Southern Colorado Plateau
Inventory and Monitoring Network

National Park Service
U.S. Department of the Interior

Natural Resource Program Center
Fort Collins, Colorado

Data Management Plan for the Southern Colorado Plateau Network

Nicole J. Tancreto and Marguerite N. Hendrie

National Park Service, Southern Colorado Plateau Network,
Northern Arizona University, P.O. Box 5765, Flagstaff, AZ 86011

October 2006

Data Management Plan for the Southern Colorado Plateau Network

Nicole J. Tancreto and Marguerite N. Hendrie

National Park Service, Southern Colorado Plateau Network,
Northern Arizona University, P.O. Box 5765, Flagstaff, AZ 86011

October 2006

Recommended Citation: N.J. Tancreto and M.N. Hendrie. 2006. Data management plan for the Southern Colorado Plateau Network. National Park Service, Flagstaff, AZ. 77 p. plus appendices.

U.S. Department of the Interior
National Park Service, Southern Colorado Plateau Network
Northern Arizona University
P.O. Box 5765
Flagstaff, AZ 86011

Acknowledgements

This plan was written with considerable help, inspiration and ideas gleaned from data management plans written and developed by other data managers throughout the NPS Inventory & Monitoring program.

EXECUTIVE SUMMARY **1**

CHAPTER 1 INTRODUCTION **6**

1.1 OVERVIEW AND SCOPE 6
1.2 PLAN REVISIONS 6
1.3 GOALS AND OBJECTIVES 7
1.4 NETWORK ORGANIZATION AND MANAGEMENT 7
1.5 SOURCES OF NATURAL RESOURCE DATA 9

CHAPTER 2 ROLES AND RESPONSIBILITIES **10**

2.1 DATA MANAGEMENT ROLES 11
2.2 PERSONNEL RESOURCES AND RESPONSIBILITIES 12
2.3 DATA MANAGEMENT COORDINATION 14

CHAPTER 3 INFRASTRUCTURE & SYSTEM ARCHITECTURE **16**

3.1 INFRASTRUCTURE 16
3.2 SCPN SYSTEM ARCHITECTURE 17
3.3 LAN AND WAN INFRASTRUCTURE 19
3.4 IT NETWORK SECURITY 19

CHAPTER 4 DATA MANAGEMENT PROCESS AND WORKFLOW **21**

4.1 TYPES OF DATA AND INFORMATION 21
4.2 DATA MANAGEMENT FRAMEWORK 21
4.3 PHASES OF DATA PROCESSING 23
4.4 WATER QUALITY DATA 24

CHAPTER 5 DATA DESIGN **25**

5.1 DATA MODELING 25
5.2 DATABASE STANDARDS 25
5.3 DATA SHEET DESIGN 26
5.4 DATABASE APPLICATION DESIGN 26

CHAPTER 6 DATA ACQUISITION AND VERIFICATION **28**

6.1 NETWORK DATA 28
6.2 NPS DATA 30
6.3 EXTERNAL DATA 32

CHAPTER 7 DATA VALIDATION, SUMMARY AND EXPORT FOR ANALYSIS **33**

7.1 DATA VALIDATION 33
7.2 DATA SUMMARIES 33
7.3 EXPORT FOR ANALYSIS 34

CHAPTER 8 DOCUMENTATION **35**

8.1 METADATA 35
8.2 DATA QUALITY DOCUMENTATION 36
8.3 DATABASE APPLICATION DOCUMENTATION 36

CHAPTER 9 DATA OWNERSHIP AND SENSITIVITY **38**

9.1 DATA OWNERSHIP 38
9.2 DATA SENSITIVITY 39

CHAPTER 10 DISSEMINATION AND ACCESS **41**

10.1 DISTRIBUTION MECHANISMS 41
10.2 NPS ACCESS AND PRODUCTS 42
10.3 PUBLIC ACCESS AND PRODUCTS 42

10.4 TRACKING AND FEEDBACK 43

CHAPTER 11 ARCHIVING AND STORAGE OF DIGITAL PRODUCTS 44

11.1 NPS AND NETWORK STANDARDS 44
11.2 ELECTRONIC ARCHIVING AND STORAGE 44
11.3 ELECTRONIC ARCHIVING PROCESS AND WORKFLOW 44
11.4 QUALITY CONTROL FOR CONVERTED DATA 44
11.5 DIGITAL DATA BACKUPS 45

CHAPTER 12 ARCHIVING AND STORAGE OF PHYSICAL PRODUCTS 46

12.1 NPS STANDARDS 46
12.2 SCPN STANDARDS 46
12.3 ARCHIVING PROCESS 46
12.4 NETWORK WORKING FILES AND LIBRARY 47

LIST OF TABLES

TABLE 1-1. ESTABLISHMENT PURPOSE AND SIZE OF SOUTHERN COLORADO PLATEAU NETWORK PARK UNITS....8
TABLE 1-2. SCPN STAFF, DATA MANAGEMENT ROLE AND POSITION TYPE. 9
TABLE 2-1. DATA MANAGEMENT TASKS AND THE POSITIONS WITH PRINCIPAL RESPONSIBILITY. 10
TABLE 2-2. ROLES AND RESPONSIBILITIES FOR DATA STEWARDSHIP 11
TABLE 3-1. SCPN INFRASTRUCTURE SERVICE OR SUPPORT PROVIDERS. 16
TABLE 3-2. GENERAL DIRECTORY STRUCTURE AND THEIR ROLES FOR MASTER LIBRARIES. 19
TABLE 4-1. CATEGORIES OF DATA 21

LIST OF FIGURES

FIGURE 1-1. OVERVIEW OF SOUTHERN COLORADO PLATEAU NETWORK PARK UNIT LOCATIONS 7
FIGURE 2-1. DATA STEWARDSHIP RESPONSIBILITIES OF THE CORE DATA MANAGEMENT TEAM. 12
FIGURE 3-1. SCPN LOCAL, AND NATIONAL INFRASTRUCTURE. 17
FIGURE 3-2. DIRECTORY ORGANIZATION OF WORKING FILES. 18
FIGURE 4-1. DATA MANAGEMENT ACTIVITIES WITHIN THE CONTEXT OF A MONITORING PROJECT. 22

APPENDICES 48

APPENDIX A. DATA MANAGEMENT PLAN REVISION LOG 49
APPENDIX B. BACKUP PROCEDURES 50
APPENDIX C. OPERATING PROCEDURES FOR UPLOADING AND DOWNLOADING DATA TO AND FROM GPS UNITS (DRAFT) 55
APPENDIX D. GIS DATA ORGANIZATION AND STANDARD OPERATING PROCEDURES FOR DATA AND METADATA CREATION (DRAFT) 59
APPENDIX E. DIGITAL PHOTO MANAGEMENT (DRAFT) 62
APPENDIX F. FILE NAMING CONVENTIONS (TO BE DEVELOPED)....67
APPENDIX G. DATA MODELING 68
APPENDIX H. STANDARD DATA SUMMARY REPORTS (TO BE DEVELOPED)....71
APPENDIX I. EXPORTING DATA FOR ANALYSIS (TO BE DEVELOPED) 72
APPENDIX J. METADATA CREATION FOR NON-SPATIAL DATA SETS 73

Executive Summary

Data become information through the process of analysis, synthesis, modeling, or other types of interpretation. Data management provides a means for organizing, documenting and archiving data so that the original information potential is maintained through time. This is particularly important for long-term programs where the lifespan of a data set will likely be longer than the careers of those who developed it. A data management system that can effectively produce, maintain and distribute monitoring results is central to the success of the Inventory & Monitoring (I&M) Program.

This data management plan outlines the general data management standards; expected roles and responsibilities; and data processing, storage and distribution guidelines for the Southern Colorado Plateau Network (SCPN). Detailed data management procedures for specific SCPN monitoring projects will be based on these general guidelines.

Goals and Objectives

The SCPN approach to data management is user oriented, the user ranging from SCPN staff to park managers. The primary goal of the SCPN data management program is to ensure the quality, clarity, security, longevity and availability of SCPN I&M data.

Quality – SCPN I&M data will be used by park staff to inform management decisions regarding park natural resources; it is essential that these data be accurate and complete. Appropriate quality assurance measures will be employed throughout the process of collecting, processing and maintaining data. Good data stewardship habits and attitudes will be encouraged.

Clarity – Confusing and cryptic data sets are of little use and can be easily misinterpreted. All data and information products will be accompanied by complete documentation so that users will be aware of the applicability and limitations of the data.

Security – All information products will be maintained so that appropriate levels of access are provided to SCPN and park staff. Existing technologies will be utilized to protect I&M data from corruption or loss, ensuring the long-term security and integrity of the data.

Longevity – Many factors combine to increase the longevity of a data set: proper documentation, organization, and standardization to modern technologies. SCPN will ensure that all data sets are completely documented. Data sets will be organized in a logical and consistent manner so that nothing is lost over time. As software and hardware technology changes, data sets will be updated so that they remain readable and accessible using new technology.

Availability – I&M data can only be useful to park managers if it is easily available in a timely manner and in a useful form. Information products will be distributed to park management on a regular schedule and, when appropriate, will be made available to a broader audience.

Sources of Natural Resource Data

The existence of numerous potential sources of ecological data about park natural resources requires SCPN to make some priorities about managing data. Some sources of natural resource data include:

- Inventories
- Monitoring
- Protocol development studies
- Special focus studies completed by parks
- External research projects
- Studies by other land management agencies on adjacent lands
- Resource impact evaluations related to park planning and compliance regulations
- Resource management and restoration work

SCPN will be able to maintain the highest level of control over data collected through the I&M Program and the data management program will focus on these data. However, the goal is to apply the same standards, procedures and attitudes about data management to other sources of natural resource data over the long-term, working toward raising the level of data management for projects originating outside the I&M Program.

Roles and Responsibilities

Data management and stewardship is the responsibility of all participants in SCPN Network I&M activities; it requires true collaboration among many people with a broad range of tasks and responsibilities. Good habits and attitudes are as important as standards and procedures. Although primary responsibility resides with the data manager, GIS specialist, and project managers who make up the core data management team, all SCPN staff and cooperators are responsible for ensuring data stewardship is practiced throughout the life of a monitoring project.

Infrastructure and System Architecture

Management and dissemination of monitoring data is made possible by information technology infrastructure and system architecture. Infrastructure refers to the network of computers and servers that information systems are built upon. System architecture refers to the application, database system, repositories, and software tools that make up the framework of the SCPN's data management enterprise.

SCPN relies on cooperative agreements with Northern Arizona University (NAU) and the USGS Colorado Plateau Research Station (CPRS) as well as NPS regional and national information technology personnel and resources for maintenance and support of computer and networking infrastructure.

SCPN System Architecture

Working files, master libraries and digital archives will be stored on SCPN file and data servers. A template project directory for databases, files, and project documentation will be used for each monitoring project. This directory will contain working files for which all project team members will have read/write access. Master libraries are repositories for final information products and certified databases. Master libraries will be organized according to type – databases will be stored together, documents will be stored together, etc. – and only certain SCPN personnel will have full access. Digital archives will be repositories for packages of data and information that can easily be redistributed and will have extremely limited write access.

Rather than developing a single integrated database system, SCPN will develop stand-alone project databases that share design standards and centralized lookup tables for data shared across projects. These modular databases allow for greater flexibility to accommodate each project's needs and sufficient standardization can ensure the ability to aggregate and summarize data across multiple projects. SCPN currently uses Microsoft Access for all project databases and is investigating the need to move to a client-server relational database management system such as Microsoft SQL Server.

National System Architecture

The national I&M Program provides several repositories for hosting SCPN information products and applications for summarizing park data at a national level. The applications are available online and allow users to access basic natural resource information for SCPN parks:

- *NatureBib* – master database for natural resource bibliographic references
- *NPSpecies* – master database for species occurrence records and evidence (voucher specimens, references, observations or data sets) at each park
- *NR-GIS Metadata and Data Store* – master database of metadata for GIS and natural resource data sets and a repository for that data

Data Management Process and Workflow

Within the context of a monitoring project, SCPN data management tasks can be divided into several types of activities. These activities provide the backbone for planning and executing data management procedures for each monitoring project. Specific procedures and guidelines for each of these activities are explained in the data management plan. Data design refers to the design and development of project data sheets, the database and database application. During data acquisition, data are collected in the field or acquired from other sources, entered into the project database, and verified. Data will then be validated to ensure they are within normal ranges, summarized and exported for analysis. Documentation will be completed, data and information products will be distributed, and both digital and analog products will be archived.

Data and documentation take different forms and are maintained in different places throughout the phases of a project. These phases can be modeled as a sequence of events and tasks which involve interaction with the following items:

- *Raw data* – Analog data recorded by hand on field data sheets and digital files from handheld computers, GPS receivers, telemetry data loggers, etc.
- *Working database* – A project-specific database for entering and processing data for the current season (or other logical time period). This may be the only database for short-term projects with no need to distinguish current season data from the full set of validated data.
- *Certified data and metadata* – Completed data and documentation for short-term projects, or one season of completed data for long-term monitoring projects. Certification is a confirmation by the project manager that the data have passed all quality assurance requirements and are complete and ready for distribution. Metadata records include detailed information about project data needed for proper use and interpretation.
- *Master database* – Project-specific database for storing the full set of validated project data, used for viewing, summarizing, and analysis. Current season data from the working database must pass all quality assurance steps prior to upload into this master project database.
- *Reports and data products* – Information that is derived from certified project data.
- *Edit log* – A means of tracking changes to certified data.
- *National databases and repositories* – Applications and repositories maintained at the national level, primarily for the purpose of integration among NPS units and for sharing information with cooperators and the public.
- *Archived data* – Digital and hard-copy data stored and maintained for the long-term

Water Quality Data

Water quality data collected as part of the network's monitoring program have distinct data management requirements. Data must be managed according to guidelines from the NPS Water Resources Division (WRD; http://www.nature.nps.gov/water/infoanddata/index.cfm). NPS WRD has selected the Environmental Protection Agency's national STOrage and RETrieval water quality database (STORET) as the standard for archiving NPS water quality data. WRD maintains a copy of STORET and requires that all I&M chemical, physical and biological water quality data be archived in this copy. SCPN will use the Microsoft Access database application (called NPSTORET) developed by WRD to fulfill this requirement.

The USGS/WRD-Colorado developed a water quality database that enables assessment of the temporal and spatial distribution characteristics of water quality data available for the 19 SCPN park units. For the purpose of developing the Water Quality Vital Signs Monitoring Plan, this database provides a useful tool for evaluation of historical water quality of the waters in and surrounding SCPN park units, and determination of historical and current water quality conditions in and near these parks. SCPN plans to maintain and regularly update the water quality database through data retrievals from NPSTORET or STORET, enabling it to serve as a dynamic tool for the network's long-term water quality monitoring and analysis needs.

Data Design

The data manager and project manager will collaborate on design of field data sheets, database structure and database application for each monitoring project. Databases will be standardized where possible following I&M recommended guidelines for database structure and naming conventions developed in the Natural Resource Database Template (NRDT) and the Recommended Naming Standards. SCPN will also develop standardized lookup tables for data elements shared across many monitoring projects. Database design will also be guided by a data modeling process involving the creation of three types of data models: conceptual, logical and physical.

Data Acquisition and Quality Control

Data managed and utilized by the network will originate from three types of sources: within the network, other NPS data collection efforts, and outside the NPS altogether.

- **Network Data** – any data produced from projects that are initiated (funded) by the SCPN I&M Program or projects that in some way involve the I&M Program.
- **NPS Data** – any data produced by the NPS that did not involve the SCPN I&M Program.
- **External Data** – any data produced by agencies or institutions other than the National Park Service.

SCPN staff are responsible for the acquisition and quality control of network data. Project crew leaders and members are primarily responsible for data collection, data entry and verification of data acquired from field data collection. Each monitoring project protocol will detail procedures for these data acquisition steps based on guidelines outlined in this plan. As data are collected and entered into a database, quality control procedures will be used to increase accuracy and limit transcription mistakes. A verification procedure will be used to check for and correct any transcription mistakes.

NPS and external data will be acquired only with complete documentation and metadata. These data will undergo limited processing and quality control by SCPN staff to ensure compatibility with SCPN project databases, if necessary. Legacy data from parks will be evaluated and prioritized for digitizing or converting to modern database formats. Some external ancillary data, such as climate data, will be acquired when needed in subsets rather than stored by SCPN in its entirety.

Quality Assurance, Data Summarization, and Export for Analysis

Quality assurance, data summary and data analysis are the responsibility of the project managers; however, the data manager will provide tools to project managers to facilitate these three activities. Data validation procedures (ensuring measures are logical and within normal ranges) will be detailed in each monitoring protocol and will generally include outlier detection and other exploratory analyses.

Routine data summaries will be produced after data have been verified on a schedule specific to each project. Summaries will generally be automated within the database application, but park-specific data reports can be produced for management needs. Automated exports will also be included in each database application to enable project managers to export subsets of data in a format ready for import into specific statistics or other analytic software programs.

Documentation

Data set documentation is the responsibility of the project manager and data manager. All data sets will be documented with formal metadata, using Federal Geographic Data Committee and USGS National Biological Information Infrastructure standards. Documentation accompanying database applications will include a manual with instructions for using the application, an entity relationship diagram, a data dictionary and programming code documentation.

Access and Archiving

Data Ownership and Sensitivity

SCPN data products are owned by the National Park Service provided under OMB Circular A-110, Section 36. The Freedom of Information Act (FOIA) establishes that the federal government, the NPS

included, must provide access to non-protected data and information of interest to the public through reading rooms or the Internet.

The NPS is directed to protect information about the nature and location of sensitive park resources under one Executive Order and four resource confidentiality laws:

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

All monitoring information products will be vetted for sensitive data prior to making them available to the general public. Classification of sensitive I&M data will be a shared responsibility that includes network staff, park resource management staff, park superintendents, and investigators working on individual projects. Park management has ultimate responsibility for deciding which information is sensitive and should not be released to the public. The network has ultimate responsibility for ensuring that sensitive data are not released to the public.

Dissemination and Access

Dissemination of monitoring and information products from SCPN will follow these guidelines:

- data will be easily located and acquired
- only data subjected to full quality control and quality assurance measures will be released
- data will be accompanied by complete metadata
- sensitive data will be identified and protected from unauthorized access

Information products will be made available primarily through websites and clearinghouses which will allow users to search for and download reports, summarized data, maps and metadata and other associated information. Distribution means will include (but may not be limited to):

- SCPN public website
- NR-GIS Metadata and Data Store
- Service-wide databases, such as NPSTORET, NPSpecies, and NatureBib
- Regional, Network, or Park data servers protected with read-only access
- FTP sites, CDs, DVDs, or hard drives, as appropriate

Archiving and Storage

Digital and analog information products will be stored, archived and maintained in a variety of repositories. Digital products resulting from monitoring projects will be archived on SCPN file servers and national file and data servers and protected from catastrophic loss by regular, automated backups to external media. Analog products will be archived to NPS standards by individual park facilities or approved non-NPS institutions. At the termination of a project or at regular milestones, an archival package will be prepared and delivered to the desired location.