

National Park Service
U.S. Department of Interior



Data Management Plan Southwest Alaska Network Inventory and Monitoring Program September 2006



Dorothy C. Mortenson

**Data Management Plan
Southwest Alaska Network
Inventory and Monitoring Program**

Dorothy C. Mortenson
National Park Service
Southwest Alaska Network
Inventory & Monitoring Program
240 West 5th Avenue #114
Anchorage, Alaska 99501

Version: September 2006

File Name: SWAN_2006_DataManagementPlan_0609.doc

Recommended Citation:

Southwest Alaska Network. 2006. Data management plan for the Inventory and Monitoring Program, Southwest Alaska Network. National Park Service. Anchorage, AK.

Topic(s):

Administrative, Interdisciplinary, management

Theme Keywords:

Reports, protocols, guidelines, data management plan, Inventory & Monitoring Program

Placename Keywords:

Alaska, Southwest Alaska, Katmai National Park and Preserve, Alagnak Wild River, Kenai Fjords National Park, Lake Clark National Park and Preserve, Aniakchak National Monument and Preserve, Alaska Peninsula, Kenai Peninsula

Acronyms:

AGDC	Alaska Geographic Data Committee	I&M	Inventory & Monitoring (Program)
ALAG	Alagnak Wild River	IT	Information Technology
AKRO	Alaska Regional Office	ITIS	Integrated Taxonomic Information System
ANCS+	Automated National Catalog System	KATM	Katmai National Park and Preserve
ANIA	Aniakchak National Monument and Preserve	KEFJ	Kenai Fjords National Park and Preserve
ARCC	Alaska Regional Curitorial Center	LACL	Lake Clark National Park and Preserve
ARLIS	Alaska Resource Library and Information Services	LDM	Logical Data Model
BOFAT	Buffet of Fields and Tables	NRDT	Natural Resource Database Template
CDM	Conceptual Data Model	NPS	National Park Service
CSDGM	Content Standard for Digital Geospatial Metadata	PDF	Adobe Portable Document Format
DMP	Data Management Plan	PDM	Physical Data Model
FGDC	Federal Geographic Data Committee	QA/QC	Quality assure/quality control
FOIA	Freedom of Information Act	SWAN	Southwest Alaska Network
GIS	Geographic Information System	SOP	Standard operating procedure
		WASO	Washington Support Office (Ft. Collins)

Initial Distribution:

Website:

<http://www.nature.nps.gov/im/units/swan/>

Katmai National Park and Preserve
Kenai Fjords National Park
Lake Clark National Park and Preserve
Alaska Regional Office, Inventory & Monitoring Program
Washington Support Office, Inventory & Monitoring Program

Cover Photograph:

Cover photograph is of Hallo Glacier on the coast of Katmai National Park and Preserve. Photo was taken June 2005 by Alan Bennett, SWAN Program Coordinator.

Plan Revisions Log:

Date:	Description of revision:
9/07/2006	Revisions made throughout DMP based on comments received from NPS peer reviewers.
12/15/2005	Final preparations for review. Formatting revisions made.
9/15/2005	Revisions made throughout DMP based on comments received from Network staff review.

Table of Contents

Executive Summary

1.	Introduction	1-1
1.1.	The Inventory and Monitoring Program	1-1
1.2.	Scope of the Data Management Plan	1-1
1.3.	Data Management Goals	1-2
1.4.	Audience	1-3
1.5.	Revisions	1-3
1.6.	Types of data covered by this Plan	1-3
1.7.	Sources and Priorities of Natural Resource Data	1-5
1.8.	Inventory and Monitoring Program Organization	1-5
1.8.1.	National and Regional Organization	1-5
1.8.2.	Southwest Alaska Network	1-6
1.9.	Organization of this Plan	1-7
1.10.	Chapter Credits	1-8
1.11.	References	1-8
2.	Data Management Roles and Responsibilities	2-1
2.1.	Data Stewardship Roles	2-1
2.2.	Data Management Team	2-4
2.2.1	Network Coordinator	2-5
2.2.2.	Project Leader	2-6
2.2.3.	Data Manager	2-6
2.2.4.	AKRO GIS Specialist	2-6
2.3.	Data Management Coordination	2-6
2.3.1.	National Park Service	2-7
2.3.2.	Alaska Data Coordination	2-8
2.4.	Chapter Credits	2-8
3.	Data Management Process and Workflow	3-1
3.1.	Data Life Cycle and Workflow	3-1
3.2.	Review Cycle	3-7
3.3.	Chapter Credits	3-7
4.	Infrastructure and System Architecture	4-1
4.1.	Infrastructure	4-2
4.1.1.	Computer Server Infrastructure	4-2
4.1.2.	Clearinghouse Infrastructure	4-4
4.2.	SWAN System Architecture	4-5
4.2.1.	Team and User Project Files	4-5
4.2.2.	Integration of Products to Master Data Libraries	4-5
4.2.3.	Integration of Long-Term Databases	4-6
4.2.4.	Application Development	4-6
4.2.5.	Internet Services	4-7
4.3.	LAN and WAN Infrastructure	4-7
4.4.	IT Network Security	4-7
4.5.	Chapter Credits	4-8
4.6.	References	4-8
5.	Database Design	5-1
5.1.	National Standards	5-1
5.2.	Network Standards	5-1
5.3.	Introduction to Data Models	5-2
5.3.1.	Conceptual Data Models	5-2

5.3.2.	Logical Data Models	5-2
5.3.3.	Physical Data Models	5-3
5.3.4.	Developing the Models	5-4
5.4.	Fundamental Database Structure	5-7
5.5.	Integration of Data Dictionary	5-7
5.6.	Documenting Databases	5-7
5.7.	Chapter Credits	5-7
5.8.	References	5-7
6.	Data Acquisition and Processing	6-1
6.1.	National Standards	6-1
6.2.	Network Standards	6-1
6.3.	Data Sources	6-1
6.4.	Data Discovery	6-2
6.5.	Data Harvest	6-7
6.6.	Data Collection for Field Studies	6-7
6.7.	Data Collection for Remote Sensing	6-10
6.8.	Changes to Data Collection Procedures	6-12
6.9.	Data Compilation, Processing and Integration	6-12
6.10.	Chapter Credits	6-12
7.	Quality Assurance/Quality Control	7-1
7.1.	NPS Standards	7-1
7.2.	Quality Assurance and Quality Control Mechanisms	7-3
7.3.	Roles and Responsibilities	7-3
7.4.	Goals and Objectives	7-4
7.5.	Data Collection	7-5
7.5.1.	Methods for Reducing Collection Errors	7-5
7.6.	Data Entry	7-7
7.6.1.	Methods for Reducing Data Entry Errors	7-8
7.7.	Verification and Validation Procedures	7-9
7.7.1.	Methods for Data Verification	7-10
7.7.2.	Methods for Data Validation	7-11
7.8.	Version Control	7-13
7.9.	Data Quality Review and Communication	7-13
7.9.1.	Value of Feedback from QA/QC Procedures	7-13
7.9.2.	Monitoring Conformance to Plans and Standards	7-14
7.9.3.	Communicating Data Quality	7-14
7.10.	Chapter Credits	7-15
7.11.	References	7-14
8.	Data Documentation	8-1
8.1.	Federal and NPS Standards	8-1
8.2.	Network Standards	8-1
8.3.	Project Documentation	8-2
8.3.1.	Project Organizer	8-2
8.3.2.	Project Tracking Database	8-2
8.4.	Spatial Data Documentation	8-2
8.5.	Tabular Database Documentation	8-2
8.6.	Programming Documentation	8-3
8.6.	Vital Signs Protocols Documentation	8-3
8.6.1.	Master Version Table	8-3
8.6.2.	Additional Vital Signs Metadata	8-3
8.7.	Chapter Credits	8-4
8.7.	References	8-4

9.	Data Summary and Analysis	9-1
9.1.	Periodic and Annual Reporting	9-1
9.2.	Long-term Trends and Analysis	9-1
9.3.	Chapter Credits	9-2
10.	Data Dissemination	10-1
10.1.	Mechanism for Distribution	10-1
10.2.	Ownership, FOIA, and Sensitive Data	10-2
10.2.1.	Access Restrictions on Sensitive Data	10-4
10.2.2.	NPS Only vs. Public	10-5
10.3.	Feedback Mechanisms	10-5
10.4.	Chapter Credits	10-5
11.	Data Maintenance, Storage and Archiving	11-1
11.1.	NPS and Network Standards	11-1
11.2.	Electronic Archiving and Storage	11-1
11.3.	Electronic Archiving Process and Workflow	11-1
11.4.	Quality Control for Converted Data	11-3
11.5.	Digital Data Backups	11-3
11.5.1.	Project Backups	11-3
11.5.2.	Data Backup for Catastrophic Event Protection	11-4
11.6.	Chapter Credits	11-5
11.7.	References	11-5
12.	Natural History Archive, Records Management	12-1
12.1.	NPS Standards	12-1
12.2.	Park Unit Standards	12-2
12.3.	Network Standards	12-2
12.4.	Role of curators in curation and museum collection storage	12-2
12.5.	Workflow	12-3
12.6.	Network Archiving Process	12-4
12.7.	Archive and Storage Locations	12-5
12.7.1.	Alaska Regional Curatorial Center (ARCC)	12-5
12.7.2.	Lake Clark/Katmai Study Center (LAKA)	12-5
12.7.3.	Kenai Fjords National Park (KEFJ) Curation Facility	12-5
12.7.4.	Non-NPS Facilities	12-6
12.7.5.	Network Working Files and Library	12-6
12.7.6.	Park Unit Working Files and Library	12-6
12.7.7.	Public Resource Libraries	12-6
12.8.	NPS Automated National Catalog System (ANCS+)	12-7
12.9.	Photographs	12-7
12.10.	Specimens	12-8
12.10.1.	Duplicate Vouchers	12-8
12.11.	Other Materials	12-8
12.12.	Chapter Credits	12-8
12.13.	References	12-9
12.14.	Additional References	12-9
13.	Implementation	13-1
13.1.	Common Data Management Sins	13-1
13.2.	Years 1-5 Implementation	13-2
13.3.	Years 5+ Implementation	13-3
13.4.	Chapter Credits	13-3

Appendices

The first number of the appendix correlates with chapters.

- 3-1** Information Deliverables Specifications for Research Southwest Alaska Network: Deliverable Checklist and Description
A template for determining deliverables and format requirements. To be used at the start of each project.
- 4-1** NPS, Program, and Network Clearinghouses
A listing and description of Clearinghouses used by the Network
- 4-2** Cheat Sheet: File Naming Standards
Provides a shortened list of file naming standards.
- 5-1** Microsoft Access Database Application Specifications
Provides specifications for application development in Access.
- 7-1** Standard Operating Procedure: Data Audit Reporting
Provides a list of data management components to check when auditing and includes an example of a Data Audit Report.
- 8-1** Project Organizer
A word document template for documenting the files for a project.
- 8-2** Transfer of Natural Resource History Files to Archives
An example of a report of the information that can be generated from the Project Tracking Database.
- 8-3** Monitoring Protocol Database Data Dictionary
A template for database documentation
- 12-1** Study Plan Agreement Template: Curation of Natural History Specimens
- 12-2** Natural History Specimen Collections, Curatorial Responsibilities for National Park Service Collectors
- 12-3** MOU Template: Agreement between the National Park Service and [Repository] on Management of NPS Natural History Collections.
- 12-4** SWAN Transfer of Natural History Files to Curation Excluding Object Collections.
- 12-5** Form 10-127. Outgoing Loan Agreement and Conditions for Outgoing Loans
- 12-6** Instructions and spreadsheet template for transferring voucher information to ANSC+
- 12-7** ANCS+ Catalog Data Fields
- 12-8** Link between I&M and Archives, Inventory Program, Alaska Regional Meeting, January 30, 2004 Minutes

Figures

Figure 1-1	Shaded relief map of Southwest Alaska Network park units	1-7
Figure 2-1	Data management responsibilities for project leaders and Network data managers	2-5
Figure 3-1	Information management process and workflow	3-2
Figure 4-1	Simplified comparison of infrastructure and system architecture	4-1
Figure 4-2	Server infrastructure	4-3
Figure 5-1	Conceptual data model	5-3
Figure 5-2	Logical data model	5-4
Figure 5-3	Physical data model	5-4
Figure 7-1	Schematic of QA/QC controls during project stages	7-2
Figure 11-1	Electronic files storage and archives diagram	11-3
Figure 12-1	Workflow of Natural History Curation	12-3

Tables

Table 1-1	Categories of data products	1-4
Table 2-1	Categories of data stewardship involving all Network personnel	2-2
Table 2-2	Primary responsibilities related to data management	2-3
Table 3-1	List of tasks per stage and staff with primary and secondary responsibility.	3-3
Table 4-1	General directory structure and their roles for Master Libraries	4-6
Table 6-1	Vital sign data collection information at a glance	6-3
Table 6-2	Satellite resolution, swath width, area, cost, and history.	6-11
Table 8-1	Example of the Master Version Table used to track changes in vital signs protocols	8-4
Table 10-1	Primary repositories for SWAN information	10-2

Executive Summary

The National Park Service implemented a strategy known as “vital signs monitoring” to develop scientifically sound information on the status and long-term trends of park ecosystems and to determine how well current management practices are sustaining those ecosystems. The Southwest Alaska Network (SWAN) consists of five Alaskan park units (Aniakchak National Monument and Preserve, Alagnak National Wild River, Katmai National Park and Preserve, Kenai Fjords National Park, and Lake Clark National Park and Preserve). Collectively these units comprise 9.4 million acres or 11.6 percent of the total land area managed by the National Park Service. Network parks encompass climatic conditions, geologic features, near pristine ecosystems, natural biodiversity, freshwater, and marine resources equaled few places in North America.

The SWAN Vital Signs Monitoring Plan (Bennett, et. al. 2006) describes the monitoring strategy for these ecosystems. This Data Management Plan (DMP) is an appendix to the Monitoring Plan and is a guide to ensure over time that relevant natural resource data collected will be entered, validated, analyzed, reported, documented, cataloged, archived, and made available to others for management decision-making, research, and education.

The DMP serves as an overarching strategy in which protocols will be developed. Each protocol will contain specific data management standard operating procedures (SOPs). Some general SOPs that cross many protocols, such as specifications for developing a database or for managing photographs, have already been written and are cited in this document. Most recent SOPs can be found on the SWAN website (<http://www.nature.nps.gov/im/units/swan/>)

The Data Management Plan focuses on the processes used to:

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

1. Data Management Goals

The goal of the NPS I&M Program is to provide scientifically and statistically sound data to support management decisions for the protection of park resources. The goal of data management is to ensure the quality, interpretability, security, longevity and availability of our natural resource data. The goal of the Data Management Plan is to outline the procedures and work practices that support effective data management.

The DMP goals are to ensure that:

- Data managed by the network are of high quality
- Network data can be easily interpreted

- Data are secure for the long term
- Network data are readily available

2. Data Stewardship Roles and Responsibilities

Every individual involved in the I&M Program is required to understand and perform data stewardship responsibilities in the production, analysis, management, and end use of the data as described in the DMP and the specific monitoring protocols. Specific roles and responsibilities for vital signs monitoring are written in each monitoring protocol. Senior staff shares the responsibility in ensuring proper data management procedures are followed.

3. Data Management Process and Workflow

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

A project is divided into the following stages:

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

4. Infrastructure and System Architecture

Infrastructure refers to the network of computers and servers that our information systems are built upon. SWAN relies heavily on the national, regional and park information technology (IT) personnel and resources to maintain its computer infrastructure.

The infrastructure needs to support these required functions:

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

SWAN will also use the infrastructure and information systems from partners who share monitoring objectives or provide source information.

5. Database Design

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

- Understand conceptually the data life cycle flow of the data collection process.
- Determine the data relationships as the implementation progresses.
- Determine how the information will be presented.

Understanding the relationships among the data components collected is the key to the success of a database and its utility. If the relationships are misunderstood, the database may become tedious in data entry and cumbersome at data output.

6. Data Acquisition and Processing

Data may be acquired by different mechanisms:

- Data harvest (such as the case of downloading data from other agencies)
- Field collection
- Remote sensing

Data acquired by any of these methods will follow individual monitoring protocols.

7. Quality Assurance/Quality Control

High quality data and information are vital to the credibility and success of the Inventory and Monitoring (I&M) Program and everyone plays a part in ensuring products conform to data quality standards.

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

- Field crew training
- Standardized field data sheets with descriptive data dictionaries
- Use of handheld computers and data loggers
- Equipment maintenance and calibration
- Procedures for handling data in the field
- Database features to minimize transcription errors, including imports from data loggers, range limit, pick lists, etc.
- Verification and validation, including automated error-checking database routines

Quality assurance methods should be in place at the inception of any project and continue through all project stages to final archiving of the dataset. It is critical that each member of the team work to ensure data quality.

The final step in project quality assurance is the preparation of summary documentation that assesses the overall data quality. A statement of data quality will be composed by the Project Leader and incorporated into formal metadata.

Metadata for each dataset will also provide information on the specific quality assurance procedures applied and the results of the review.

8. Data Documentation

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

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

Data documentation will be available and searchable in conjunction with related data and reports via the SWAN website as well as with the national I&M Program NR-GIS Data Store.

9. Data Summary and Analysis

Providing meaningful results from data summary and analysis is a cornerstone of the I&M Program and characterizes the network's data management mission to provide useful information for managers and scientists. Each monitoring protocol establishes requirements for on-demand and scheduled data analysis and reporting. Based on these requirements, the associated databases for the protocols include functions to summarize and report directly from the database as well as output formats for import to other analysis software programs. In addition to tabular and charted summaries, the network provides maps of natural resource data and geographic information system (GIS) analysis products to communicate spatial locations, relationships and geospatial model results. See Chapter 7 for a more detailed description of the Network's analysis and reporting schedule and procedures.

10. Data Dissemination

SWAN data dissemination strategy aims to ensure that:

- Data are easily discoverable and obtainable
- Only data subjected to complete quality control are released, unless necessary in response to a Freedom of Information Act (FOIA) request
- Distributed data are accompanied by appropriate documentation
- Sensitive data are identified and protected from unauthorized access and inappropriate use

SWAN data products will be accessible by:

- The SWAN website (<http://www.nature.nps.gov/im/units/swan/>).
- NR-GIS Data Store
- Service-wide databases, such as NPSTORET, NPSpecies, and NatureBIB.
- Regional, Network, or Park data servers
- External repositories such as the Alaska Resource Library and Information Service, US Geological Survey, University of Alaska, Alaska Department of Fish and Game, Western Regional Climatic Center, Exxon Valdez Oil Spill Trustee Council, and many others.
- FTP sites, CDs or DVDs, as appropriate.

11. Data Maintenance, Storage and Archiving

SWAN data maintenance, storage and archiving procedures aim to ensure that data and related documents (digital and analog) are:

- Kept up-to-date with regards to content and format such that the data are easily accessed and their heritage and quality easily learned.
- Physically secure against environmental hazards, catastrophe, and human malice.

Primary data maintenance will be performed on the central Alaska Regional Office (AKRO) server or Network server and will follow AKRO backup procedures. Data and information content of SWAN files stored on this server will be kept current.

Project data will be electronically archived as stand-alone products and will include:

- Project documentation
- Data in raw, verified, and analyzed conditions
- Respective metadata
- Supporting files, such as photographs, maps, etc.
- All associated reports

Final deliverables from project data will be integrated with ongoing libraries and databases.

12. Natural History Archiving, Curation and Records Management

In most instances, administrative documents, natural history specimens, photographs, audio tapes and other materials are essential companions to the digital data. Direction for managing many of these materials (as well as digital materials) is provided in NPS Director's Order 19: Records Management (2001) and its appendix, NPS Records Disposition Schedule (NPS-19 Appendix B, revised 5-2003). NPS-19 states that all records of natural and cultural resources and their management are considered mission-critical records, that is, necessary for fulfillment of the NPS mission and must be permanently archived.

The SWAN DMP guides project leaders in complying with archival directives. Physical items considered project products such as reports, maps, photographs, or notebooks will be cataloged and filed in the Network's main office and accessioned through the NPS Rediscovery curatorial database. A copy of the accessioned material will be archived according to NPS Standards and follow the procedures outline in the SWAN DMP. Physical specimens, such as plants, animals, or tree core samples, will be accessioned and housed at the appropriate and accepted archival institution.

13. Implementation

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

During the development stages of the SWAN I&M Program, data mining was done at each of the parks and the regional office. This consisted of looking through archives, natural resource folders, libraries, and electronic files. Some examples of common data management problems are:

- Unfinished projects
- Lack of header information
- Lack of documentation
- Lack of file naming standard and version control
- No archive or final storage

Good data management practices will take time. Chapter 13 lists short- and long-term goals for implementing the DMP.