. Under the terms of FOIA, agencies must make non-protected data and information products available for inspection and copying in public reading rooms or on 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 (http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=1996_register&docid=fr29my96-149.pdf)
- National Parks Omnibus Management Act (16 U.S.C. 5937) (http://www4.law.cornell.edu/uscode/html/uscode16/usc_sup_01_16_10_79.html)
- National Historic Preservation Act (16 U.S.C. 470w-3) (http://www4.law.cornell.edu/uscode/html/uscode16/usc_sup_01_16_10_1A_20_II.html)
- Federal Cave Resources Protection Act (16 U.S.C. 4304) (http://www4.law.cornell.edu/uscode/html/uscode16/usc_sup_01_16_10_63.html)
- Archaeological Resources Protection Act (16 U.S.C. 470hh) (http://www4.law.cornell.edu/uscode/html/uscode16/usc_sup_01_16_10_1B.html)

The OMB ensures that grants and cooperative agreements are managed properly. Federal funding must be disbursed in accordance with applicable laws and regulations. OMB circulars establish some degree of standardization government-wide to achieve consistency and uniformity in the development and administration of grants and cooperative agreements. OMB Circular A-110 establishes property standards within cooperative agreements with higher institutions and non-profit organizations. Section 36, "Intangible Property," describes the administrative requirements pertinent to data and ownership, stipulating that all data and materials collected or generated using National Park Service personnel or funds become the property of the park unit from which they were collected.

For additional information, see:

- The Freedom of Information Act 5 U.S.C. § 552, As Amended By Public Law No. 104-231, 110 Stat. 3048 (Appendix 9A). (http://www4.law.cornell.edu/uscode/5/552.html)
- A Synopsis of Laws and Policies Governing Public Dissemination of Natural Resources Data (Appendix 9C). Also contains the text to Section 36 of OMB Circular A-110.
- U.S. Department of the Interior FOIA regulations, instructions, and general information at http://www.doi.gov/foia/.
- Draft Director's Order #66: Freedom of Information Act and the Protection of Exempted Information (National Park Service, 2004a). (http://www1.nrintra.nps.gov/DO66/DO_66_Final_ Draft.doc)
- Reference Manual RM-66A: FOIA processing (National Park Service, 2004b). (http://www1. nrintra.nps.gov/DO66/RM-66A-Draft.doc)
- Draft Reference Manual RM-66B: Handling protected information (National Park Service, 2004c). (http://www1.nrintra.nps.gov/DO66/RM-66B-Draft.doc)

9.3 General Standards and Guidelines

Data and information will be shared internally through I&M network channels, externally through online databases and repositories, and by special request as appropriate, but only after quality control procedures have been fully implemented and data sensitivity assessed.

9.3.1 Collaborative Agreements

To ensure that proper ownership, format, and development of I&M network products are maintained, all cooperative or interagency work must be conducted as part of a signed collaborative agreement. Every cooperative or interagency agreement or contract involving an I&M network must cite OMB Circular A-110 under the *Reports and Deliverables* Section. The following shows appropriate language to use when citing Circular A-110:

> As the performing organization of this agreement, [institution or organization name] shall follow the procedures and policies set forth in OMB Circular A-110.

Every cooperative or interagency agreement or contract must include a list of deliverables and products clearly defined within each agreement or contract. Details on formatting and media types that will be required for final submission must be included. Agreements and contracts must list all products expected to result from the project. These include, but are not

limited to, field notebooks, photographs (hard copy and digital), specimens, raw data, and reports.

Project leaders should provide a schedule of deliverables that includes sufficient time for the NPS to review drafts before scheduled final submissions.

9.3.2 Due Notification

Network staff will notify investigators prior to making data sets available to the public. This will allow each investigator the opportunity to make a request in writing to further restrict access to the data set by the public. Network staff will review the investigator's request and determine whether the request will be granted and for how long the data set will remain restricted. Details on how data and information products will be made available to the public are provided in Chapter Ten.

9.3.3 Access Restrictions to Sensitive Data

With regard to natural and cultural resources, sensitive data are those that could be used to harm, remove, or destroy sensitive resources protected by units of the National Park System. Network staff members are responsible for managing access to sensitive data handled by their programs. These responsibilities include:

- Identify and classify all potentially sensitive park resources and information.
- Provide to superintendents a complete list of potentially sensitive park resources for which data may be collected as part of any project. The superintendent then determines what information should be protected.
- Ensure that all protected information, digital and hardcopy, is properly identified and marked.
- Ensure that all references to protected information are removed or obscured in any reports, publications, maps, data, or other form made available to the public.
- Inform investigators working on network projects 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 National Park Service.

The network coordinator, NPS project liaison, or network data manager will identify all potentially sensitive park resources to the principal investigator for each project. Conversely, the principal investigators for each project must identify any known references to potentially sensitive park resources.

Agreements with partners will include a standard confidentiality agreement which states that, among other things, the researcher will not share protected data or information with any other public, private, or academic party without channeling requests through the NPS project lead. Researchers may fulfill such requests provided there is a confidentiality agreement established between NPS and the new holder of the data.

When sharing data and information on sensitive resources with other federal agencies, the other agency will be told of the need to channel any requests for the data or information

through the NPS, as per NPS Director's Order #66 (National Park Service 2004a) and accompanying Reference Manuals 66A (National Park Service 2004b) and 66B (National Park Service 2004c).

9.3.4 Answering Requests under the Freedom of Information Act

Data stewards will forward records requested pursuant to the FOIA to the requester via the Park FOIA Officer in a timely manner. Data stewards will maintain records subject to the FOIA in a manner so as to make them rapidly available and to leave no doubt that the most recent version of accurate, non-sensitive information is provided to requesters. Data stewards are responsible for reading, understanding, and following applicable laws, policy, and procedural guides.

If information requested under the FOIA has not already undergone sensitivity classification, it will be thoroughly but rapidly assessed for presence of sensitive information and, if such information is found, the data steward will work with the appropriate park and regional FOIA Officers to formulate a response to the requester, which may include complete denial, complete fulfillment, or partial fulfillment once sensitive information have been cleared from the requested records.

9.4 References and Resources

Archaeological Resources Protection Act (16 U.S.C. 470hh) (http://www4.law.cornell.edu/uscode/html/uscode16/usc_sup_01_16_10_1B.html). Accessed 23 January 2008.

Executive Order 13007: Indian Sacred Sites (http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=1996_register&docid=fr29my96-149.pdf). Accessed 23 January 2008.

Federal Cave Resources Protection Act (16 U.S.C. 4304) (http://www4.law.cornell.edu/uscode/html/uscode16/usc_sup_01_16_10_63.html). Accessed on 23 January 2008.

National Historic Preservation Act 16 U.S.C. 470w-3. (http://www4.law.cornell.edu/uscode/html/uscode16/usc_sup_01_16_10_1A_20_II.html). Accessed 23 January 2008.

National Park Service, 2004a. Draft Director's Order #66: Freedom of Information Act and the Protection of Exempted Information. (http://www1.nrintra.nps.gov/DO66/DO_66_Final_ Draft.doc).

National Park Service, 2004b. Draft Reference Manual RM-66A: FOIA processing. (http://www1. nrintra.nps.gov/DO66/RM-66A-Draft.doc).

National Park Service, 2004c. Draft Reference Manual RM-66B: Handling protected information. (http://www1.nrintra.nps.gov/DO66/RM-66B-Draft.doc).

National Parks Omnibus Management Act 16 U.S.C. 5937. (http://www4.law.cornell.edu/uscode/html/uscode16/usc_sup_01_16_10_79.html). Accessed 23 January 2008.

Office of Management and the Budget, 1999. Circular A-110. (http://www.whitehouse.gov/omb/ circulars/a110/a110.html#1)

The Freedom of Information Act 5 U.S.C. § 552, As Amended By Public Law No. 104-231, 110 Stat. 3048 (http://www4.law.cornell.edu/uscode/5/552.html). Accessed 23 January 2008.

U.S. Department of the Interior FOIA regulations, instructions, and general information. (http://www.doi.gov/foia/). Accessed 23 January 2008.

9.5 Related Network-level SOPs

- General guidance: project deliverables
- Approving information for distribution
- Data sharing agreement

CHAPTER 10. Data Dissemination

Providing well-documented data in a timely manner is a primary goal of the I&M Program, and is essential to program success. This chapter describes the methods by which quality natural resource data and information collected by an I&M network are made available to park managers, researchers, educators, and the general public.

10.1 Objectives

The objectives of data dissemination are:

- Ensure non-sensitive data are easily discoverable and obtainable
- Ensure data that have not yet been subjected to full quality control are not released to the public, unless necessary in response to a FOIA request
- Distribute data with complete and accurate metadata that clearly identify who collected the data, what data were collected, and where, when and how the data were collected
- Identify and protect sensitive data from unauthorized access

10.2 Laws and Policies

Three major pieces of legislation form the basis for information dissemination and protection by federal agencies. They are the Freedom of Information Act, 5 USC Section 552 (FOIA), the National Parks Omnibus Management Act of 1998, 16 USC Section 5937 (NPOMA), and OMB Circular A-110 (Office of Management and the Budget, 1999). Each provides for the wide and proactive sharing of information and each recognizes instances where such sharing may be harmful.

The FOIA requires that federal agencies provide access to most data and information holdings, whether developed by employees or by cooperators. Agencies are directed to post commonly requested items in public reading rooms and/or on public web sites. Nine exemptions provide for withholding information, such as that relating to law enforcement actions and confidential personnel records that are protected. Of these, the third exemption allows withholding information that is protected under other statutes. These include:

- The Archaeological Resources Protection Act, which requires information on the location and nature of archaeological resources to be protected. (http://www4.law.cornell.edu/uscode/html/uscode16/usc_sup_01_16_10_1B.html)
- The National Historic Preservation Act, which directs that information regarding historic resources should be withheld if disclosure would cause significant invasion of privacy or risk harm to historic resources. (http://www4.law.cornell.edu/uscode/html/uscode16/usc_sup_01_16_10_1A_20_II.html)
- The Federal Cave Resources Protection Act, which prohibits release of information regarding the nature and location of caves. (http://www4.law.cornell.edu/uscode/html/uscode16/usc_sup_01_16_10_63.html)
- Executive Order 13007, Protection and Accommodation of Access to Indian Sacred Sites, which allows for maintaining the confidentiality of Indian Sacred Sites. (http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=1996_register&docid=fr29my96-149.pdf)

- The National Parks Omnibus Management Act (NPOMA), which provides protection of sensitive natural and cultural resource information. (http://www4.law.cornell.edu/uscode/html/uscode16/usc_sup_01_16_10_79.html)

The NPOMA, heralded as the NPS "research mandate" (Harmon 1999), established the Cooperative Ecosystem Studies Unit and I&M programs. While NPOMA clearly requires a higher level of information sharing with the research community and other interested parties as well as use in management decisions, the exemptions in Section 207 allow for protection of information about resources not previously specified in other legislation.

Another key policy statement, OMB Circular A-110 establishes federal government rights to use and share information gathered with federal support. The Circular also specifies that such information is subject to the FOIA and its exemptions. See Chapter 9 for details.

10.3 General Standards and Guidelines

I&M networks employ a number of distribution methods that ensure information collected and developed as part of network programs are made widely available to park employees, cooperators, and the public.

Data and information are disseminated to the public principally through the Internet and only after quality control procedures have been fully implemented, unless a request is made pursuant to the FOIA with consideration to sensitivity and ownership classification. Data that are draft, under review, or otherwise not finalized may be posted on I&M network intranet sites for access by park or other NPS staff.

Network staff will notify investigators prior to making data sets available to the public. This will allow each investigator the opportunity to request in writing to further restrict access to the data set by the public. Network staff will review the investigator's request and determine whether the request will be granted and for how long the data set will remain restricted.

10.4 Data Distribution Mechanisms

According to FOIA (specifically the 1996 amendments), all information routinely requested must be made available to the public via reading rooms and/or the Internet. An I&M network's principal means of distributing inventory and monitoring data will be the Internet (Table 10-1). A number of Internet-based databases and repositories have been developed as part of the NPS I&M Program to store and disseminate a variety of natural resource information. They include:

- I&M public web sites
- National applications, including NPSpecies and NatureBib (public release pending)
- The NPS Data Store

Table 10-1. Distribution mechanisms for I&M network-related data

Item	Repository
Reports -digital	NatureBib / NPS Data Store; Network web site
Reports - hard copy	Park libraries, park archives, network office
Protocols	NPS Protocol Database, NPS Data Store
Digital data sets – draft	Network server; network Intranet
Digital data sets – final / certified	Network Internet; NPS Data Store
Digital data – water quality	NPStoret
Digital data sets – sensitive	Network server
Project products Specimen vouchers Photographic film	Park museum or archive, or other academic or NPS museum facility

10.4.1 Network Web sites

I&M networks will regularly provide updated information about inventories and monitoring projects, including annual reports and detailed project reports through their individual public web sites (science.nature.nps.gov/im/units/) and intranet sites (nrintra.nps.gov/im/units). Networks will follow established I&M design and content standards.

10.4.2 National Data Distribution Tools

NPSpecies will be updated with observations and vouchers collected during the course of network projects, if these data are deemed important to the further documentation of species already present on certified park species lists, or constitute additions to these lists. NatureBib will be updated as new documents are produced or obtained via data mining efforts. Records in these databases are not publicly available; however, as they are transitioned to Service-oriented Architecture a public component will be developed.

Water quality data collected by I&M networks will be uploaded to EPA's STORET National Data Warehouse. The NPS Water Resources Division has developed NPSTORET, which is a desktop database application for I&M networks to manage data entry, documentation, and transfer of water quality data. Data from NPSTORET are transferred periodically to STORET (Figure 10-1). Individual networks are free to use NPSTORET for data entry and maintenance, or to develop a customized database compatible for data exchange and delivery.

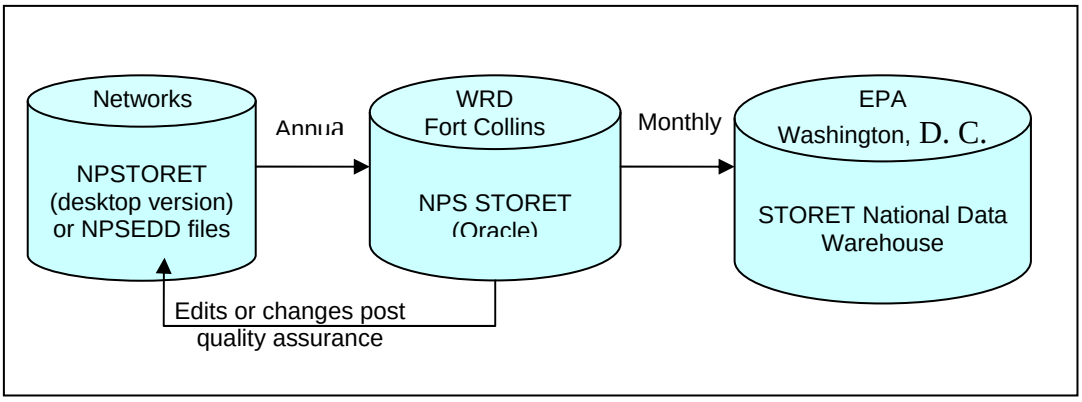

Figure 10-1. Flow diagram for water quality data from I&M networks to the STORET National Data Warehouse.

Data from short-term studies (such as biological inventories) should be posted to NatureBib or the NPS Data Store within two years after data collection was completed, or following publication of the investigator's results (whichever comes first).

Data from long-term air quality monitoring conducted by network parks or by cooperators working with the parks, including visibility, gaseous pollutants (mainly ozone), atmospheric deposition (wet and dry), and meteorology are available through the NPS Air Resources Division web site at: http://www2.nature.nps.gov/air/monitoring.

Long-term monitoring data collected as part of the I&M Vital Signs Monitoring Program will be posted to the NPS Data Store at appropriate regular intervals, or when trends analyses have been completed and reported on by the network. Before data are posted, the investigator or project leader will be asked to verify the final data set and metadata, and to identify any sensitive data or information. Sensitivity will be documented in the metadata which will accompany all data sets.

Most data products resulting from the Natural Resources Inventories conducted through the NRPC and the I&M Program can be accessed via the Inventories web site at http://science.nature.nps.gov/im/inventory/index.cfm.

10.4.3 Sensitive Data

Data released outside of the federal government without a confidentiality agreement is considered a release of information under the Freedom of Information Act and triggers the application of the "release to one, release to all" principle. There are many venues for release of sensitive data that constitute release to the public. Therefore, information should be reviewed for sensitivity by qualified NPS staff prior to inclusion in these media. They include:

- Internet posting
- Interpretive programs, media, and activities
- News releases
- Resource management storing and sharing activities

- NEPA compliance documents

Inadvertent access may be deemed a release depending on the circumstances. Network staff will institute security measures to ensure that any person uploading records to public databases is familiar with the procedures for identifying and abridging protected sensitive information.

10.4.4 Feedback Mechanisms

I&M network web sites will provide an opportunity for NPS staff, cooperators and the public to provide feedback on data and information gathered as part of I&M programs. A "comments and questions" link will be provided on the main page of the site for general questions and comments. A more specific "data error feedback" link will direct comments to network staff pertaining to errors found in web site-accessible data.

The following procedure describes a process for networks to receive, verify, and correct data errors identified by public users:

- Users send a notification through the network web site.
- An acknowledgment of receipt is sent to the user.
- Information is input into an error log table incorporated either in the project database or a specific error tracking database developed for the network.
- Network staff determine whether the data in question are correct or incorrect. If correct, the user is informed that no corrections are to be made and the information stands. If incorrect, network staff make the appropriate corrections and notify the user and original data collectors.
- The network web site is refreshed with the corrected information.

10.5 References and Resources

Archaeological Resources Protection Act (16 U.S.C. 470hh) (http://www4.law.cornell.edu/uscode/html/uscode16/usc_sup_01_16_10_1B.html). Accessed 23 January 2008.

Executive Order 13007: Indian Sacred Sites (http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=1996_register&docid=fr29my96-149.pdf). Accessed 23 January 2008.

Federal Cave Resources Protection Act (16 U.S.C. 4304) (http://www4.law.cornell.edu/uscode/html/uscode16/usc_sup_01_16_10_63.html). Accessed on 23 January 2008.

Harmon, D. 1999. The new research mandate for America's National Park System: Where it came from and what it could mean. George Wright Forum 16(1):8-23.

National Historic Preservation Act 16 U.S.C. 470w-3. (http://www4.law.cornell.edu/uscode/html/uscode16/usc_sup_01_16_10_1A_20_II.html). Accessed 23 January 2008.

National Park Service. Air Resources Division.
(http://www2.nature.nps.gov/air/monitoring/). Accessed 16 January 2008.

National Park Service. Natural Resources Inventory & Monitoring Program, inventories.
(http://www1.nrintra.nps.gov/im/inventory/index.cfm). Accessed 16 January 2008.
[Intranet)

National Parks Omnibus Management Act 16 U.S.C. 5937.
(http://www4.law.cornell.edu/uscode/html/uscode16/usc_sup_01_16_10_79.html).
Accessed 23 January 2008.

Office of Management and the Budget, 1999. Circular A-110.
(http://www.whitehouse.gov/omb/circulars/a110/a110.html#1)

10.6 Related Network-level SOPs

- Approving information for distribution
- Data sharing agreement
- Procedures for developing, posting, and updating network web site content
- Guidance document: project deliverables
- Data summary and analysis

CHAPTER 11. Records Management, Data Maintenance, and Archiving

I&M networks are responsible for managing the documents, specimens, photographs, digital files, and other products resulting from network projects, administration, and activities. Much of the information generated by I&M networks will only increase in value and applicability over time. The potential for loss of data, documents, or objects can come from a variety of sources, including catastrophic events (e.g., fire, flood, and earthquake), user error, hardware failure, software failure or data corruption, theft, and intentional acts of vandalism. A fundamental responsibility of network data management is to establish procedures that will guarantee that I&M information will be accessible and usable for future generations.

11.1 Objective

The objective of I&M network records management, data maintenance, and archiving procedures is:

- To ensure that I&M network information can be easily found, obtained, shared, and properly interpreted by a broad range of users in perpetuity.

11.2 Laws and Policies

Records are all documentary materials, including books, electronic data, maps, moving images, papers, photographs and sound recordings, made or received by the National Park Service during the transaction of public business. Directions for managing these materials, as well as electronic files, are provided in NPS Director's Order 19: Records Management, and its appendix, NPS Records Disposition Schedule. 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, specifically, the products from Inventory and Monitoring 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."

Direction for the curation of specimens or other objects related to I&M research, and the archiving of reports, publications, maps, or other documents, are contained with the National Park Service Museum Handbook, Part II, and individual park museum Scope of Collection documents.

The NPS General Support System (GSS) Backup Activity Plan (NPS 2007) and the NPS GSS Standard Procedures for Backup, Storage, & Recovery (NPS 2007b) both cover information relevant to the implementation of a network backup of electronic files by providing a minimum set of standards. These documents are provided by the NPS Office of the CIO and apply to all "headquarters, regions, and parks."

11.3 General Standards and Guidelines

NPS Director's Order 19, Appendix B – Records Disposition Schedule has 40 pages devoted to classifying and filing NPS federal records and disposition specifications for each NPS federal record category. Generally, permanent federal records should be transferred to the National Archives when they are 30 years old. To the extent possible, an I&M network should comply with the most current NPS disposition schedule.

I&M network procedures should also be established and documented to meet the following standards:

- Where practical, maintain mission critical and permanent paper records in both hardcopy and electronic file format, and protect and preserve them indefinitely
- Maintain data sets in file formats no more than two versions behind current software versions, or store in American Standard Code for Information Interchange (ASCII) format, complete with data and file documentation
- Ensure information can be easily obtained, shared, and properly interpreted by a broad range of users
- Ensure backup, storage, and recovery practices for archived electronic files equal or exceed the minimum standards established by the NPS Office of Chief Information Officer
- Store all electronic files on servers in a networked environment using approved file-naming standards and file directory structures
- Maintain all data, programmatic, and administrative electronic files indefinitely or to comply with Director's Order 19 Appendix B, Records Disposition Schedule
- Ensure all short- and long-term projects are well-documented, organized, and protected according to local and national standards and guidelines
- Ensure all collected natural history specimens and objects are well-documented and preserved in perpetuity.

11.3.1 Digital Data Maintenance

Digital data pose particular challenges to long-term maintenance and preservation. Technological obsolescence is a significant cause of information loss, and data can quickly become inaccessible to users if stored in out-of-date software formats or on outmoded media. Effective maintenance of digital files depends on 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 (Vogt-O'Connor 2000). As software and hardware evolve, data sets must be continuously migrated to new platforms, or they must be saved in formats that are independent of specific platforms or software (e.g., ASCII delimited files).

An essential part of any archival holding (digital or non-digital) is its accompanying explanatory materials (Olson and McCord 1998). Effective long-term maintenance of information products depends on thoughtful and appropriate data documentation.

Directory Structures

Electronic files on I&M network file servers should be managed within a hierarchical set of directory structures, including folders for administrative files, data, publications, and reports. The primary repository for all electronic files should also be identified and access provided to all program staff. Key aspects of the file management strategy include the following:

- To the extent possible, all applications work (e.g., spreadsheets, GIS, documents, databases) should be conducted from a common server
- The server-based copy of all electronic files will be the primary version. Local electronic file versions on personal computers are not authoritative
- Generally, all electronic files should be maintained in a networked environment that enables immediate retrieval. Although there are some exceptions, off-line storage and retrieval of electronic files from external media will be discouraged (except for the off-site backup storage of data files for security)
- Working files are kept separate from archived products
- Finished products are read-only
- Standards such as naming conventions and directory structures should be enforced within shared and archived sections (e.g., libraries, GIS, databases). Although less stringent in other sections, these conventions are encouraged as good practice
- Version control is implemented in both active and archived directories.

File Naming

All electronic files should be named according to a set of guidelines and standards that apply to all electronic files created or maintained by staff or cooperators. The following guidance for file naming standards should be considered:

- File names should be as short as possible, but include all mandatory information
- The electronic file name should be unique, such that the chances for another electronic file having the same name on a local area network are unlikely
- A versioning system of numbers (e.g., version1, version2, final, or v1, v2, etc.) should be used on documents
- Generally, use lower case characters. However, it is an acceptable practice to mix upper and lower case letters in a filename to separate a concatenation of words, such as words in a title
- Do not use spaces in file names
- Do not use any special characters other than the underscore
- Use the underscore to separate categories in an electronic file name
- Use generally recognizable abbreviations or spell out words
- Avoid uncommon acronyms, abbreviations and codes. Generally, avoid codes that require users to refer to another source for the code description. Agency-wide recognized codes such as park or network codes are acceptable
- Use leading zeroes for numbers 1 through 9

- Use the default file extension recommended by the software application associated with the file (Examples: .doc, .xls, .mdb, .rtf).

Version Control
Version control allows I&M network staff to manage changes made to data sets or documents, and provides an auditable trail of individual versions. As a general rule, to safeguard against loss or damage to data, changes or edits should not begin without having in place the ability to reverse them. In its simplest form, version control is simply saving a copy of a file before beginning work on it, and using an established version naming convention. Embedding dates or sequential numbers into file names, or establishing specific file directory structures are all techniques that can be used for simple version control. Version control can also include the MS Word "track changes" feature, or the SharePoint versioning features.

The I&M network data manager determines the version control method and conventions that will be used. Software tools that assist in file management can be helpful; for example, a database that tracks revision histories, or backup routines that allow for automatic file renaming and archiving. Commercially-developed or open-source version control software may be an option. A version control system should:

- Preserve previous versions of documents for possible recovery
- Track documents as they change during the course of the developmental and editorial phases of document/report creation
- Prevent conflicts between multiple collaborators by prohibiting multiple edits to the same file at the same time
- Audit the document creation process by tracking who changes a file, when they make the change, and what changes they make
- Reduce storage requirements by eliminating multiple copies of complete documents

Backup and Recovery
The need for backup and restore capabilities is critical to ensure proper safeguarding of irreplaceable information. Data recovery can be required due to catastrophic events (e.g., flood, hurricane, fire), software or hardware failure, or malicious activities. Every I&M network must have formal backup procedures documented and in place for both system- and workstation-level information. Storage procedures for data backup must include media rotations to both onsite and offsite locations. In addition, an essential component of any backup system also includes the regular testing of restore capabilities.

I&M networks should discourage electronic file storage on personal computers and, to the extent possible, essential electronic files should be managed on a network server. In situations where an employee works with a computer that is off-network, each employee will be responsible for archiving and storing electronic files, with scheduled transfers of files to a network server that is backed up.

11.3.2 Digital Archives

Both electronic and hardcopy files should be consolidated and packaged for archiving when a project is complete or when milestones are reached. Project leaders are responsible for packaging electronic files and data, and for preparing materials for the curator. Network data

managers can be of assistance and provide guidance and clarification concerning the process or workflow. Project protocols should designate who will be responsible for product archiving, integration, backup, and distribution. Figure 11-1 illustrates the recommended paths that product archiving and storage follow. Procedures for hard-copy archiving are described in section 11.3.3 (below).

Electronic File Archiving

Digital files destined for archiving should meet the following requirements:

- Comply with the network's electronic file naming standards and hierarchical directory structure
- Comply with data documentation and metadata standards
- Clearly distinguish between sensitive and non-sensitive information
- Integrate all deliverables, such as final reports, into appropriate catalogs, electronic libraries and clearinghouses, as described in this plan or specific project protocols.

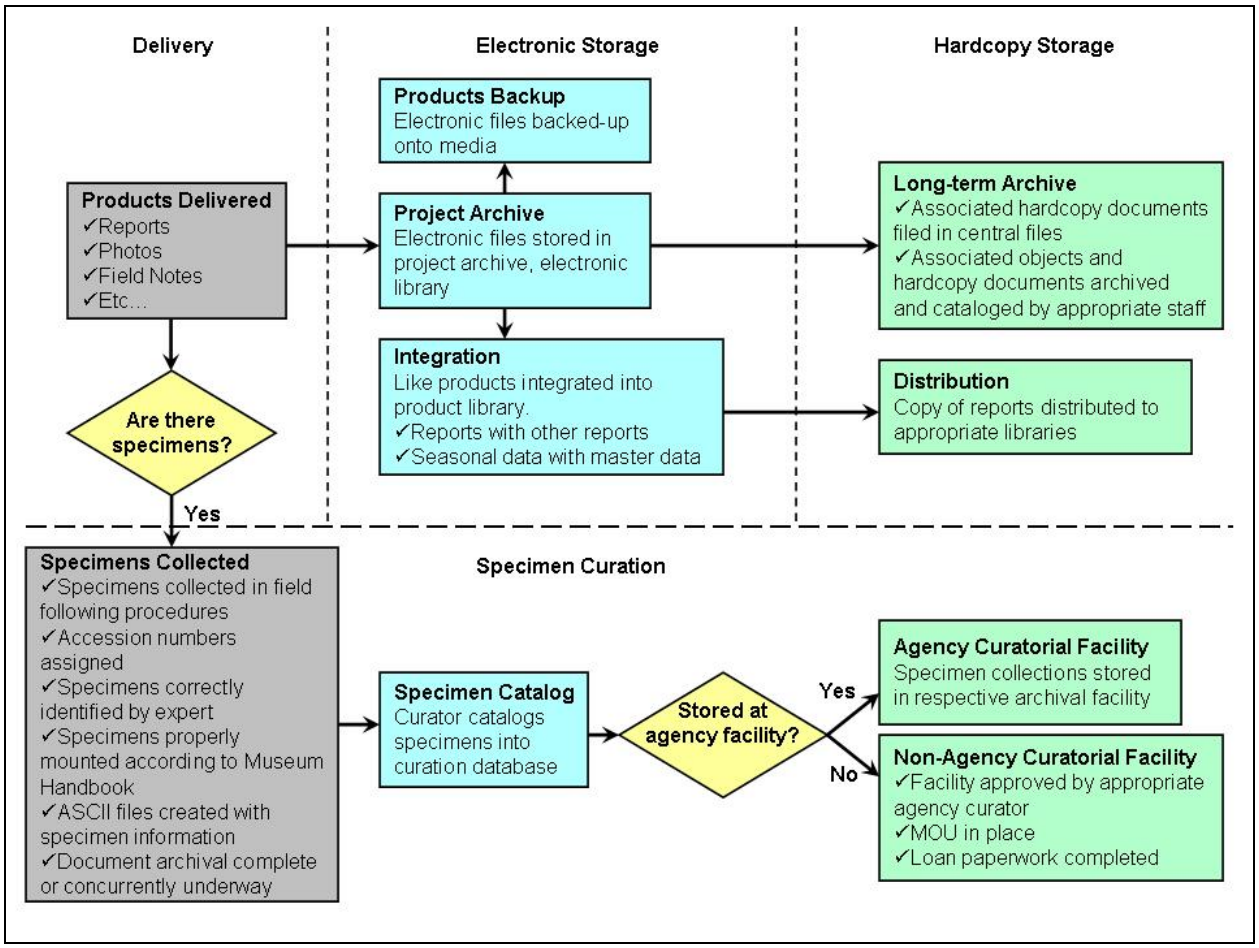

Figure 11-1. General workflow for storage and archiving of electronic (blue) and hard copy (green) scientific research products.

Short-term Projects

Short-term are those projects that usually last from one to three years and include individual research projects, inventories, or pilot work done in preparation for long-term research or monitoring.

Upon project finalization, it is highly recommended that a set of ASCII comma-delimited text files 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 data set (typically in flat or relational database formats) and will help ensure the data are usable in a wide range of applications or platforms. A tool such as Access_to_ ASCII.mdb (available from http://www1.nrintra.nps.gov/im/datamgmt/dtbases/links_sources/index.cfm), which automates the creation of these ASCII files from MS Access databases, will help with file conversion.

A second reviewer (preferably the project leader) 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. All finalized files will be stored in the archive section of the project directory folders.

In addition to creating ASCII files, staff should regularly update completed and archived data sets in older versions of MS Access, with the goal of having no data set more than two versions behind the current version. Converted databases will require thorough quality control if they are actively being used for data entry or analysis and if no documentation of prior quality assurance/quality control measures is available. Forms, queries, reports, and data entry will be thoroughly tested and corrected where any problems occur. All previous versions of the data set will be saved.

Long-term Projects

Long-term projects include vital signs monitoring and other multi-year research and monitoring projects performed by the parks, other agencies, or cooperators, for which data acquisition and entry will continue for a period of four years or longer.

Long-term monitoring databases will require periodic conversion to current database formats. Converted databases will require careful QA/QC to correct problems caused by modifications in programming code or other software changes. All active or long-term databases will conform to the current NPS software version standards.

Long-term projects also will have variable long-term data archiving requirements. Raw data should be stored in perpetuity. Processed data sets should be archived as a complete unit before any modifications are made to project protocols. Depending on the project, it might also be necessary to preserve interim data sets (data “milestones”) over the long-term. Archived data sets or subsets destined for long-term archival will be saved, whenever possible, in their native formats in addition to ASCII text files. Specific data archiving requirements for ongoing projects should be spelled out in the Standard Operating Procedures for each project.

Spatial Data
Spatial data sets essential to I&M networks will be maintained in a format that remains fully-accessible by the current ArcGIS version. ArcGIS has maintained compatibility with previous data formats and, while shapefiles have retained functionality in ArcGIS, coverages may require conversion if they are no longer supported.

Both uncorrected and corrected GPS data (e.g., .ssf and .cor files) should be archived in their native format in addition to the corresponding GIS files that are created.

11.3.3 Object Curation and Document Archiving (non-digital products)

The NPS Museum Management Program is responsible for the curation of physical objects related to park units (e.g., specimens) and the archiving of documents, publications, and other printed materials. The Automated National Catalog System (ANCS) is the data system used by the Museum Management Program to manage this information; a stand-alone copy of the application is maintained at each park with museum facilities. NPS museum management staff are responsible for managing this application and museum holdings.

Park museum curators are a source of expertise, advice, and guidance on archiving and curatorial issues and have a role in almost all projects undertaken by I&M networks. Project leaders will involve park curators when projects are in the planning stage, to ensure all aspects of specimen preservation or document archiving are considered and any associated expenses are included in project budgets. Section 4.3.16 of Director's Order 24 requires "*…project budgets to include funding for the basic management of collections that are project-generated."* The costs for curation include cataloging, labeling, specimen preparation, initial storage of objects and specimens, and the organization and storage of project documentation.

The overarching technical guidance for archival procedures is found in the NPS Museum Handbook. In particular, Part II, Appendix A – Mandates and Standards for NPS Museum Collections lists the cultural and natural history laws, regulations and conventions for NPS museum collections and should be reviewed prior to object collections.

The long term management of physical objects and records will require the close cooperation of the individual collectors (whether NPS personnel, contractors, or permitted researchers) and local park museum personnel. Much of the responsibility will rest with the collector who remains the sole source for all curated documentation and the principal authority for essential collection information.

Successful collaboration with NPS museum staff includes adequately preparing materials for accession. The types of information neeed by curators include:

- A project summary sheet listing project name, research permit, abstract and purpose, contact information, sensitivity, use of materials, and all materials included in the archive submission
- Clearly labeled materials with appropriate administrative boundary name (e.g., park, network), date(s), accession number, catalog number (if assigned), and project number.
- Compliance with archival paper and folder type requirements
- Printed and labeled photos

- Voucher specimens labeled according to the NPS Museum Handbook and any loan paperwork if specimens will reside outside an NPS facility
- A .csv file containing voucher attribute information formatted for automated data transfer to the ANCS data system
- A CD or DVD of all related, electronic materials and labeled with appropriate administrative boundary name, date created, date(s) of information, accession number, project number, list of contents.

Natural History Specimens

Natural history specimens collected under the auspices of an I&M network, along with associated field notes, should be transferred to the park in which they were collected for either inclusion in the parks' museum collections or for transfer to a repository approved by the park. In this latter case, the park will enter into a formal, documented relationship with the repository and the specimens will be considered to be "on loan" with clearly stated terms and responsibilities. All loans are subject to the approval of the park superintendent.

Park curators should be involved at an early stage in the permitting or agreements process; early cooperation between curators and collectors will facilitate compliance with the current NPS Management Policies. Park curators can provide the collector or network with a pre-formatted MS Access database template, the use of which will greatly simplify eventual importation of catalog data into the ANCS. At a minimum, parks will need such data in comma-delimited format (.cvs) for automated uploading.

Photographs

Archivists have been reluctant to fully embrace digital photography because of a concern that, with the accelerating rate of technological change, documentary heritage is in danger of being lost in the information age (Cox 2000). While a network maybe prepared to accept both digital and analog photographs, the particular photographic methods to be used by any given project should be developed jointly by the project leader and network staff or cooperators. Early involvement of curatorial staff will insure that the specific preservation considerations of particular photographic methods are understood and considered.

For projects using analog photos, networks are strongly encouraged to use Kodachrome 35mm slide film, which has the longest dark-storage dye stability of any color film (Wilhelm and Brower 1993), or other professional-grade slide film. Slides should be labeled using indelible pigment, pH-neutral ink (e.g., Pigma Micron), or using laser-printed archival-quality slide labels. Labels must include the following information:

- Title – who or what is in the image
- Location – where the image was taken
- Date – when the image was taken
- Park Code
- Access Constraints – indication if the image should be restricted due to sensitivity

Additional important information to associate with images includes copyright information (restrictions on using the image) and contact information (who to contact for further information); the geographic location with UTM NAD83 coordinates (based upon a

network's mapping standard) and datum. All slides should be placed in archival polypropylene slide-sleeves and stored away from direct light, in climate-controlled conditions, and preferably in an archival box.

Additionally, slides may be scanned and saved digitally in Tagged Image File Format (TIFF). If scans are used as the primary means of distributing or reproducing the images, the physical life of the original slides will be greatly extended.

If photographic prints are desired, staff and cooperators are strongly encouraged to use true black-and-white films (e.g., Ilford Delta, Kodak T-Max or Tri-X). Prints and negatives will be stored in individual polypropylene sleeves and within archival boxes. Each print will be labeled on the back, using archival-quality labels that are either laser-printed or hand-labeled with the same information elements required for slides. If a contractor is submitting photographs, corresponding TIFF files may also be submitted.

Every image, regardless of format, should be entered into a network photo management database, where attributes such as electronic file name, keywords, project, photo description, photographer, date, and location are cataloged. The National Park Service has developed and adopted a set of suggested metadata attributes (http://focus.inside.nps.gov/docs/DigitalPhotoMetadata/index.html). All photo files and the associated photo database should be housed on the archive portion of the network server.

11.4 References and Resources

36 CFR 2.5: Research Specimens (Code of Federal Regulations, 2001). Provides direction for managing natural history collections

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).

National Park Service, 2001b. Director's Order 19, Appendix B: Records Disposition Schedule. Revised May, 2003. U.S. Department of the Interior, National Park Service, Washington, D.C.

National Park Service. Natural Resources Inventory & Monitoring Program, I&M network data management: code and tools. (http://www1.nrintra.nps.gov/im/datamgmt/dtbases/links_sources/index.cfm). Accessed 16 January 2008. [Intranet site]

National Park Service. NPS Digital Photo Metadata Standard. (http://focus.inside.nps.gov/docs/DigitalPhotoMetadata/index.html). Accessed 16 January 2008. [Intranet site]

National Park Service, 2007. NPS General Support System, Backup Activity Plan. U.S. Department of Interior, National Park Service, Washington, D.C.

National Park Service, 2007b. NPS GSS Standard Procedures for Backup, Storage, & Recovery. U.S. Department of Interior, National Park Service, Washington, D.C.

National Park Service. 2000. NPS Director's Order 24: Museum Collections Management.

Olson, R. J. and R. A. McCord, 1998. Data Archival. Pages 53-57 in Michener, W. K., Porter, J. H., and S. G. Stafford. Data and information management in the ecological sciences: a resource guide.

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. (http://www.cr.nps.gov/museum/publications/conserveogram/19-21.pdf)

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

11.5 Related Network-level SOPs

- Network I&M Program Checklist of Project Specifications
- Digital Photo Management Strategy
- Electronic File Naming Guidelines and Standards
- Program Guidance and Rules for File Directory Structure
- Network Project Template Directory Structure
- Standards and Guidelines for Backup, Storage, and Recovery
- Network Archive and Storage Locations

CHAPTER 12. Project Tracking and Documentation

I&M networks face the challenge of managing multiple projects, contracts, and agreements; budgets; timelines and deliverables; and permanent, term and temporary staff. As networks develop and become fully staffed, many have found that a project tracking and documentation system is necessary to function effectively.

12.1 Objectives

The objective of project tracking and documentation procedures are:

- Efficiently manage and coordinate contracts, permits, agreements, and associated timelines, schedules, and budgets
- Ensure project products are delivered on time and per specifications
- Manage information needed for annual reporting and accountability
- Improve accountability for I&M network natural resource inventory and monitoring efforts and provide a means of tracking accomplishments

12.2 Laws and Policies

There are no overarching NPS or DOI policies related to specific project tracking methods or requirements. Individual agreements established by I&M networks contain specific requirements, timelines and deliverables that must be attained.

The NPS Associate Director of Stewardship and Science established in 2001 the requirement that each I&M network submit an Annual Administrative Report and Workplan at the end of each fiscal year.

12.3 General Standards and Guidelines

12.3.1 Project Organization

All project electronic files should be stored in an intuitive project directory structure that is clearly understood by all network staff. Network digital directory structures should be organized at the project level, such that most or all digital files associated with a project are filed under a common root directory. Project file names will adhere to the naming conventions established by the network. Physical objects acquired as part of a project will be stored according to the network's specification for records management and object curation.

12.3.2 Project Tracking

Networks are encouraged to implement a project tracking system or database that will serve as the primary organizational tool for cataloging and searching information on I&M projects. This system should, at a minimum:

Maintain a master list of projects: project leaders and other users will be able to locate rapidly information related to completed, current, or upcoming projects (e.g., objectives, contact information, status, funding sources, project codes from various NPS systems)

Track product deliverables: a list of project deliverables will allow oversight of permitting contracting, content and formatting specifications, delivery schedules, and disposition schedules.

A project tracking database should be hosted on a network file server and be accessible by all appropriate staff. Custom queries or reports will help administrators manage projects and facilitate reporting on project status, accomplishments and delivered products.

12.3.3 Additional Project Metadata

Long-term monitoring projects may require documentation beyond the scope of established FGDC standards. Documentation should be provided for algorithms, output files, and analytical products which may reside in different systems and formats, and could potentially be overlooked when distributing or applying project data. Data use and data request histories, and information on secondary research or publications resulting from long-term monitoring projects, should also be maintained. The network's project tracking database potentially could be used to organize this information. Specific methods for documenting this information will be tested for ease of use and are likely to evolve over time as needs become apparent and new solutions are developed.

12.4 References and Resources

12.5 Related Network-level SOPs

- Checklist for initiating a new project
- Project specifications and deliverables
- Project management database application
- Project data life cycle guidance

CHAPTER 13. Implementation

The Inventory and Monitoring Program data management plans contain practices that may be new to many NPS staff and cooperators. While this document and the accompanying network-level data management plan and SOPs offer practical guidelines and solutions, successful implementation will require continuous formal and informal education of staff and cooperators, and a long-term and disciplined commitment on the part of data managers and network staff to make it all work.

13.1 Objectives

The objectives for a plan to implement I&M network data management practices and procedures are:

- Ensure that the procedures outlined in this plan and related SOPs are put into practice
- Provide a framework to realistically assess related network procedures and adjust them as needed
- Establish benchmarks so that implementation progress is measured and evaluated

13.2 Laws and Policies

Implementation of an I&M network data management plan follows requirements stemming from federal law, Executive Orders, Director's Orders, or national Inventory and Monitoring Program guidance. The Inventory and Monitoring data management plan is a component of a broader decision support system (NPS-75) required for evaluating the state of national parkland resources. The network level data management plan puts these requirements into context, and provides operational guidance for implementing them.

13.3 General Standards and Guidelines

Implementing a data management plan requires that many related components come together efficiently. Each network data management plan should ensure or accomplish all of the following:

- Data management objectives are clear, quantifiable, achievable, and supportive of the I&M network goals
- Protocols and standard operating procedures are followed so that data are consistent, comparable, secure, and of highest quality
- Data management standard operating procedures illustrate tasks, roles, and responsibilities; are reasonable and achievable, and minimize duplication of effort
- I&M network data management programs provide superior infrastructure and support to I&M network and park staff
- Milestones are set periodically to allow for regular, critical evaluation of objectives and time for planning alternative solutions
- Maintain data management staff expertise in current technology, and provide opportunities for training in new technologies.

13.4 Education and Training

Training and education of both data management staff and network staff will be critical in consistently obtaining high quality data. Data management staff will require solid working knowledge of software, hardware and business processes. Park staff and cooperators must be familiar with the tools, procedures, and guidelines outlined in SOPs, which can be accomplished through formal training sessions and more informal one-on-one assistance.

13.5 Milestones

I&M network data management planning documents and SOPs should be dynamic, with accommodation made for periodic review of progress. The following milestones are suggested for implementing each plan.

13.5.1 Years One through Five:

The following outcomes should be achieved within the first five years of implementing a network data management plan:

- All network staff understand the fundamentals of data management, which include:
    - File and folder management
    - Network shared drive organization
    - Documentation (e.g., metadata formats and process)
    - Quality assurance and quality control
    - Storage, backup, and archiving
- Data management practices are improved by implementing:
    - Conceptual Data Models for Vital Signs monitoring protocols
    - Testing of data entry prior to field work.
- Common standard operating procedures are used for multiple protocols where possible
- Data management staff are directly engaged, in the earliest possible stages, in all aspects of vital signs monitoring protocols

13.5.2 Five Years and Beyond

After five years, the network data management plan and associated SOPs will be fully reviewed and revised, and will incorporate the lessons and practices learned over the first five years. Emphasis will likely increase in the following areas:

- Data reporting, synthesis, and analysis
- Integration of data with other offices, agencies, or programs
- Direct access to data via web services or service-oriented architecture frameworks
- Migration of data to new platforms and applications.

13.6 References and Resources

13.7 Related Network-level SOPs

- File and folder naming conventions
- Documentation
- Quality Assurance/Quality Control

Appendix A. Glossary of Terms

ArcIMS, ArcSDE (Arc Internet Map Server and Arc Spatial Database Engine) Geographic information system (GIS) software applications developed by the Environmental Systems Research Institute (ESRI). These applications provide tools for managing large, complex spatial data sets within a relational database management system (ArcSDE), and serving the data over the Internet (ArcIMS).

Best practice A technique, methodology, or standard operating procedure that, through experience and research, reliably leads to a desired result.

Certified data and metadata Completed data and documentation for short-term projects, or one season of completed data for long-term monitoring projects. Certification is a confirmation by the project leader that 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 proper use and interpretation.

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

Data Distinct measurements or observations of a variable, usually formatted in a special way. They include symbols or representations of facts or ideas that can be communicated, interpreted, or processes by manual or automated means.

Tabular data are usually organized into logical tables of records and fields, arranged in a matrix of rows and columns. Tabular data can be displayed, manipulated, and stored as simple text files or in applications software (e.g., spreadsheets, relational databases).

Spatial data are any data that reference geographic coordinates. GIS data always contain these references; tabular data that contain spatial references are also considered spatial. The terms *GIS data* and *spatial data* are often used interchangeably.

Raw data are data in their original form, i.e., data that have not been altered, summarized, or grouped into broader categories. Raw data can exist in many forms: as hand-written information on field data forms and in notebooks; as unaltered photographs; sound and video recordings; remote sensing imagery; or Global Positioning System (GPS) files.

Derived data are raw data that have been processed, or converted to another form using some automated or manual process. Raw natural resources data are often processed and packaged for summation, statistical analysis, and graphical display, or the production of maps and other information products.

GIS (geographic information system) data contain information about the location and shape of, and relationships among, features on the surface of the earth and are usually stored as

geographic coordinates and topology. Topology is used to compare the geographic locations of features relative to one another (e.g., roads connected to a highway, two vegetation polygons adjacent to one another).

Legacy data are data that are at risk of becoming lost, unusable, or obsolete due to software or metadata limitations.

Programmatic data are data developed specifically by or for a program or project. Programmatic data are usually acquired based on well-defined objectives and specific requirements.

Non-programmatic data are data obtained from an external source, but that are of value to a program or project (e.g., county land parcel information, weather data managed by the National Weather Service, or taxonomic information managed by the U.S. Department of Agriculture).

Sensitive data are data that through loss, unauthorized access, or modification, could be used in such as way as to adversely affect valuable resources, the national interest, the conduct of federal programs, or individual privacy. Examples of sensitive natural resources data might include the locations of rare flora or fauna species, caves, or cultural sites.

Data set A grouping of related data, such that the asemblage of the information will be meaningful to prospective users.

Dataset Catalog A Microsoft Access database application developed by NPS to accommodate spatial and non-spatial metadata. The Dataset Catalog is not fully FCDC-compliant; therefore, has been used primarily for documenting non-spatial data sets. Ongoing development was stopped in 2007 and the database will be replaced as WASO completes the transition of its systems to a service-oriented architecture framework.

Edit log A means of tracking changes to certified data.

Document Recorded information regardless of physical form or characteristics. Often used interchangeably with the term “record.”

Ecological indicators A subset of the physical, chemical, and biological elements and processes of natural systems selected to represent the overall health or condition of the system.

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

Information Created from data as a result of processing, manipulating, synthesizing, or organizing data in a way that provides interpretation or meaning.

Inventories Extensive point-in-time effort to determine location or condition of a resource, including the presence, class, distribution, and status of plants, animals, and abiotic components such as water, soils, landforms, and climate. Inventories contribute to a statement of park resources, which is best described in relation to a standard condition such as the natural or unimpaired state. Inventories may involve both the compilation of existing information and the acquisition of new information. They may be relative either to a particular point in space (synoptic) or time (temporal).

IRMA Integration of Resource Management Applications. The project name for the transition of multiple "silo" applications (e.g., NPSpecies, NatureBib, NPS Data Store) to a service-oriented architecture structure, which is being conducted by the NPS Natural Resource Program Center in Fort Collins.

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

Master database Central repository for project data, used for viewing, summarizing, and analysis. Contains only data that have passed all quality assurance/quality control.

Metadata Information about data. A metadata document contains specific and detailed information about a data set, including who, what, where, when, why and how the data were collected, analyzed, or manipulated. Metadata are considered an essential component of any good data set.

Monitoring The collection and analysis of repeated observations or measurements to evaluate changes in condition and progress toward meeting a management objective. Detection of a change or trend may trigger a management action, or it may generate a new line of inquiry. Monitoring is often done by sampling the same sites over time, and these sites may be a subset of the sites sampled for the initial inventory.

NatureBib The National Park Service bibliographic database, which is used to catalog, search, and manage natural resource-related information sources pertaining to national parks.

National databases and repositories Applications and repositories maintained at the NPS national level, primarily for the purpose of integration among NPS units and for sharing information with cooperators and the public.

Natural Resource Database Template (NRDT) is a core set of database tables that serves as a foundation for building relational databases for the NPS Inventory and Monitoring Program.

Natural Resource Program Center (NRPC) The core of the NPS Natural Resource Stewardship and Science Directorate. NRPC staff are located in Fort Collins and Lakewood, Colorado, and in Washington, D.C. NRPC has five divisions: Air Resources; Biological Resources Management; Environmental Quality; Geologic Resources; and Water Resources. NRPC also includes the Office of Inventory, Monitoring and Evaluation; the Office of Education and Outreach; and the Office of Natural Resource Information Systems.

Normalization The process of reducing a complex data structure into its simplest, most stable structure, thereby eliminating redundant attributes, keys and relationships.

NPS Data Store An online graphical search interface that links dataset metadata to a searchable data server on which datasets are organized by NPS units, offices and programs.

NPS Data Clearinghouse – the central repository for NPS GIS data available to the public, implemented through the NPS Focus gateway.

NPS Focus A decentralized digital imagery and data management system, implemented through a central Internet portal sponsored by the NPS Office of the Chief Information Officer. Includes to the NPS Data Clearinghouse.

NPSpecies The master species database for NPS. The database lists the species that occur in or near each park, and the physical or written evidence for the occurrence of the species (e.g., references, vouchers, and observations). NPSpecies is implemented online through secure and public servers.

NPSTORET The NPS version of the EPA (Environmental Protection Agency) STORET database, used for transferring water quality data collected by NPS programs to STORET. (see STORET)

Physical data model A representation of a data design which takes into account the facilities and constraints of a given database management system. In the lifecycle of a project it is typically derived from a logical data model, though it may be reverse-engineered from a given database implementation.

Protocols Highly detailed, formal documents that explain how data are to be collected, managed, analyzed and reported. Protocols are a key component of consistency and quality assurance for natural resource monitoring programs

Relational Database Management System (RDMS or RDBMS) A system of storing data in related tables, which allows data to be structurally organized for maximum efficiency and a minimum of redundancy.

Research Permit and Reporting System (RPRS) A web-based tool that provides a researcher with applications, procedures, and guidelines for submitting and obtaining a scientific research and collecting permit within NPS units. RPRS also includes Investigator Annual Reports, a report required by all receiving a permit.

Sensitive data Any information which, through loss, unauthorized access, or modification could adversely affect a park resource or program, or the privacy of individuals.

Service-oriented architecture A software architecture that modularizes business functions and allows different applications to easily share and exchange data. SOA separates functions into distinct services, which can be shared, combined, and reused to create applications. NRPC is transitioning many natural resource-related applications to SOA, which is an alternative to previous "silo" application architecture.

Spatial data Information about the location, shape, and relationships among geographic features, usually stored as coordinates and topology. Can be stored in tabular or GIS file format.

Standard Operating Procedures (SOPs) Detailed step-by-step instructions that outline a formal set of procedures for performing specific tasks.

STORET (STORage and RETrieval) A database application maintained by the U.S. Environmental Protection Agency that contains raw biological, chemical and physical data on surface and ground water quality collected by federal, state and local agencies, Indian Tribes, volunteer groups, academics, and others. (see NPSTORET)

Structured Query Language (SQL) A language used to query and process data in a relational database. All database systems designed for client/server environments support SQL.

Vital Signs Ecological elements or processes chosen to represent the overall health or condition of park ecosystems, known or hypothesized effects of environmental stressors, or elements of value to humans, and are the subject of long-term monitoring by the Inventory & Monitoring Program.

Working database – A project-specific database for entering and processing data for the current season (or other logical period of time). May also constitute the master database for short-term projects

Appendix B. List of Acronyms

Acronym	Definition
ANCS+	Automated National Catalog System
CDM	Conceptual Data Model
CESU	Cooperative Ecosystems Studies Unit
CSDGM	Concept Standards for Digital Geospatial Metadata
DMP	Data Management Plan
EPA	U.S. Environmental Protection Agency
ERD	Entity Relationship Diagram
ESRI	Environmental Systems Research Institute
FGDC	Federal Geographic Data Committee
FOIA	Freedom of Information Act
FTP	File Transfer Protocol
GIS	Geographic Information System
GPS	Global Positioning System
HTML	Hypertext Markup Language
I&M	Inventory & Monitoring (Program)
IMS	Internet Map Server
IAR	Investigator's Annual Report
IRMA	Integration of Resource Management Applications
IT	Information Technology
ITIS	Integrated Taxonomic Information System
JPEG	Joint Photographic Experts Group
LAN	Local Area Network
LDM	Logical Data Model
LiDAR	Light Detection and Ranging
MOU	Memorandum of Understanding
mp	Metadata Parser
MS	Microsoft
NBII	National Biological Inventory Infrastructure
NOAA	National Oceanic and Atmospheric Administration
NPS	National Park Service
NRDT	Natural Resource Database Template

Acronym	Definition
NRPC	Natural Resource Program Center
NRTS	Natural Resource Table Standards
OMB	Office of Management and Budget
PC	Personal Computer
PDA	Portable Digital Assistant
PDF	Adobe Portable Document Format
PDM	Physical Data Model
QA/QC	Quality Assurance/Quality Control
RDMS/RDBMS	Relational Database Management System
RFID	Radio Frequency Identification
SDE	Spatial Database Engine
SOA	Service-oriented Architecture
SOP	Standard Operating Procedure
SQL	Structured Query Language
TIFF	Tagged Image File Format
TSN	Taxonomic Serial Number
USGS	U.S. Geological Survey
WAN	Wide Area Network
WASO	Washington Support Office
XML	Extensible Markup Language

Appendix C. Checklist for completing or updating network-level data management plans

The following checklist originated as a series of questions used by peer reviewers to evaluate draft data management plans. The checklist was adapted for network-level data management plans in 2008. Networks should evaluate their plans against the elements in this list to ensure that all are addressed.

#	✓	Description
Organization and Presentation		
1		The document is well-organized, clearly written, and professionally-presented. It has been reviewed by at least two persons other than the author before final submission.
2		The Table of Contents includes a list of tables, a list of figures, and a complete list of appendices or supporting documents.
3		An Executive Summary provides a concise overview of the plan, suitable for a non-technical reader.
4		The plan includes (or provides URLs to) appendices, guidance documents, or SOPs for specific details on procedures or standards referenced in the plan. If all associated materials are not submitted with the plan, it is clear when and where they will be provided.
5		All references mentioned in the plan are included in a Literature Cited section.
6		A map showing the network location and network parks is included in the plan.
7		A revision history log is included as part of the plan document.
8		The scope and purpose of the plan is described, as well as how it functions in conjunction with the national data management plan, the network monitoring plan, and network-specific guidelines and SOPs.
Content		
While the following topics are broadly addressed in the national plan, the network plan should provide greater specificity than the national plan, and describes any items or procedures that diverge from the national plan.		
9		Data management-related staff roles and responsibilities are defined and explained.
10		The network IT infrastructure supporting data management functions is described and clearly presented.
11		The document provides file naming conventions and a network directory structure.
12		There is a clear process for data mining at network parks, and for integrating this information into NPS-wide applications (NatureBib, NPSpecies, NPS Data Store).
13		The acquisition, management, and integration of spatial data are described.

#	✓	Description
14		There is a clear process for managing data resulting from inventory and monitoring projects, and integrating data into NPS-wide applications and programs (e.g., NatureBib, NPSpecies, NPS Data Store, NPS museum management).
15		Procedures for managing photographs (digital and hard-copy) are presented and described.
16		Projects and project management standards and procedures are defined and related to the management of information throughout the project life-cycle.
17		There is a clear process for how the Natural Resource Database Template is used in the development of network-level inventory or monitoring databases.
18		Quality assurance and quality control guidelines are provided for data collection and data entry. It is clear in the plan who is responsible for QA/QC, verification, and validation of network data.
19		The plan specifies data documentation processes and tools that will ensure that metadata are comprehensive and developed in compliance with standards.
20		Version control systems are documented for monitoring protocols, associated databases, and other documents or applications as needed.
21		The general procedures for data analysis and reporting are explained in sufficient detail.
22		Network internet and intranet website structure, content, and management is described and documented.
23		Electronic data storage and backup procedures are clearly described, sufficient to protect all network data, and include provisions for off-site rotation of data.
24		Filing and archive systems are sufficient to protect all network documents, and include integration with park or regional museum management programs.
25		Implementation of the plan is described. Checks and balances are presented that will ensure, in subsequent years, that plan requirements are being met. The procedure for updating the plan is described.

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

NPS D-1968, May 2008.

National Park Service
U.S. Department of the Interior

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

www.nature.nps.gov