



Sonoran Desert Network Data Management Plan

October 2005

Deborah L. Angell
Data Manager

National Park Service
Inventory and Monitoring Program
Sonoran Desert Network
7660 E. Broadway Blvd.
Tucson, Arizona 85710-3776



**SONORAN
DESERT
NETWORK**

Inventory and Monitoring Program

Recommended citation:

Angell, Deborah L. 2005. Sonoran Desert Network Data Management Plan. National Park Service, Inventory and Monitoring Program, Sonoran Desert Network, Tucson, AZ. 80 pp.

Initial distribution:

SODN website: <http://www1.nature.nps.gov/im/units/sodn/data.htm>

Revision history log:

Previous Version	Date of Revision	Author	Changes Made	Reason for Change	New Version Number

Table of Contents

Table of Contents	i
List of Tables	iii
List of Figures	iv
Acknowledgements	v
Executive Summary	vi
Chapter 1. Introduction	1
1.1 Data and Data Management: An Overview	2
1.2 Data Management Goals and Objectives	4
1.3 Sonoran Desert Network Organization and Management	5
1.4 Intended Audience and Layout for the SODN Data Management Plan	6
Chapter 2. Data Stewardship Roles and Responsibilities	8
2.1 Roles and Responsibilities	9
2.2 Data Management Coordination	11
2.3 Project Stewardship	12
Chapter 3. Data Management Resources: Infrastructure and Systems Architecture	16
3.1 Computer Resources Infrastructure	16
3.2 National Information Management Systems	18
3.3 Network Systems Architecture	21
Chapter 4. Data Management Process and Work Flow	25
4.1 Project Work Flow	25
4.2 Data Life Cycle	28
4.3 Integrating and Sharing Data Products	30
Chapter 5. Data Acquisition and Processing	34
5.1 Programmatic Data	34
5.2 Non-Programmatic NPS Data	38
5.3 Non-Programmatic External Data	39
5.4 Data Discovery/Data Mining	40
Chapter 6. Data Quality Assurance / Quality Control	43
6.1 National Park Service Quality Mandate	43
6.2 Quality Assurance and Quality Control Mechanisms	44
6.3 Roles and Responsibilities	45
6.4 Goals and Objectives	46
6.5 Data Collection	47
6.6 Data Entry	48
6.7 Verification and Validation Procedures	50
6.8 Version Control	52
6.9 Data Quality Review and Communication	53
Chapter 7. Data Documentation	56
7.1 Data Set Documentation	56
7.2 Project Documentation	60
Chapter 8. Data Management Support for Analysis and Reporting	61
8.1 Timeline for Analysis and Reporting	61
8.2 Coordination with Project Leaders	61
8.3 Annual Analyses and Reports	62
8.4 Long-Term Analyses and Reports	62
8.5 Special Analyses and Reports	63
Chapter 9. Data Dissemination	64

9.1 Data Ownership.....	64
9.2 Data Distribution	66
9.3 Data Feedback Mechanisms	71
Chapter 10. Data Maintenance, Storage, and Archiving	73
10.1 Digital Data Maintenance.....	73
10.2 Storage and Archiving Procedures for Digital Data.....	75
10.3 Storage and Archiving Procedures for Documents and Objects	77
Chapter 11. References	80

List of Tables

Table 1.1. Categories of data and project products.....	2
Table 2.1. Categories of data stewardship involving all network personnel.....	8
Table 2.2. Summary of roles and responsibilities.....	9
Table 3.1. Groupings for common lookup tables.....	22
Table 4.1. Repositories for Network products.....	31
Table 9.1. Types of data uploaded to web applications/repositories.....	67
Table 10.1. Backup schedule for Network servers.....	76

List of Figures

Figure 1.1. Parks in the Sonoran Desert Network.....	5
Figure 1.2. Layout of the SODN Data Management Plan.	6
Figure 2.1. Shared responsibilities for each project.....	13
Figure 3.1. Schematic representing the logical layout and connectivity of computer resources.	16
Figure 3.2. Model of the national-level application architecture.	19
Figure 3.3. Common lookup tables and satellite databases.	22
Figure 3.4. Different levels of data standards and their corresponding degree of implementation variability.....	23
Figure 4.1. Conceptual model of project work flow.	26
Figure 4.2. Diagram of the typical project data life cycle.....	29
Figure 4.3. Storing and disseminating project information.....	31
Figure 4.4. Steps involved in product distribution.....	32
Figure 4.5. Data flow diagram for water quality data.....	33
Figure 6.1. QA/QC controls applied at progressive stages of a project.....	45
Figure 7.1. Natural Resources Integrated Metadata System (http://science.nature.nps.gov/im/datamgmt/metaplan.html).	58
Figure 10.1. Current directory structure for data files. Note: All directories are not shown.	75

Acknowledgements

Much of the material in this data management plan was developed through the collaboration of a hardworking group of data managers and others. I would like to acknowledge and extend my appreciation to Margaret Beer (Northern Colorado Plateau Network), John Boetsch (North Coast and Cascades Network), Rob Daley (Greater Yellowstone Network), Fred Dieffenbach (Northeast Temperate Network), Patrick Flaherty (Appalachian Highlands Network), Teresa Leibfreid (Cumberland Piedmont Network), Bill Moore (Mammoth Cave National Park), Dorothy Mortenson (Southwest Alaska Network), Velma Potash (Cape Cod National Seashore), Gareth Rowell (Heartland Network), Geoffrey Sanders (National Capital Region Network), Sara Stevens (Northeast Coastal and Barrier Network), Doug Wilder (Central Alaska Network), and Mike Williams (Heartland Network).

A special thank you to John Boetsch, who agreed to take on the tremendous responsibility of coordinating this group. John kept us on the right track throughout this process, and I doubt we would have accomplished as much as we did without his commitment.

Thanks to Lisa Nelson, Joe Gregson, and Steve Fancy for their comments and suggestions on initial drafts of the individual plan chapters.

Finally, I appreciate the contributions network staff and cooperators Andy Hubbard, Theresa Mau-Crimmins, Natasha Kline, Emily Dellinger, Jason Welborn, Bill Halvorson, and Brian Powell.

Executive Summary

Data and information are the basic products of scientific research. In ecological research, where field experiments and data collections can rarely be replicated under identical conditions, data represent a valuable and, often, irreplaceable resource . . . In long-term ecological studies, retention and documentation of high quality data are the foundation upon which the success of the overall project rests.

Brunt 2000

Information is the common currency among the many different activities and people involved in the stewardship of National Park Service (NPS) natural resources. As part of the Service's effort to "improve park management through greater reliance on scientific knowledge," a primary purpose of the Inventory and Monitoring (I&M) Program is to develop, organize, and make available natural resource data and to contribute to the Service's institutional knowledge. The I&M Program's efforts to identify, catalog, organize, structure, archive, and disseminate relevant natural resource information will largely determine the Program's efficacy and image among critics, peers, and advocates.

The NPS is a highly decentralized agency with complex data requirements. The primary audience for many of the products from the I&M Program is at the park level – providing park managers with the information they need to make better-informed decisions and to work more effectively with other agencies and individuals for the benefit of park resources. However, certain data are also needed at the regional or national level for a variety of purposes, and as stated by the National Park Advisory Board (2001), the findings "must be communicated to the public, for it is the broader public that will decide the fate of these resources."

The Sonoran Desert Network (SODN) Data Management Plan presents the overarching strategy for ensuring that I&M Program data are documented, secure, accessible, and useful for decades into the future. The plan refers to other guidance documents, standard operating procedures, and detailed protocols that convey more specific standards and steps for achieving our data management goals. The plan acts as a foundation upon which to build as new protocols are developed, advances in technology are adopted, and new concepts in data management philosophy are accepted.

Data and Data Management: An Overview

Collecting natural resource data is a critical step toward understanding the structure and function of the evolving ecosystems within our National Parks. We analyze and synthesize these 'raw' data to model various aspects of ecosystems. In turn, we use our results and interpretations to make management decisions about the Park's vital natural resources. Thus, *data* collected by researchers and managed by the Sonoran Desert Network according to this Plan is transformed into *information* through analyses, syntheses, and modeling.

Any good set of data – whether collected last week or 20 years ago – must be accompanied by enough explanatory documentation (*e.g.*, how and why it was collected) so that we can understand it and use it with confidence. Therefore, our Network data management system cannot simply attend to the tables, fields, and values that make up a data set. It must also provide a process for developing, preserving, and integrating the context that makes data interpretable and valuable. Although thoroughly documenting a data set is time-intensive, it results in clearer preservation and presentation of the data.

We sometimes use the term 'data' in a broader sense that encompasses other products that are generated alongside primary tabular and spatial data. These products fall into five general categories: raw data,

derived data, documentation, reports, and administrative records. To meet I&M Program goals, and to ensure adequate context for primary data products, these categories of products all require some level of management to ensure their quality and availability. We will use a more ‘holistic view’ about how natural resource data are generated, processed, finalized, and provided. All phases of data and information processing are integrated, and information about each phase and its processes must be shared through good documentation.

There are many potential sources of important data and information about the condition of natural resources in our parks. The types of work that may generate these natural resource data include:

- Inventories
- Monitoring
- Protocol development pilot studies
- Special focus studies done by internal staff, contractors, or cooperators
- External research projects
- Monitoring or research studies done by other agencies on park or adjacent lands
- Resource impact evaluations related to park planning and compliance with regulations
- Resource management and restoration work

Because the I&M Program focuses on long-term monitoring and natural resource inventories, our first priority is to produce and curate high-quality, well-documented data that we derive from these primary efforts. However, we can apply the same standards, procedures, infrastructure, and attitudes about data management to other natural resource data sources. As time and resources permit, we will work toward raising the level of data management for current projects, legacy data, and data originating from outside the I&M Program. We will place the greatest emphasis on those projects that are just beginning development and implementation because applying new data management practices to an ongoing project can be difficult and will generally meet with less success.

Goals and Objectives of Data Management

The data-related mission of the I&M Program is to provide scientifically and statistically sound data to support management decisions for the protection of park resources. The Program’s success at identifying, cataloging, organizing, structuring, archiving, and disseminating relevant natural resource information will largely determine its effectiveness and standing among critics, peers, and advocates. The principal goal of the SODN Data Management Plan is to elucidate the driving concepts, principles, procedures, and processes for ensuring the quality, interpretability, security, longevity, and availability of ecological data and derived information produced by our inventory and monitoring efforts. Our objectives are centered around five main principles:

- *Quality* – ensure that appropriate quality assurance measures are taken during all phases of data development: acquisition, processing, summary and analysis, reporting, documenting, and archiving.
- *Interpretability* – ensure that complete documentation accompanies each data set so that users will be aware of its context, applicability, and limitations.
- *Security* – ensure that both digital and analog data are maintained and archived in a secure environment that provides appropriate levels of access to project leaders, technicians, network staff, and other users.
- *Longevity* – ensure that data sets are maintained in an accessible and interpretable format, accompanied by sufficient documentation.
- *Availability* – ensure that the data and information from our I&M studies are made available and easily accessible to managers and other users.

Data Stewardship Roles and Responsibilities

Everyone within the SODN I&M Program uses or manages data and information, and each of us has our roles and responsibilities in this process. This new and *crucial* emphasis on data management, analysis, and the reporting of results will require a large investment of personnel, time, and money, and the Network expects to invest at least thirty percent of available resources in developing and operating its data management system.

For the SODN I&M Program to work effectively, everyone within the Network will have stewardship responsibilities related to the production, analysis, management, and/or end use of the data.

The fundamental role of the Network data manager will be to coordinate these tasks. This requires understanding and determining program and project requirements, creating and maintaining data management infrastructure and standards, and communicating and working with all responsible individuals.

The data manager and the project leader are the personnel primarily responsible for data management. The network coordinator also assists by ensuring that project leaders meet timelines for data entry, verification, validation, summarization/analysis, and reporting. Figure 1 illustrates the core data management duties of the data manager and project leader and where those duties overlap.

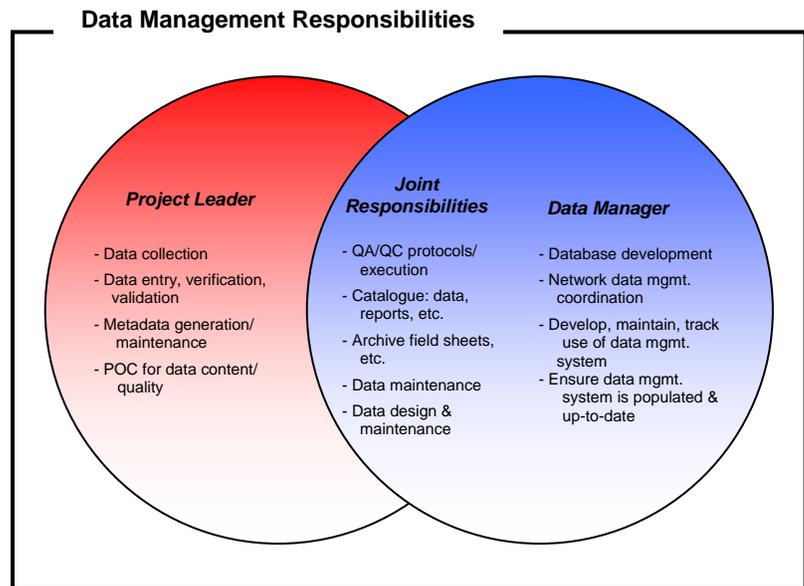


Figure 1. Core data stewardship responsibilities of project leaders and data managers.

Data Management Infrastructure/Architecture

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

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

The national servers host and maintain online applications that provide storage and access to basic natural resource data and information collected by the I&M Program:

- *NatureBib* is the master database for natural resource-related bibliographic references.
- *NPSpecies* is the master database for species that occur in or near each park and the physical or written evidence for their occurrence (*e.g.*, vouchers, observations, references, and data sets).
- *NR-GIS Metadata Database* is the master metadata database for natural resource data sets.
- *NR-GIS Data Store* is a graphical search interface that links data set metadata to a searchable data server on which natural resource data sets are organized by NPS units, offices, and programs.

The network data servers host the following types of data and information:

- *Master project databases* – compiled data sets for monitoring projects and other multi-year efforts that have been certified for data quality.
- *Common lookup tables* – *e.g.*, parks, personnel, projects, species, etc.
- *Network digital library* – network repository for finished versions of products for Network projects (*e.g.*, reports, data set documentation, data files, formal metadata, etc.).
- *GIS files* – base spatial data, imagery, and project-specific themes.
- *Working files* – working databases, draft geospatial themes, draft reports, administrative records, etc.
- *Project tracking application* – database that tracks project status, contact information, product due dates.

Database Design Strategy

Rather than developing a single integrated database system, our approach uses modular, standalone project databases that share design standards and links to common lookup 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 effect on data integration. In addition, one project database can be modified without affecting the functionality of other project databases.
- By working up from modular data sets, we avoid a large initial investment in a centralized database and the concomitant difficulties of integrating among project areas with very different – and often unforeseen – structural requirements. Furthermore, the payoff for this initial investment is not always realized down the road by greater efficiency for interdisciplinary use.

Project database standards ensure compatibility among data sets, which is vital given the often unpredictable ways in which data sets will be aggregated and summarized. Well-thought-out standards also encourage sound database design and facilitate interpretability of data sets. The Network will follow the standards for database objects used by the Natural Resource Database Template (<http://science.nature.nps.gov/im/apps/template/index.htm>), to the extent possible. Databases that are developed for park and network projects will all contain the following main components:

- *Common lookup tables* that contain lists of parks, personnel, and species
- *Core tables and fields based on network and national templates* that contain ‘who, where, and when’ for project data collection
- *Project-specific fields and tables* containing recorded observations

Project Work Flow

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

- *Short-term projects*, which may include individual park research projects, inventories, or pilot studies done in preparation for long-term monitoring.
- *Long-term projects*, which will primarily include the implemented monitoring studies central to the I&M Program, but may also include multi-year research projects and monitoring performed by other agencies and cooperators. Long-term projects often require a higher level of documentation, review, and infrastructure development.

From a data management standpoint, a primary difference between short-term and long-term projects is an increased need to adhere to standards for long-term projects to ensure internal compatibility over time. This does not, however, negate the need to follow standards for short-term projects whenever possible and practical. Furthermore, both short- and long-term projects share many workflow characteristics, and both generate data products that must be managed and made available.

A project can be divided into five primary stages, each characterized by a particular set of activities carried out by staff involved with the project:

- *Planning and approval* – although this phase lacks specific data management activities, data managers must be kept informed of projects in this phase, particularly as timelines for products are finalized.
- *Design and testing* – collaboration between the project leader and the data manager is critical during this phase to assure data quality and integrity.
- *Implementation* – all aspects of this phase are overseen by the project leader; data management staff acts primarily as facilitators to support database applications, GIS, GPS, data validation, summarization, and analysis.
- *Product integration* – data management staff integrates products into national and network databases and repositories; data from working databases are merged into master databases.
- *Evaluation and closure* – the network coordinator, project leader, and data manager work together to assess how well the project met the stated objectives and what steps might be taken to make improvements.

During various phases of a project, the data take on different forms and are maintained in different places as they are acquired, processed, documented, and archived. Key points of this data life cycle are as follows:

- All raw data are archived intact.
- Working databases are the focal point of all modification, processing, and documentation of data collected for a given season (or other period that makes sense for a given project).
- Upon data certification, indicating that the data have passed all documentation and quality assurance requirements, the data are archived and posted or otherwise integrated with the national data applications.
- Data for long-term monitoring projects are uploaded into a master database that includes multiple years of data.
- Certified data sets are used to develop reports and other data products, which are also archived and posted to the appropriate national repositories.
- All subsequent revisions to certified data sets are documented in an edit log, which is distributed with the data.

Data Acquisition & Processing

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

- *Programmatic data* are produced from projects that are either initiated (funded) by or substantially involve the I&M Program.
- *Non-programmatic NPS data* are produced by the NPS but did not involve the I&M Program.
- *Non-programmatic external data* are produced by agencies or institutions other than the NPS.

The importance or value placed on a data set in any of these categories will be based on its quality, completeness, relevance, and potential usefulness of the data set itself, as well as the impact it has on the SODN I&M Program and parks.

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

The Network will conform to NPS standards and mandates, as well as to national I&M Program standards and procedures, to facilitate program integration and data/information sharing. General and protocol-specific SOPs will provide detailed instructions for processing specific types of data.

Quality Assurance/Quality Control

The perception that the data we collect during our inventory and monitoring studies are valuable resources to be used over the long-term is justified only if we have confidence in our data. Our efforts to detect trends and patterns in ecosystem processes require data of documented quality that minimize error and bias. Data of inconsistent or poor quality can result in loss of sensitivity and lead to incorrect interpretations and conclusions. We must remember that high quality data and information are vital to the credibility and success of the I&M Program, and everyone plays a part in ensuring that our products conform to these standards.

NPS Director's Order #11B: Ensuring Quality of Information Disseminated by the National Park Service specifies that information produced by the NPS must be of the highest quality and be based on reliable data sources that are accurate, timely, and representative of the most current information available. Therefore, we will establish and document procedures for quality assurance (QA) and quality control (QC) to identify and reduce the frequency and significance of errors at all stages in the data life cycle. When these procedures are followed, the progression from raw data to verified data to validated data implies increasing confidence in the quality of those data. The data manager will establish SOPs to ensure compliance with DO #11B. These procedures will document both internal and external review processes for data and information disseminated outside the network, as well as guidance for handling complaints about data quality.

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

- Standardized field data collection forms
- Use of field computers and automated data loggers
- Proper calibration and maintenance of equipment

- Field crew and data technician training
- Database features such as built-in pick lists and range limits to reduce data entry errors
- Automated error-checking routines

We evaluate data quality by applying verification and validation procedures. *Data verification* checks that the digitized data match the source data, and *data validation* checks that the data make sense. The Data Management Plan describes several methods for verifying and validating data, and each monitoring protocol will include specific procedures for assuring data quality.

A final report on data quality will be incorporated into the documentation for each project. This will include a listing of the specific methods used to assess data quality and an assessment of overall data quality prepared by the project leader.

Data Documentation

Data documentation is a critical step toward ensuring that all data sets retain their integrity and utility well into the future. Complete, thorough, and accurate documentation should be of the highest priority for long-term studies, and since long-term data sets are continually changing, this documentation must remain up-to-date. Data documentation refers to the development of metadata, which at the most basic level can be defined as ‘data about data,’ or more specifically as information about the content, context, structure, quality, and other characteristics of a data set. Additionally, standardized metadata provide a means to catalog data sets within intranet and internet systems, thus making these data sets available to a broad range of potential users.

Without metadata, potential users of a data set have little or no information regarding the quality, completeness, or manipulations performed on a particular ‘copy’ of a data set. Such ambiguity results in lost productivity as the user must invest time in tracking down information, or, worst case scenario, renders the data set useless because answers to these and other critical questions cannot be found. As such, data documentation must include an upfront investment in planning and organization. At a minimum, we will require the following elements for documentation of all data managed by the Network:

- Data dictionaries and Entity Relationship Diagrams (ERDs) for all tabular databases
- Formal metadata compliant with Federal Geographic Data Committee (FGDC) standards, the National Biological Information Infrastructure (NBII) Profile (where appropriate), and the NPS Metadata Profile for all geospatial and biological data sets
- Project metadata

We will create all metadata according to NPS standards and guidelines. Formal metadata will be created using either Dataset Catalog, an NPS tool for producing abbreviated metadata, or the ArcCatalog data management application included with ArcGIS software, supplemented by the NPS Metadata Tools & Editor. We will publish metadata to the online NR-GIS Metadata Database, and all documentation will also be maintained with its accompanying data set(s) on the network data servers.

Support for Analysis & Reporting

Creating meaningful information from data sets through summaries and analyses is a critical component of the I&M Program and characterizes the Network’s data management mission to provide useful information for park personnel. Close coordination between the project leader and data manager is important to identify opportunities and methods to streamline data extraction and exports from databases

based on project objectives, protocols, and data management and analysis SOPs. Where possible, project databases will include automated summary and report routines.

To make data sets available for subsequent analysis by third parties, the Network will establish a timeline of data processing steps including error-checking, summarizing, analyzing, and distributing data. Monitoring project leaders will be responsible for their project databases, but once a year they will review and certify the data set, write an annual report, and make the data available in a common repository for others to use in syntheses and further analyses.

Data Dissemination

One of the most important goals of the Inventory and Monitoring Program is to *integrate natural resource inventory and monitoring information into National Park Service planning, management, and decision-making*. To accomplish this goal, the Network will use a variety of distribution methods to make data and information collected and developed as part of the Program available to a wide community of users, including park staff, other researchers and scientists, and the public. We will ensure that:

- Data are easily discoverable and obtainable.
- Distributed data are accompanied by complete metadata that clearly establishes the data as a product of the NPS I&M Program.
- Data that have not yet been subjected to full quality control will not be released by the Network, unless necessary in response to a FOIA request *or* unless accompanied by a data quality disclaimer.
- Sensitive data are identified and protected from unauthorized access and inappropriate use.
- A complete record of data distribution/dissemination is maintained.

Distribution options include the Network data server and digital libraries, along with several online interfaces. The national I&M Program has developed several web-based applications and repositories to store different types of natural resource information:

- *NPSpecies* – data on park biodiversity (species information)
- *NatureBib* – park-related scientific citations
- *Biodiversity Data Store* – raw or manipulated data products that document the presence/absence, distribution, and/or abundance of any taxa in NPS units
- *NR-GIS Metadata and Data Store* – spatial and non-spatial metadata and accompanying data sets
- *Sonoran Desert Network website* – reports and metadata for all I&M data produced by the Network

Data Ownership

The NPS defines conditions for the ownership and sharing of collections, data, and results based on research funded by the United States government. All contracts and cooperative or interagency agreements should include clear provisions for data ownership and sharing as defined by the NPS:

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

As such, the Network has established guidelines for ensuring its ownership of data and other research information.

FOIA and Sensitive Data

The Freedom of Information Act (FOIA) stipulates that federal agencies, including the NPS, must provide access to agency records that are not protected from disclosure by exemptions. The NPS is directed to protect information about the nature and location of sensitive park resources under one Executive Order and four resource confidentiality laws:

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

When any of these regulations are applicable, public access to data can be restricted. If disclosure could result in harm, information about the following natural resources may be classified as ‘protected’ or ‘sensitive’ and information withheld:

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

The Network will comply with all FOIA restrictions regarding the release of data and information, as instructed in NPS Director’s Order #66 and accompanying Reference Manuals 66A and 66B (currently in draft). Classification of sensitive data will be the responsibility of Network staff, park superintendents, and project leaders. Network staff will classify sensitive data on a case-by-case, project-by-project basis and will work closely with project leaders to ensure that potentially sensitive park resources are identified, that information about these resources is tracked throughout the project, and that potentially sensitive information is removed from documents and products that will be released outside the Network.

Data Maintenance, Storage, and Archiving

Data, documents, and any other products that result from projects and activities that use Network data are all crucial pieces of information. Directions for managing these materials are provided in NPS Director’s Order #19: Records Management (2001) and the accompanying NPS Records Disposition Schedule (NPS-19 Appendix B, revised 5-2003). This guidance states that records of natural and cultural resources are considered ‘mission-critical’ records (permanent records that are to be transferred to the National Archives when 30 years old) and that copies of these materials “should not, in any instance, be destroyed.”

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

- Kept up-to-date in content and format so they remain easily accessible and usable

- Protected from catastrophic events (*e.g.*, fire and flood), user error, hardware failure, software failure or corruption, security breaches, and vandalism

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

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