

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.

1. Introduction

1.1. The Inventory and Monitoring Program

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

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

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

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

1.2. Scope of the Data Management Plan

The Data Management Plan (DMP) is a guide for the SWAN Inventory and Vital Signs Monitoring Program to ensure the continuity and documentation of data management methods and procedures over time. It may provide a model for parks to follow with other projects, but its primary focus is managing the large integrated data sets developed by the Network.

The DMP focuses on the processes used to:

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

The I&M Program anticipates 30% of the program's efforts will involve data management. The DMP provides the overarching guide for this effort.

The DMP provides guidance for ensuring that program data are documented, secure, and remain accessible and useful for future generations. The DMP, in turn, refers to other guidance documents and standard operating procedures that convey the specific standards and steps for achieving the data management goals. The DMP is an appendix to the SWAN Monitoring Plan (Bennett et al. 2006), but is designed to be a stand-alone document that may be used by the parks or other programs.

Data management specific to vital signs monitoring will be addressed in detailed protocols. Specifics for methods, frequency and timing of data acquisition and plans for the primary data management for each vital sign measure will be outlined in these protocols. Each protocol will incorporate the strategies and guidelines of the DMP.

1.3. Data Management Goals

One of the five service-wide goals of the NPS I&M Program is to provide sound data to support management decisions for the protection of park resources. The goal of data management is to ensure the quality, interpretability, security, longevity and availability of our natural resource data. The goal of the DMP is to outline the procedures and work practices that support effective data management. The DMP may also serve as a guide to the enhancement of legacy information to match formats and standards described in this document.

The DMP goals are to ensure that:

- **Data managed by the network are of high quality**, including designing standardized data entry, importation, and handling procedures that effectively screen for inappropriate data and minimize transcription and translation errors;
- **Network data are readily available**, by implementing standard procedures for distributing data, while protecting sensitive data; and designing a standardized filing system for organizing I&M information;
- **Network data can be easily interpreted**, by considering the users' needs as the primary factor driving the design of summary reports and analyses; establishing rigorous data documentation standards; integrating common data tables and fields in NPS or regional standards; and making summary information available in formats tailored to the variety of audiences interested in I&M program results;
- **Data are secure for the long term**, including instituting standard procedures for versioning, data storage and archiving; and natural history archiving, curation and records management are provided to NPS curators.

1.4. Audience

The intended audiences of the DMP are:

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

1.5. Revisions

The DMP is scheduled for a full review and revision every five years, coinciding with the SWAN I&M Program review schedule. Until then, minor revisions may be made as needed and will be documented in the Plan Revision Log at the beginning of this document.

1.6. Types of Data Covered by this Plan

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

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

This plan encompasses a range of products that are coordinated or managed by SWAN. These products fall into four general categories: data, documentation, reports, and administrative records (Table 1-1).

Table 1-1. Categories of data products

Category	Description	Examples
1) Data		
Raw Data	Data obtained from the environment and that has not been subjected to any quality assurance or control beyond those applied during field work.	- field data sheets - specimens - remotely sensed data - data gathered electronically on field computers - GPS rover files - photographic imagery
Verified and Validated Data	Data that have been evaluated for completeness, correctness, and conformance/compliance of a specific data set against the standard operating procedure (verified), as well as reviewed for specific analytic quality (validated).	- Relational databases - Tabular data files - Laboratory results - GIS layers - Maps
Analyzed Data	Data that have been analyzed after field collection and verification. This includes statistical analyses to estimate population trends or a compilation of analyzed data from different sources or time periods to derive new information.	Summarized reports, data and maps from statistical analyses or queries
2) Documentation	Documentation provides the information required to understand the context of the data.	- Data collection protocols - Data processing/analysis protocols - Record of protocol changes - Data dictionary - FGDC metadata - Database design documents - QA/QC reports - Catalogs
3) Reports	Reports provide a means of presenting and publishing the methods and the results of analysis in the context for which it was intended.	- Annual progress reports - Final reports - Trend analysis reports - Publications
4) Administrative Records	Administrative records supplement the context of a project and should be considered part of the project's deliverables.	- Contracts and agreements - Study and work plans - Research permit - Critical administrative correspondence - Investigator Annual Reports

1.7. Sources and Priorities of Natural Resource Data

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

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

The priorities for Network data management efforts are:

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

1.8. I&M Program Organization

1.8.1. National and Regional Organization

NPS has implemented a strategy designed to institutionalize natural resources I&M on a programmatic basis throughout the agency. The 270 national parks have been grouped into 32 vital signs networks linked by geographic similarities, common natural resources, and resource protection challenges. The network approach facilitates collaboration, information sharing, and economies of scale in natural resource monitoring. The approach also provides parks with a “minimum infrastructure” to initiate natural resource monitoring.

Alaska National Parks have been grouped into four I&M networks:

- Arctic Network (ARCN), headquarters stationed in Fairbanks
- Central Alaska Network (CAKN), headquarters stationed in Fairbanks
- Southeast Alaska Network (SEAN), headquarters stationed in Juneau
- Southwest Alaska Network (SWAN), headquarters stationed in Anchorage

The regional I&M Program coordinator provides oversight to these networks and is located in the Alaska Regional Office (AKRO) in Anchorage. Other regional natural resource scientists, regional and park GIS staff, and the regional science advisor are also located at AKRO.

All Alaska Networks have their respective Board of Directors comprised of superintendents, the regional I&M coordinator, and regional science advisor, and a Technical Committee comprised of park resource staff, selected regional resource staff, and a liaison from USGS-Biological Resources Division. Some

members of the Technical Committee serve in more than one network, providing cross coordination. More detailed information on the Board of Directors and Technical Committee can be found in the respective Network Charter.

Each Alaska network manages its respective program. However, the Alaska networks may share common data resources such as support from the regional office and from Alaska-based partners. Alaska networks also share common challenges in implementation with the remote nature of the parks and are working together to find solutions to these hurdles.

1.8.2. Southwest Alaska Network

SWAN includes five park units:

- Katmai National Park and Preserve (KATM), offices in King Salmon
- Alagnak Wild River (ALAG), managed by the KATM offices
- Aniakchak National Monument and Preserve (ANIA), managed by the KATM offices
- Kenai Fjords National Park (KEFJ), offices in Seward
- Lake Clark National Park and Preserve (LACL), offices in Port Alsworth, Anchorage, and Homer

Collectively these units comprise 9.4 million acres (see map in Figure 1-1).

Staff will consist of the following positions. Staff composition is addressed in more detail in Chapter 8 of the Vital Signs Monitoring Plan (Bennett et al. 2006).

Network Staff	Alaska Regional Staff	Park Staff
• Network Coordinator • Assistant Network Coordinator • Data Manager • Assistant Data Manager • Biometrician/Wildlife Biologist • Marine Nearshore Ecologist • Landscape Ecologist • Botanist • Freshwater Ecologist • Biological Technicians • Physical Science Technician • Clerk Assistant	• Physical Scientist • Ancillary support from regional Information Technology (IT) staff, Curators, and GIS staff.	• Ecologists • Wildlife Biologists • Fisheries Biologist • Coastal Biologist • Biological Technicians • Education/Outreach Specialist • Ancillary support from park IT staff, Curators, and Permitting Officers

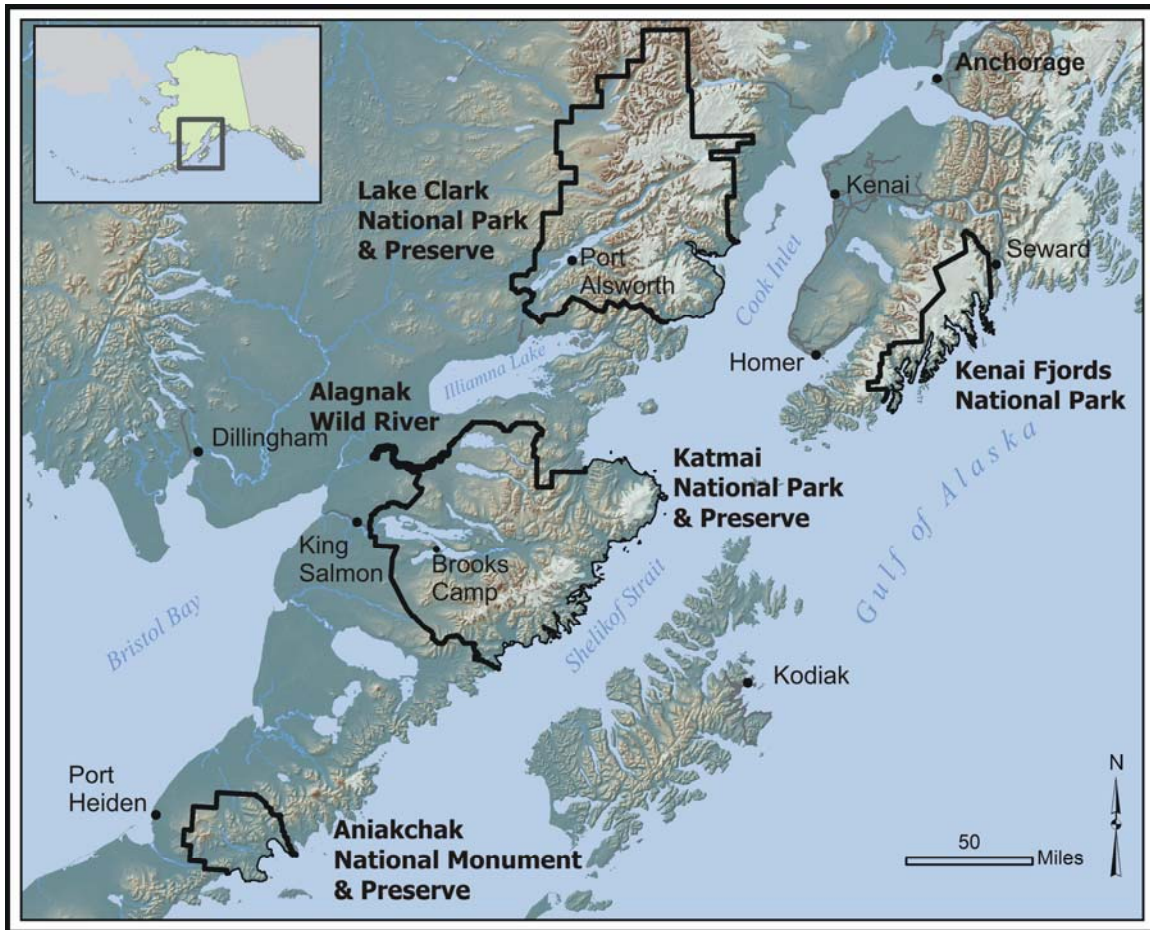


Figure 1-1. Shaded relief map of Southwest Alaska Network park units.

1.9. Organization of this Plan

Chapters 1 through 4 are overview chapters describing roles, processes and framework. Chapters 5 through 12 discuss data management operations. Each chapter begins with mandates, when applicable, and ends with Chapter Credits and References. The writing of the DMP for the I&M Program involved over a year of many discussions among I&M data managers. The Chapter Credits acknowledges individuals who contributed to the development of the chapter and provides initial contacts for further discussion and development. References listed in a chapter provide a selected list of documents, guidelines or policies that are also helpful in further discussion and development of data management operations. Appendices are prefixed with the chapter number (i.e., Appendix 4-1, for an appendix referenced in Chapter 4) for easy reference.

1.10. Chapter Credits

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

1.11. References

Bennett, A.B., W.L. Thompson, and D.C. Mortenson. 2006. Vitals signs monitoring plan, Southwest Alaska Network. National Park Service. Anchorage, AK.

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

2. Data Management Roles and Responsibilities

The production, interpretation, management, and dissemination of natural resource data is a team effort. Every staff member has data stewardship responsibilities. These responsibilities go well beyond the past organizational requirements due to the long-term nature and importance to the I&M Program. Data management, analysis, and reporting will require a large investment of time for each staff member. I&M Program managers expect that at least 30 percent of staff time will be spent on data management related tasks. With such an investment of time and money, it is essential that each project team member understand their roles and responsibilities to data management. This chapter clarifies roles and outlines the principal and ancillary data management responsibilities of network staff, cooperators, and other project participants. Monitoring protocols and the inventory study plan may contain specific tasks or additional roles that supersede those described in this document. Table 2.1 lists the types of data stewardship activities that may take place during a project and the associated team members.

2.1. Data Stewardship Roles

The current and expected capacity, diversity, and rate of change in information technology make managing any large amount of data a greater task than any person can expect to do alone. Producing high quality data and information about natural resources and ecosystem functions requires a group of people working together to steward data and information assets. Knowledgeable individuals from scientific, administrative, and technological disciplines must work together to ensure that data are collected using appropriate methods, and that resulting data sets, reports, maps, models, and other derived products are well managed. Data sets and the presentation of these data must be credible, representative, and available for current and future needs.

Chapter 8 of the Monitoring Plan (Bennett et al. 2006) describes staffing as part of the implementation plan for the Program. Described here, Table 2-2 generalizes the positions as roles and describes the data management responsibilities.

Table 2-1. Categories of data stewardship involving all Network personnel. Each position is listed in only one category according to overriding responsibilities. However, most positions contribute in each category.

Stewardship Category	Related Activities	Principal jobs or positions
Production	Collecting data or information from any original or derived source. This includes recording locations, images, measurements, and observations in the field, digitizing source maps, keying in data from a hardcopy source, converting existing data sources, image processing, and preparing and delivering informative products, such as summary tables, maps, charts, and reports.	Project Crew Member Project Crew Leader Data/GIS Specialist or Technician
Analysis	Using data to predict, qualify, and quantify ecosystem elements, structure, and function as part of the effort to understand these components, address monitoring objectives, and inform park and ecosystem management.	Network Ecologist Resource Specialist Network Biometrician
Management	Preparing and executing policies, procedures, and activities that keep data and information resources organized, available, useful, compliant, and safe.	Network Data Manager Project Leader GIS Manager Information Technology Specialist Database Manager National Level I&M Data Manager
End Use	Obtaining and applying available information to develop knowledge that contributes to understanding and managing park resources.	Network Coordinator Park managers and superintendents Others

Table 2-2. Primary responsibilities related to data management

Role	Primary responsibilities related to data management
<i>Principal Roles</i>	
Network Coordinator	Ensure programmatic data and information management requirements are met as part of overall Network business
Network Biometrician	Develop overall Network sampling design. Responsible for data analysis associated with monitoring and research, the development of models, and the documentation of procedures used. Review proposed methods for sound statistical practice.
Network Ecologist or Project Leader	Ensure useful data are collected and managed by integrating natural resource science in network activities and products, including specifying objectives, sample design, data analysis, synthesis, and reporting. Oversee and direct operations, including data management requirements and protocols, for one or more Network projects. Maintain communication with project staff, Network Data Manager, and resource specialist regarding data management. Responsible for the final submission and review of all products and deliverables.
Resource Specialist	Understand the objectives and the protocols of the project, the resulting data, and their scientific and management relevance. Make decisions about data with regard to validity, utility, sensitivity, and availability. Describe, publish, release, and discuss the data and associated informational products. Note: The Resource Specialist serving as a Project Leader is also responsible for the duties listed with that role.
Project Crew Leader	Supervise crew members to ensure their data collection, protocols, and management obligations are met, including data verification and documentation. This role is likely to be a park-based employee.
Project Crew Member	Record and verify measurements and observations based on project objectives and protocols. Document methods, procedures and anomalies. This role is likely to be a park-based employee.
Network Data Manager	Oversee development, implementation, and maintenance of data infrastructure and standards. Facilitate and integrate data and metadata across monitoring and research projects. Oversee long-term storage and maintenance of the data. Work with others to design and develop monitoring databases, applications, and products.
Assistant Data Manager/GIS Specialist	Support the Network Data Manager and project leaders in implementing data management procedures. Apply particular knowledge and abilities related to database software and associated application(s).
GIS Specialist	Assignments may include GIS activities, cataloging, data conversion and integration, documentation, and reporting.

<i>Ancillary Roles (Non-programmatic)</i>	
Park and Regional Curator	Oversee all aspects of natural history archiving and curation (specimens, final reports, field notes, etc.). This may include the acquisition, cataloging and documentation, preservation, and use of park collections.
Park and Regional Information Technology Staff	Maintain the local area network, establish and maintain system security, update software and hardware, and implement a secure backup system of server files and data.
Regional GIS Liaisons	Provide support for long-term storage of GIS data. Update and maintain GIS software and tools. Provide technical assistance.
National I&M Data Manager	Provide service-wide database design, support, and services, including receiving and processing to convert, store, and archive data in service-wide databases
Park Permitting Officers	Requires permits to be completed before field work is done. Requires Investigator Annual Reports to be complete at the end of each calendar year. Permitting officers ensure appropriate park staff are informed of field collections and work.
<i>Information Customers</i>	
Education / Outreach Specialist	Directs education and outreach program for the Network. Coordinates interpretation of I&M Program information for educational and outreach purposes. Oversees public outreach information through school materials, web sites and other means.
Other End Users	These information consumers may include Board of Directors (superintendents), Technical Committee, network coordinators, resource managers, science advisors, and the science community. Provide feedback on information, presentation needs, and interpretation.

2.2. Data Management Team

The Network coordinator, project leaders, biometrician, data manager and GIS specialists comprise the central data management team for a given inventory and monitoring project. The project leader and the data manager provide the primary leadership and may share responsibility for some overlapping tasks (see Figure 2-1).

Although this chapter describes roles and responsibilities in general terms, the specific tasks assigned to each position is best understood in the context of a project. Chapter 3 describes the workflow of a project and the data management needs at each stage of a project. Necessary tasks and staff responsibilities are described in Table 3-1.

Because of the collaborative nature of project data management, good communication among these personnel is essential to meeting program goals.

Communication is promoted by providing:

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

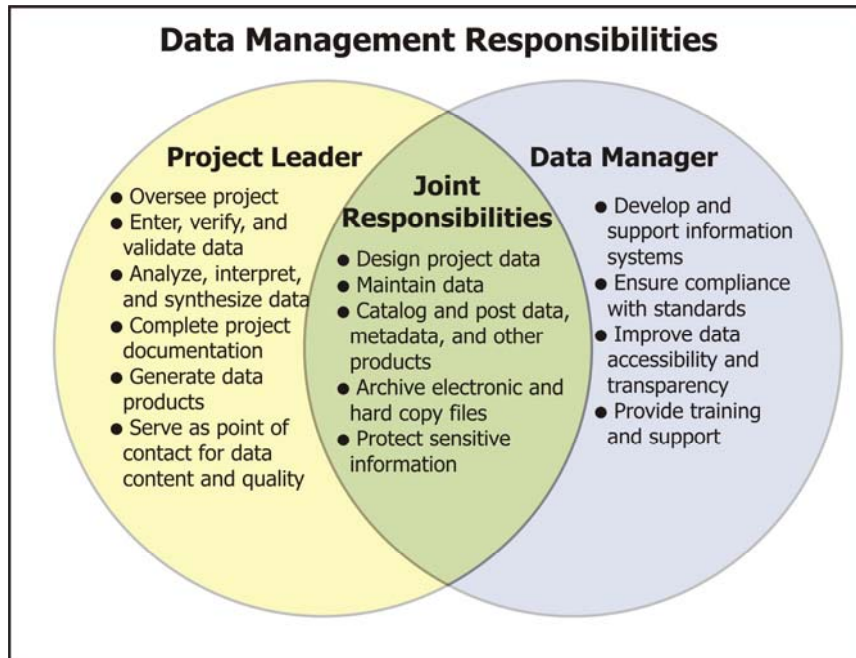


Figure 2-1. Data management responsibilities for project leaders and network data managers.

2.2.1. Network Coordinator

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

2.2.2. Project Leader

Project leaders oversee and supervise all phases of an I&M project from initiation to product delivery. Their active involvement in data management determines the quality and usefulness of the project data and overall success and longevity of the I&M Program.

In terms of data management, the project leader assumes ownership of the data. In terms of the SWAN monitoring program, each vital sign is considered a project and will be assigned leadership by the Network technical committee. To ensure the quality of each project, project leaders are responsible for designating an alternate leader who is capable of maintaining project operations in his or her absence.

2.2.3. Network Data Manager

The Network data manager oversees the development, implementation, logical validation and long-term storage of I&M data. The data manager works with project leaders and other program staff to design databases, applications, and products, and to facilitate dissemination of data-related products.

The Network data manager exercises stewardship over the data to ensure data are archived, documented, made compatible with other program data and made available/discoverable. The data manager ensures that the mechanisms in place for making data discoverable and available are maintained. Both the data manager and the project leader ensure that the information conveyed via these mechanisms is up-to-date and accurate.

2.2.4. AKRO GIS Specialists

The GIS specialists manage spatial data themes associated with Network I&M projects, as well as other spatial data related to the full range of park resources. They incorporate spatial data into the GIS. They also maintain standards for geographic data and are responsible for sharing and disseminating GIS data throughout the Network.

2.3. Data Management Coordination

The Natural Resource Challenge stresses the importance of collaboration among NPS, other public agencies, universities, and non-governmental organizations. The I&M Program encourages coordination among participants at all levels to help ensure that data collected by NPS staff, cooperators, researchers, and others are entered, quality-checked, documented, analyzed, reported, archived, cataloged, and made available for management decision-making, research, and education.

Data management is an adaptive process that facilitates science and fosters partnerships among data managers and scientists within the network and across Alaska. SWAN benefits from the data coordination and collaboration of all organization levels within NPS and with other agencies located in Alaska. The SWAN Monitoring Plan (Bennett et al. 2006) discusses in more detail partnerships with other organizations or agencies.

2.3.1. National Park Service

National to Network

Other divisions within NPS provide baseline inventory data products that will benefit the Network, such as:

- NPSTORET by the Water Resources Division for water quality data
- ARIS by the Air Resources Division for air quality data
- NPSpecies from Biological Resource Management Division for biological inventories
- NatureBIB from Natural Resource Program Center, Geological Resources Division, and Biological Resource Division for bibliography citations
- GIS Theme Manager, NR-GIS Data Store, and NR-GIS Clearinghouse from Natural Resource Information Division, Geological Resources Division, and Biological Resource Division for GIS and other data

At a national level, the I&M Program has designed and built data tools such as NPSpecies, Dataset Catalog, NatureBIB, and the Natural Resource Database Template, which are used by all of the Networks.

Network to Network

Data managers throughout the I&M Program regularly coordinate with each other and national program staff via annual meetings, conference calls, workgroups, listserv, web sites, and informal communication. The Network data manager contributes to these national discussions regarding data standards, data integration and analysis methods, and improving data management awareness.

SWAN maintains an active role in promoting consistency among protocols and data sets involving other networks. Cooperative projects are being discussed with other networks.

Network to Parks

Direct involvement by park-based resource managers and staff is essential to the long-term success of the I&M Program. Data-related training is provided to the parks and discussions on good data management practices are frequent. Recent improvements to park computer networks allow the data manager and park staff to view the same project files and maintain quality information.

2.3.2. Alaska Data Coordination

Partnerships

Partnerships with organizations outside of NPS are discussed in Chapter 8 of the SWAN Vital Signs Monitoring Plan (Bennett et al. 2006) and should be used as a reference. Each protocol will indicate specific data sharing and coordination among organizations. These organizations may include:

- Alaska Resource Library and Information System (ARLIS)
- Exxon Valdez Oil Spill Trustee Council
- State of Alaska, Department of Fish and Game
- State of Alaska, Department of Environmental Conservation
- State of Alaska, Department of Natural Resources
- US Fish and Wildlife Service
- US Geological Survey
- University of Alaska

Alaska Geospatial Data Committee

The Alaska Geospatial Data Committee (AGDC), mirroring a similar structure as the Federal Geospatial Data Committee (FGDC), is a multi-agency committee focused on geospatial standards and cooperative efforts for basic cartographic themes (hydrography, transportation, cadastral, etc.). Meetings are generally held quarterly.

Alaska Arc User Group

The Alaska Arc User Group (AK-AUG) is an informal group that meets in Anchorage once a month to share GIS ideas, problems and solutions. The AK-AUG also jointly sponsors an annual Survey and Mapping Conference. These events are a means to get to know other data managers and technicians in other agencies and to stay current on projects and data sets being done locally.

2.4. Chapter Credits

Rob Daley (Greater Yellowstone Network), Velma Decker (Cape Code National Seashore), Margaret Beer (Northern Colorado Plateau Network) and Doug Wilder (Central Alaska Network).

3. Data Management Process and Workflow

This section describes the stages of a project and how data are affected by each stage. A project may be considered short term with a finite ending date (1 to 5 years) or long term with ongoing collection supporting monitoring goals. Both short- and long-term projects deliver products that must be managed. A short-term project may include research, inventory, or pilot projects preparing for long-term monitoring. Most vital signs monitoring will be long-term projects. In this case, planning, designing, developing, and testing have been completed during the pilot projects and protocols have been formalized for long-term monitoring. The primary difference in data management between short-term projects and long-term vital signs monitoring is the requirements for adhering to standards, protocols, and peer review. Short-term projects are encouraged to follow similar standards as long-term monitoring, but this may not be reasonable due to scope and budget of the project. Whether the projects are short- or long-term, the life cycles of the data and products are similar for both and are described in this chapter.

3.1. Data Life Cycle 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. This chapter provides a sequence of project stages and anticipated deliverables at each stage. For data management to be effective, it must occur throughout the project life cycle.

A project is divided here into the following stages (see Figure 3-1 and Table 3-1):

1. Planning and Approval
2. Design and Testing
3. Implementation
4. Product Development and Delivery
5. Product Integration
6. Evaluation and Administrative Reporting

SWAN uses a project tracking database to document and support the progress of a project. This database provides an overview of the project, identifies primary contacts and due dates, lists deliverables, and notes the status and location of the deliverables. It also helps to generate reports used in the final archiving of a project and in data audits.

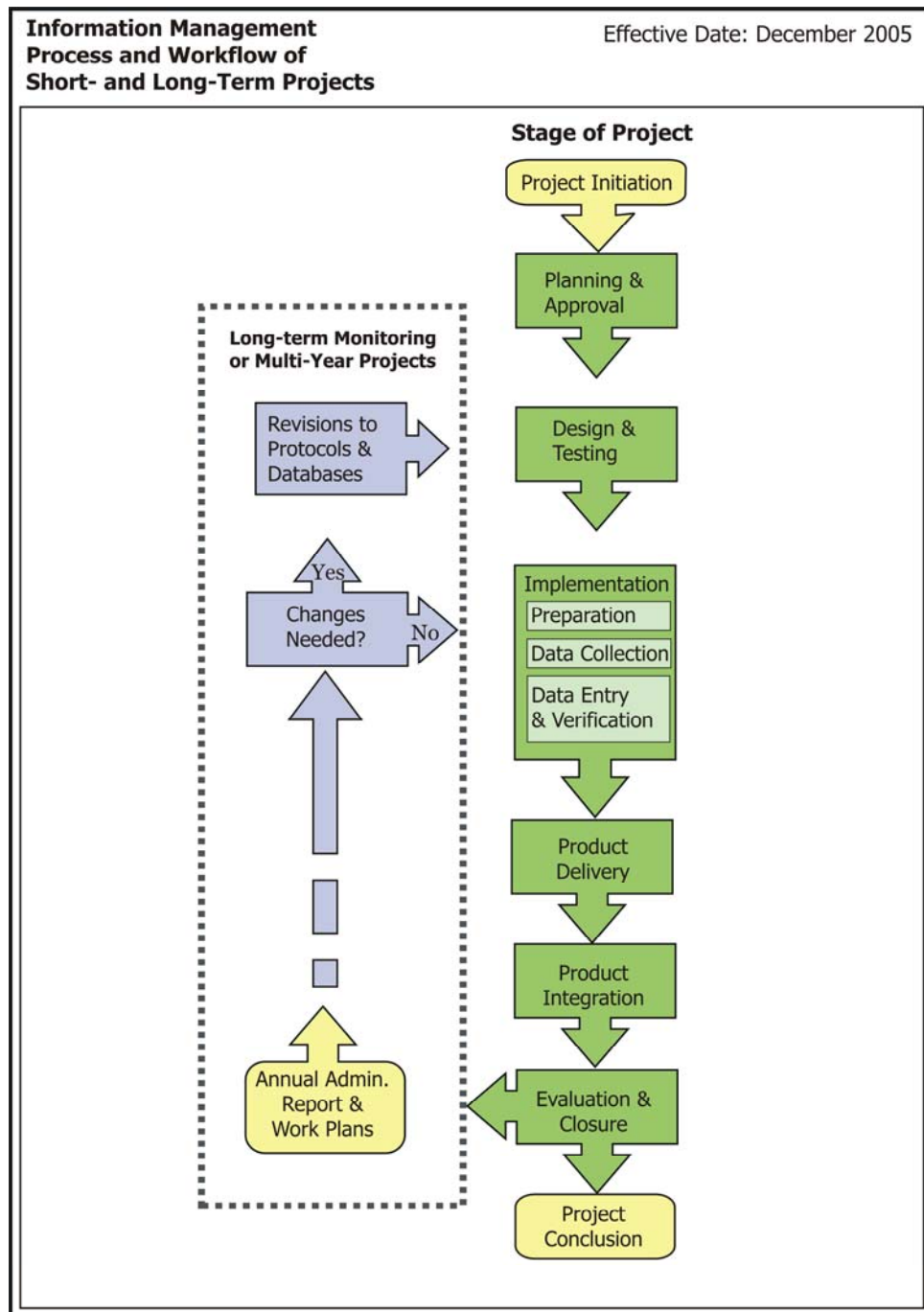


Figure 3-1: Information management process and workflow. This diagram illustrates the workflow for short-term projects (1-5 years) and is characteristic of research, inventories, and protocol development/pilot projects. Long-term vital signs monitoring will mostly be in the Planning & Approval and Design & Testing stages for years 2005-2009, and will be in the other stages for years 2010+.

Table 3-1. List of tasks per stage and staff with primary and secondary responsibility. PL = Project leader; BM = Biometrician; DM = Data Manager; GIS = GIS specialist. * = Primary role; - = Secondary role.

Project Stage	Task	Task Description	PL	BM	DM	GIS
Planning & Approval	1.1	Notify data manager of new project in planning stage	*			
	1.2	Develop study plan and budget	*	-	-	
	1.3	Create a project tracking record and assign a project code (status: planned)	-		*	
	1.4	Identify project deliverables with links to SWAN specifications	*		-	-
	1.5	Determine data sensitivity and ownership	*		-	-
	1.6	Develop language for contracts/agreements and permits	*	-	-	-
	1.7	Apply for permits in the park and note in project tracking	*			
Design & Testing	2.1	Notify data manager of changing status	*			
	2.2	Update project tracking record (status: design or rejected)			*	
	2.3	Clarify roles and responsibilities	*	*	*	*
	2.4	Scoping for information systems design	*	*	*	*
	2.5	Initiate metadata record	*		-	-
	2.6	Create data dictionary and data forms	*		*	*
	2.7	Define quality assurance measures	*	-	-	
	2.8	Database and application development	*	-	*	*
	2.9	Sample site selection and documentation	*	*	-	*
	2.10	Provide training as needed and time permits			*	*
Implementation	3.1	Notify data manager of changing status	*			
	3.2	Training	*			
	3.3	Update project tracking record (status: active or abandoned)			*	
	3.4	Data acquisition, data entry, upload data logger files	*			
	3.5	Upload GPS and other data logger data into database				*
	3.6	Quality assurance: verification, validation and processing, data certification	*	*		
	3.7	Update metadata and documentation	*			
	3.8	Deliver data and metadata to DM/GIS (see Task 5.2)	*			
Product Development & Delivery	4.1	Summarization, analysis, map production	*	*	*	*
	4.2	Interpret results, reporting, peer review	*	*		
	4.3	Assemble products and deliver to DM	*			
	4.4	Route products to appropriate product handler(s)			*	
	4.5	Check for completeness, proper format, metadata and certification			*	*
	4.6	Rename files, store in digital library			*	*
	4.7	Update project tracking database			*	

		(deliverable status: checked in)				
	4.8	Upload certified data into master project database			*	-
	4.9	Finalize and parse metadata record	-		*	*
	4.10	Annual data audit report (may occur at any stage of project)	*		*	
Product Integration	5.1	Publish products and integrate with national applications (NatureBib, NPSpecies, NPSTORET, NR-GIS Data Store)			*	*
	5.2	Update project tracking database (deliverable status: published)			*	*
	5.3	Electronic permanent storage			*	*
	5.4	Hardcopy archival and distribution	*		-	
Evaluation & Administrative Reporting	6.1	Verify that all products are accounted for	*		*	*
	6.2	Administrative evaluation and reporting, decision on continuation and methods changes for long-term projects	*	-		
	6.3	Communicate administrative decisions and information management needs to data manager/GIS specialist	*			
	6.4	Update project tracking database (project status: complete, design, active or suspended)	-		*	
	6.5	Final data audit report	*		*	

1. Planning and Approval

During planning and approval, a request for proposals package may be distributed to likely candidates or a proposal may be formulated in-house. The proposal provides the initial ideas of a project providing enough description and budget information to determine if the idea should be continued. If it is, then a more detailed study plan is written and reviewed. Funding is provided either in-house or via a contract or cooperative agreement. A contract or cooperative agreement will make reference to the detailed study plan to describe the scope of work. The annual work plan is a supplemental to the study plan and provides specific tasks and identifies deliverables for a specific year. These plans and cooperative agreements should include specifications of deliverables (Appendix 3-1). Permits applications are completed after the study plan is approved and must be in hand before any work is started.

2. Design and Testing

During design and testing, a project is ready to begin formalizing and testing sampling methods, procedures, and analysis. A more thorough data discovery search is done, using the many sources readily available. Methods and sampling design are developed or modified, field forms and datasheets/databases are created, and standard operating procedures are written. These methods, data,

and SOPs are practiced, and are tested for quality control/quality assurance and preliminary integration into permanent storage.

Many SOPs and guidelines will be written for the implementation, product delivery and product integration of the I&M Program. Review of these existing documents at this stage is advised.

It is critical at this stage for the project leader and the data manager to work together. The dialog between these two will help to ensure information is consistently captured and easy to input into a database, and results can be easily retrieved. Scheduling of a meeting as early as possible (shortly after approval) is encouraged, and is the responsibility of the project leader.

3. Implementation

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

Preparation

Preparation for a project includes logistics, staffing, training, acquiring and maintaining equipment, and preparation for archival accessioning. In addition the project leader should ensure all appropriate research permits have been acquired and are in hand. No data deliverables are anticipated at this time.

Logistics:

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

Training:

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

Equipment:

Equipment purchases should follow the standard NPS operating procedures. Equipment should be tracked and "checked out," then "check in" at completion. Ensure equipment is properly calibrated before and during its use. SWAN has an equipment database to help manage the reservation, check-out and check-in of items.

Accessioning:

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

Data Collection

Acquisition of data sets from non-NPS sources may be collected and analyzed. Field collection should be conducted as specified in the study and work plans, or as specified in protocols and supporting SOPs. The project leader is responsible for policing the collection of data. Any deviation from these protocols should be carefully documented, such as in the case of equipment failure or adverse conditions.

Data Entry and Verification

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

4. Product Development and Delivery

Product delivery is the completion of the deliverables as specified in the study and work plans or Information Deliverables Specifications checklist (Appendix 3-1), organized, packaged and documented both in hard-copy and electronic format, typically as a stand-alone product. In almost all projects, a report is required to summarize and analyze the project and to make recommendations. Each product should follow the guidelines regarding that particular type of deliverable.

5. Product Integration

Products for a single project should be integrated into a larger repository of information. These include entering reports into the appropriate bibliographies and libraries, data into appropriate master databases, and materials into the archives. This will allow the information from the project to be searchable in the

servicewide search engines and retrievable by others, and will provide the long-term storage of these products.

Descriptions of how each product should be cataloged and integrated are described later in Chapters 11 and 12. A project may, however, have additional integration needs, such as when working jointly with other agencies for a common database. Ensuring this step of the project is done is a joint responsibility between the project manager and the data manager. This stage is usually not completed by the contractor.

6. Evaluation and Administrative Reporting

To close out a project, all deliverables, administrative tasks and management tasks must be met to the project/program manager's specifications. A checklist of these items is used for the Project Manager, the Data Manager, and the NPS Permitting Officer, using the Information Management Deliverables for Science Projects and the Project Tracking Database as a guideline (see Appendix A3-1).

From the point of implementation forward, lessons are learned along the way that may change how future projects are implemented. Did the project answer the questions? Should anything be modified? As a project approaches termination or logical milestone, this evaluation and feedback should be conducted. Ideally this has already been written in the annual and final reports.

If this is a long-term monitoring project, a more formal review will be done prior to the continued implementation. Modifications may be minor and include updating SOPs, protocols, and databases, or they may be more significant requiring the Monitoring Plan itself to be revised.

No specific deliverable is required at this stage. However, careful documentation that can trace the changes made is required for long-term monitoring. This may be accomplished in meeting minutes, protocols tracking record, metadata, or other documentation.

3.2. Review Cycle

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

3.3. Chapter Credits

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

4. Infrastructure and System Architecture

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

This chapter discusses the components of the information system architecture and infrastructure needed to reach the scientific community. It includes all infrastructure components required from project development through information dissemination via a clearinghouse.

Infrastructure refers to the network of computers and servers that our information systems are built upon. System architecture refers to the application, database system, repositories, and software tools that make up the framework of the SWAN's data management enterprise. See Figure 4-1 for a simplified comparison of the two.

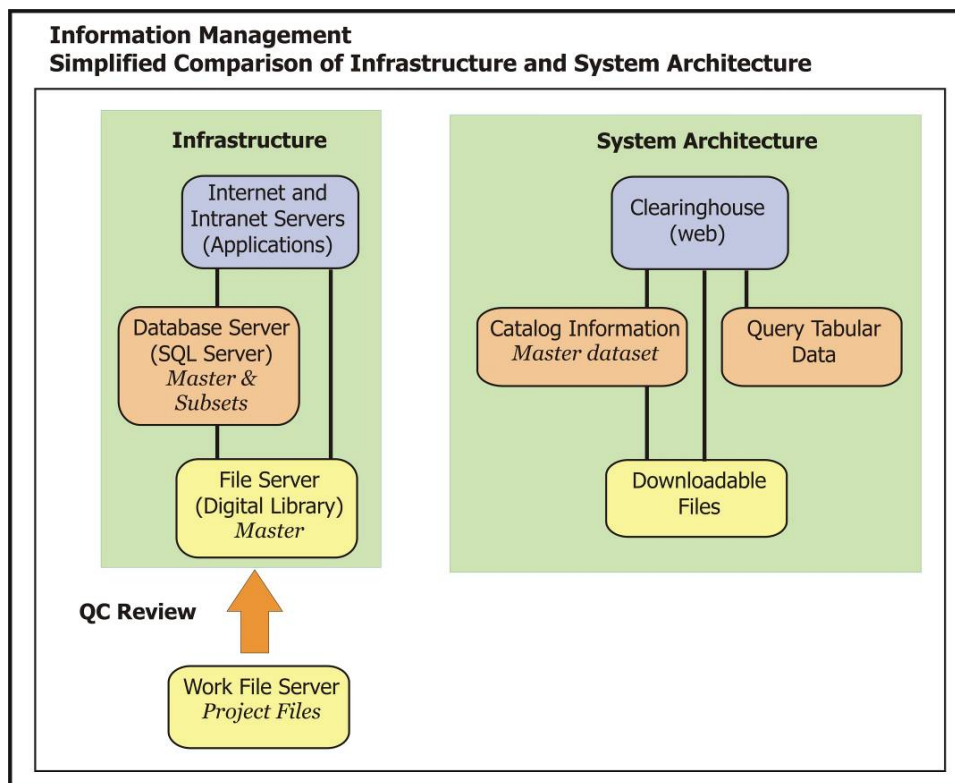


Figure 4-1: Simplified comparison of infrastructure and system architecture

4.1. Infrastructure

SWAN relies heavily on the national, regional and park information technology (IT) personnel and resources to maintain its computer infrastructure. This includes, but is not limited to: computers, servers and other related hardware; software installation and support; email administration; security updates; virus protection; telecommunications; computer networking; and backups of servers.

The infrastructure needs to support these required functions:

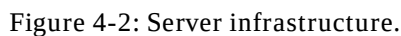
- Provide a central repository for master data sets
- Provide controlled subsets of data for local computing
- Provide a means for uploading and downloading data for both NPS and public
- Support desktop and internet applications
- Provide security, stability, and backups

4.1.1. Computer Server Infrastructure

Computer servers can be divided into five primary functions:

- 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, and 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 website pages and database applications, Blue Angel for clearinghouse search engines, and ArcIMS for internet interactive mapping.

The infrastructure used by SWAN takes advantage of existing infrastructure established at the national, regional and park level. Each level hosts some



4.1.2. Clearinghouse Infrastructure

Clearinghouses are portals to searching and accessing information. There are four primary components that make this a success:

- Gateway
- Node
- Metadata records
- Data

When visiting a public library, the user searches a **gateway** to find a publication. The gateway searches the index of its own **node**, as well as other nodes around the country to find a reference to that publication. When it finds the reference, it will return the full **metadata record** describing that reference, including where it is physically located. If the publication were available in electronic format and stored in a repository, the user may be able to link directly to the publication file (**data**).

Gateways require considerable resources to maintain and are broad based. Nodes provide the framework for subject experts to maintain information (metadata records and data) while eliminating the need to maintain a separate and costly gateway. By following standards, the gateways are able to access the information located at each node and provide robust searching. The NPS GIS Data Store is the official NSDI Clearinghouse node for NPS, and will feed into the Geospatial One Stop (GOS).

The term clearinghouse may be used more loosely to mean a one-stop-shop website listing a suite of information items to choose. This is not the same as a clearinghouse gateway; however it serves a very important and critical function. Although a clearinghouse gateway provides sophisticated searching, they often provide too much information with little or no evaluation. These one-stop-shop websites, or pseudo clearinghouses, provides easy browsing, often supplemented with an evaluation of the quality.

4.2. SWAN System Architecture

SWAN offices are located at the Alaska Regional Office and are treated as a team under the same domain, where SWAN has its own team work files. The SWAN data manager has limited write access to the other servers, such as the SQL Server, Intranet server, and the data file server (X drive). Natural Resource park staff and the SWAN staff may access team work file servers distributed at the parks. This enables active projects to be worked on locally, while allowing park and network staff to easily share files. Completed projects and products are moved to the data file server and redistributed as read-only to all parks. Electronic files of final products are made to the NPS GIS Data Store for public distribution.

4.2.1. Team and User Project Files

Initial project files, for both short and long term, are typically stored on the work files server. The project files must:

- Contain a Project Organizer Document – a document containing contact information and an explanation the directory structure and what files are stored
- Conform to file naming standards (Appendix 4-2)
- Conform to documentation standards (Chapter 8)

The data manager will create a working directory for the projects. Project leaders will modify as needed, such as deleting directories that are not being used. Project directories may be “checked out” to project leaders located at parks, but should remain in a location accessible to the network data manager (i.e., on the park resources active projects directory). Project directories may also be “checked out” to contractors with the understanding a similar structure and requirements are required as part of the deliverables.

Project template directory structures for short-term projects and protocol development have been generated. Contact the Network data manager for the most current version.

4.2.2. Integration of Products to Master Data Libraries

Electronic version of the deliverables, as listed in the project tracking database, cooperative agreements, permits, study plans or other administrative records will be electronically stored by two online methods:

- 1) Project archive, where all records are “packaged” as a stand alone. Types of information to be included are listed in Chapter 1, section 1.5.
- 2) Product Library, where like products are stored with other products (such as the final report stored with other reports).

These files will be centrally stored in the master data libraries on the Alaska Regional Office (AKRO) central servers. Table 4-1 illustrates this general directory structure. Files will be reviewed for quality and standards prior to storing in the libraries.

Table 4-1: General directory structure and their roles for Master Libraries. Exact directory structure is subject to change.

Location	Description	Custodian
X:\GIS_Data	Storage of GIS data, metadata, and Theme Manager information. Best access to this information is through the Theme Manager in ArcGIS.	AKRO – GIS Team
X:\Libraries\Archived_Projects	Electronic archive of “packaged” project information. Includes administrative documents, raw, QC and verified versions of data. Extra misc. files have been deleted.	SWAN Data Manager
X:\Libraries\ \Maps \Photos \Photos_Aerial \Reports	Storage of final products with other like final products, such as final reports with all other reports.	SWAN Data Manager
Redistribution to Parks X:\GIS_Data X:\Libraries	Data stored in the above directories are redistributed to the parks for more direct and speedy access.	AKRO – GIS Team and Park IT staff
NR-GIS Data Store \GIS_Data \Bibliography_PDFs \Biodiversity_Data	A copy of the electronic deliverables will also reside on the WASO Data Stores	WASO

4.2.3. Integration of Long-Term Databases

Long-term datasets will likely be stored into a database server such as SQL Server administered by the AKRO GIS Team or on the Oracle database server administered by the Washington Office (WASO). The data sets maintained on the AKRO SQL Server may include information downloaded from non-NPS sources or information designed specifically for the SWAN monitoring program. The data sets maintained on the WASO Oracle database server may include servicewide information, such as NatureBIB, NPSpecies or NPSTORET.

4.2.4. Application Development

Applications are the “front end” or interface with which we interact with data. The “back end,” or the data themselves, may be stored on a database server as described above or as a stand-alone MS Access database.

SWAN will use one or both methods:

1. Desktop Microsoft Access front-end database – where the desktop provide the interface to forms, reports or standard queries. The “back end” data will either be a direct link to the database servers described above or another Microsoft Access database.
2. ColdFusion Web Application – where the Internet provides the interface to the data. The “back end” data are a direct link to the database servers described above.

Option 1 allows the ability to create a “stand alone” database and is used by NPSpecies and NPS STORET. Strict procedures for data check out and for data uploading are critical to maintain the integrity of the master database.

Option 2 provides direct access to the most current information stored in the centralized database and does not require any more than an internet connection. NPSpecies also provides a direct Internet connection to the master Oracle database.

4.2.5. Internet Servers

Applications, products, and links to the clearinghouses discussed in this chapter will be available on the SWAN internet website. Information in draft or subject to security will be made available on the SWAN intranet or extranet websites, where applicable. SWAN internet, intranet, and extranet servers are hosted on the WASO servers. AKRO provides an intranet server and may be used for development purposes. The AKRO servers are restricted by the Department of Interior and can only provide intranet services.

4.3. Local and Wide Area Network Infrastructures

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

4.4. IT Network Security

IT security is managed by the IT Team. SWAN staff will comply with all required security training and procedures advised by the AKRO, IT Team.

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

The regional IT staff participates in a number of security related teams. One staff person is dedicated to security, ensuring passwords have time restraints, and virus protection is current. Additional staff ensure firewalls and networks are secure.

4.5. Chapter Credits

David Oradei (Alaska Regional Office) and Joe Gregson (Washington Support Office).

4.6. References

For more information and current status on infrastructure and system architecture, the following references and websites are suggested.

Infrastructure

Department of Interior. 1991. DOI Department Manual 375 DM 12.

- <http://elips.doi.gov/elips/release/2915.htm>
- This chapter describes the development, application, and maintenance of information resources standards

Mortenson, D. 2005. Infrastructure to Support Integration and Dissemination. Austin, TX. (Powerpoint 11pp)

Tipton, W.H. 2002. Department of Interior, Office of the Secretary, Memorandum: Standardizing on Microsoft Desktop and Server Software.

Clearinghouses

- Federal Geographic Data Committee
 - o <http://www.fgdc.gov>
- Executive Order 12906 – EO defining a clearinghouse (for geospatial data)
 - o http://www.fgdc.gov/policyandplanning/executive_order
- Clearinghouse information
 - o <http://www.fgdc.gov/dataandservices/>
- Alaska Geographic Data Committee
 - o <http://agdc.usgs.gov>
- NPS Focus
 - o <http://npsfocus.nps.gov>
- NR-GIS
 - o <http://science.nature.nps.gov/nrgis/>
- NatureBIB
 - o <http://www.nature.nps.gov/nrbib/index.htm>

Security

Executive Office of the President of the United States. 2005. Information System Security Line of Business: Statutory Requirements Framework. 12 pp.

Tipton, W.H. 2004. OCIO Directive 2004-011. The Department of Interior Interim Web Governance Policy. Washington DC. 9pp.

- o Moratorium on web application development for serving data on a public website

Tipton, W.H. 2002. Securing Interior's Wide Area Networks. Washington DC. 3pp.

- o Memo that states NPS cannot contract web/database development to a contractor

Other System Security

- o <http://inside.nps.gov/waso/custommenu.cfm?lv=3&prg=308&id=2168>
- o <http://www.doi.gov/ocio/security/index.html>

5. Database Design

5.1. National Standards

Currently, there are no servicewide standards on database design. The I&M Program, however, recognized early on the need for database standards and detailed documentation. Database standards promote compatibility among data sets that will be aggregated and summarized in the future. Well thought-out standards help to encourage sound database design and facilitate interpretability of data sets.

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

- Database Specifications for I&M Studies
- Recommended Database Strategies, including the I&M Database Template
- Recommended Naming Standards
- Natural Resource Database Template (NRDT) Data Standards

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

5.2. Network Standards

In addition to the above database design guidelines, SWAN is developing application standards for the user interface. The user interface of the Microsoft Access databases developed should have a consistent appearance with standard functions. A draft of the Microsoft Access Database Application Specifications can be found in Appendix 5-1. It consists of standards on:

- General appearance
- Color palette and font
- Standard header and title bar for forms
- Switchboard or main menu layout
- Command buttons or menu buttons
- Standard report format
- Standard utilities, such as linking tools, backups, search, etc.

The WASO I&M Program has developed an intranet website where snippets of code, similar to those used in the SWAN standards, may be downloaded. SWAN will use and contribute to these codes, as applicable.

5.3. Introduction to Data Models

Communication is a vital part of developing a suitable data design for individual projects. A possible mechanism for this communication is by developing data models that can be developed in collaboration with project leaders and data managers. Data models combine diagrams with associated descriptions. Data modeling is completed in three stages: conceptual, logical, and physical.

5.3.1. Conceptual Data Models (CDM)

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 (QA/QC), 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 (see Figure 5-1).

CDMs should contain the following:

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

5.3.2. Logical Data Models (LDM)

An LDM is an abstract representation of a set of data entities and their relationship, usually including their key attributes. The LDM 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. See Figure 5-2.

LDMs are made up of five main components:

1. *Data entities* – distinct features, events, observations, and objects that are the building blocks of a data set, such as:
 - sample sites
 - sampling events
 - sampling units (transects, plots, etc.)
 - watersheds
 - species
 - habitat types

Information Management Data Models

Effective Date: December 2005

Conceptual Data Model: How the Data are Collected

Description: Provide a general description in layman's terms of what is actually going to happen. Person goes to the field and does what and with what equipment. Explain a little about the environment, such as very remote location with adverse weather, or nightly visits back to an office or other source of power.

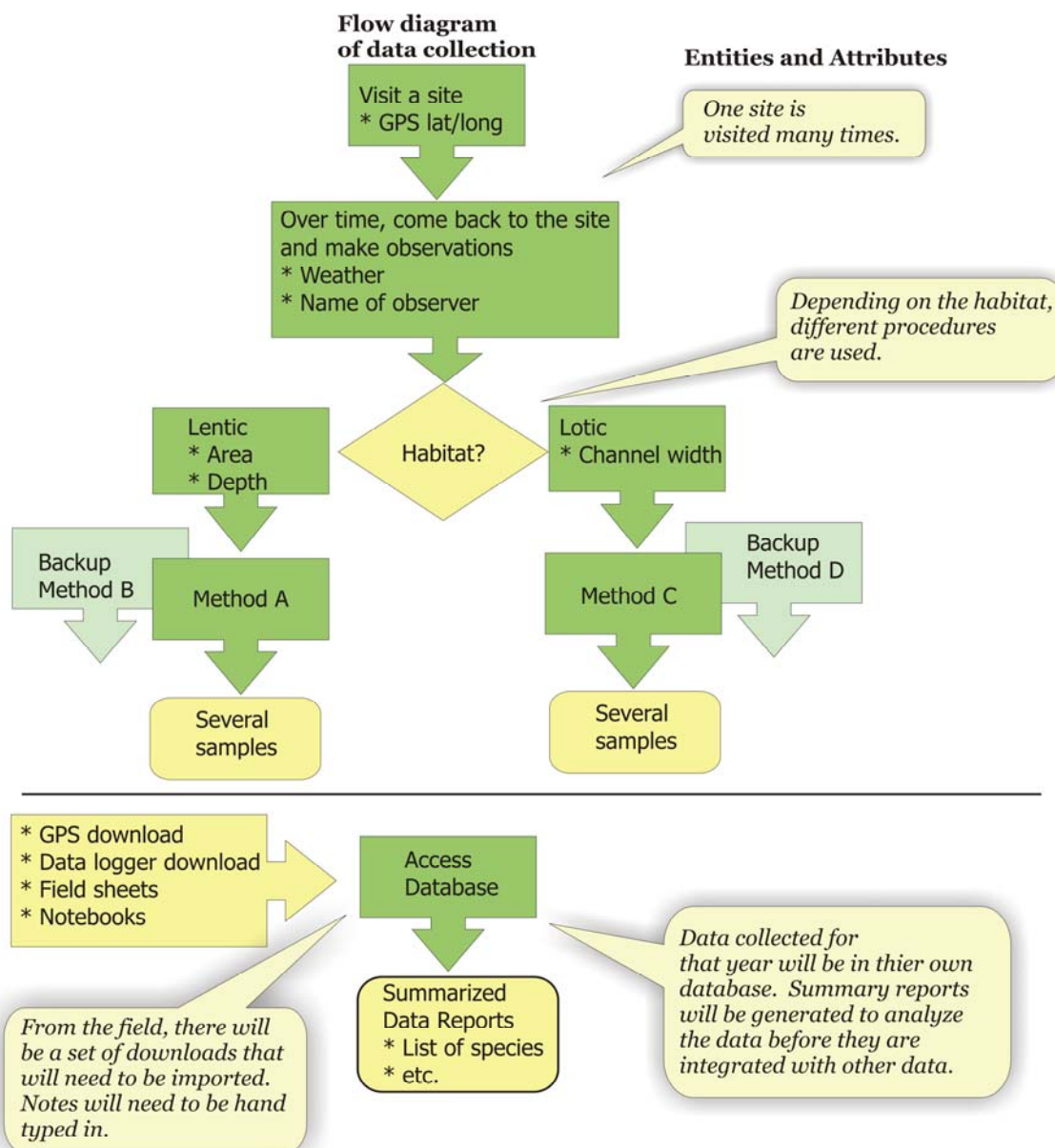


Figure 5-1: Conceptual Data Model

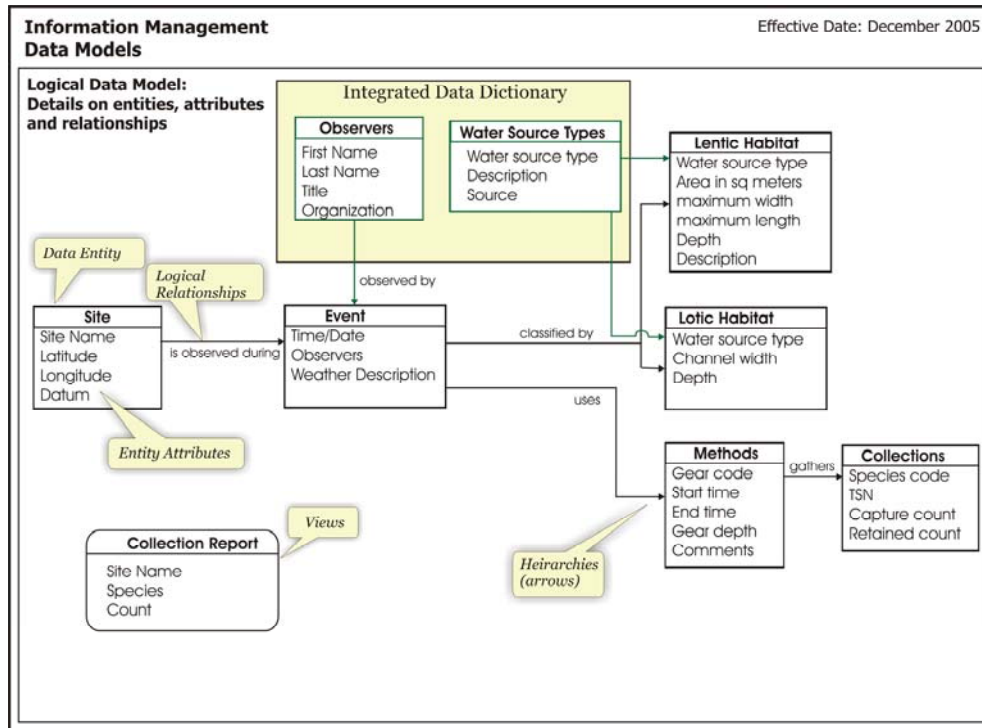


Figure 5-2: Logical Data Model

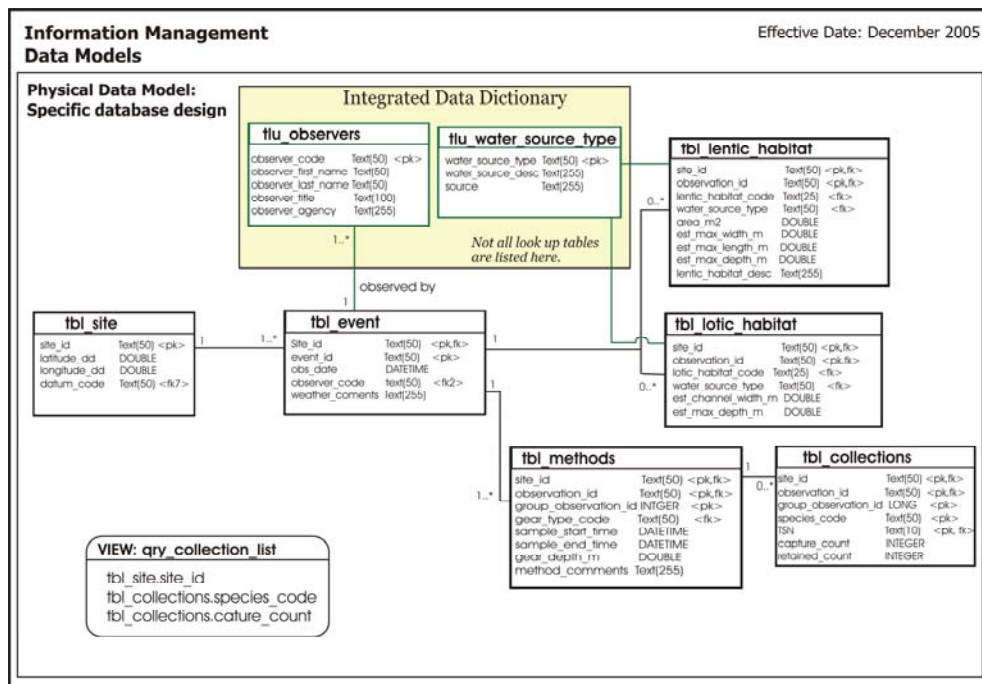


Figure 5-3: Physical Data Model

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

5.3.3. Physical Data Models (PDM)

The PDM is used to design the actual database, depicting data tables, fields and definitions, and relationships between tables. Although LDMS and PDMs are similar, the LDM only provides enough detail to communicate the information to be stored in the database. The PDM provides very specific details and definitions, such as primary keys and field types (Figure 5-3). The PDM should be applicable to MS Access or SQLServer.

5.3.4. Developing the Models

Four elements are required to develop the data models:

1. *People* – will need to be involved in the model development. Discussions may begin between a few, with review from a larger group or start with a large group working towards the specifics with a smaller group. Affected people may include the Network coordinator, scientists, partners, field crews, biometricians, and data managers.
2. *Protocols* – The protocols for the vital sign will provide the greatest substance to the models. The protocols provide the goals, objectives, methods, standards, analysis, and reporting.
3. *Reference materials* – Reference materials such as field forms, drawings, mock-up reports, references to classifications to be used will play a significant role in the data models.
4. *Frequent interactions* - Frequent discussions between the project 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 models are not flexible by themselves. It is how these techniques are applied that count. Not everything can be thought of from the start. To support change, data modeling should be iterative and interactive.

5.4. Fundamental Database Structure

The Natural Resource Database Template (NRDT) provides an example of the fundamentals for database structure to be used by the I&M Program. These fundamentals are as follows:

1. *Mandatory (core) tables and fields* – These tables and fields are used to manage the information describing the “who, where and when” of project data. They contain critical data fields that are standardized with regard to data types, field names, and domain ranges. For example: tbl_Sites, tbl_Locations, tbl_Events.
2. *Mandatory if Applicable (master look-up tables) tables and fields* – Described in more detail below (Integrated Data Dictionary), these tables contain commonly used attributes, such as park code, and are centrally located to minimize redundant information in each database.
3. *Optional (project specific) tables and fields* – These are tables and fields that are needed for a specific project or protocol, but will not likely be used in other databases or in integration.

The NRDT has been extensively reviewed by the I&M Program data managers and has been applied to some of the biological I&M projects. The fundamentals of this database structure will be applied to SWAN databases where applicable.

5.5. Integrated Data Dictionary

SWAN will be developing a master look-up table database, or Integrated Data Dictionary (also referred to as a Buffet of Fields and Tables (BOFAT) by WASO), that

will allow databases across disciplines to stay as consistent as possible. This data dictionary will contain standard attributes and values such as park code, watershed codes, and vegetation classifications.

As the protocols are developed, databases should be able to be integrated across disciplines. These databases must contain key linking attributes. It is unknown at this time what these will be for SWAN. The physical features in SWAN park units, such as lakes, glaciers, or mountains, are numerous, ill-defined, and largely unnamed. Park code, latitude and longitude, watershed basins, USGS quadrangles, township/range, sampling blocks/segments, or a combination are being considered.

5.6. Documenting Databases

Database documentation is more fully described in Chapter 8 of this Plan.

5.7. Chapter Credits

John Boetsch (Olympic National Park), Jeff Cotterman (Southwest Alaska Network), and Angie Southwould (Alaska Regional Office).

5.8. References

Southwould, Angie. 2002. NPS database specifications for Inventory and Monitoring Studies. 13 pp.

Southwould, Angie. 2002. NPS recommended database strategies, including the I&M database template 53 pp.

Stup, Richard. 2001. Dairy farm business, standard operating procedures: a writing guide. Pennsylvania State University, University Park, PA. 8pp.

Washington Support Office, Inventory and Monitoring Program. 2005. Recommended naming standards for inventory and monitoring databases. National Park Service. Fort Collins, CO. 14 pg.

Natural Resource Database Template (NRDT)

- <http://science.nature.nps.gov/im/apps/template/index.cfm>
- A set of relational database tables in Microsoft Access that standardize core tables and fields.

I&M Program Data Management Website

- <http://science.nature.nps.gov/im/datamgmt/index.cfm>
Provides many links to service-wide guidance on data design
- http://science.nature.nps.gov/im/datamgmt/data_tools.cfm (For Snippet and Tools Code)

6. Data Acquisition and Processing

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

Steps for data acquisition are as follows:

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

6.1. National Standards

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

6.2. Network Standards

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

6.3. Data Sources

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

1. **Programmatic Data** – any data produced from projects that are initiated (funded) by the I&M Program or projects that in some way involve the I&M Program.
2. **Non-programmatic Data** – includes data collected from NPS sources and data produced by external non-NPS sources.
 - o **Non-programmatic NPS Data** – any data produced by the NPS that did not involve the inventory and monitoring program, such as park visitor use information.
 - o **Non-programmatic External Data** – any data produced by agencies or institutions other than NPS, such as forest health

inventories completed by the US Forest Service and the Alaska Department of Natural Resources.

In all likelihood, the vital signs will use a combination of both programmatic and non-programmatic data to meet the goals of the I&M Program. Please refer to Table 6-1 at the end of this chapter for a listing of data acquisition for each vital sign.

6.4. Data Discovery

Data discovery, or data mining, is the process of searching for existing data/information that may be useful to the I&M Program's mission and that is related to the natural resources of the network parks. This is a continual process that begins with the collection of background information and data that assist in the development of project methodologies and designs. The process involves reviewing many different sources for varying types of information. Data mining is an important part of any I&M project, but certain vital-signs monitoring projects depend largely on data acquired via data mining endeavors (e.g., insect outbreak, air quality). Many of the data sources listed below are accessible via the internet, but some require visiting local archives, research or academic institutions, museums, or local parks to conduct the searches.

Bibliographic/Literature

- National NPS Databases (e.g. NatureBib)
- Online literature databases (e.g. First Search or Biosis)
- Alaska Resources Library and Information Services (ARLIS)
- Park archives through ANCS+

Geographic Data

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

Natural Resources Data

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

Relevant information collected during a data discovery process is maintained at the Network either electronically or in hard copy format depending on how it was collected. Any data collected during data discovery should be accompanied by as much documentation (metadata) as possible. Geographic data sets collected during this process are documented with FGDC compliant metadata.

Table 6-1. Vital signs data collection information at a glance.

9/7/2006

Project	Vital Sign	Parks					Measures	Partner Agencies	Primary Data Source	Primary Repository	Frequency	Collection Tools & Mechanisms													Products							Key Identifier(s)			
		ALAG	ANIA	KATM	KEFJ	LACL						Data Harvest	Data Sheet	GPS	Data Logger	Camera	Imagery	Vouchers	Genetics	Lab Work	Audio	Video	Radio telemetry	Other	Reports	GIS	Tabular	Photos	Geo-Imagery	Vouchers	Audio		Video	Other	
Weather and Climate	1 Weather and Climate			X	X	X	Air temperature, precipitation, surface wind, solar radiation, relative humidity, snow depth.	SWAN & WRCC-DRI	WRCC and SWAN weather stations; weather models	WRCC	Continuous data collected. Trend analysis 5 years.	X				X	X											X	X						Site name, lat/long
Landscape Dynamics & Terrestrial Vegetation	2 Glacier Extent			X	X	X	Areal extent of glacial ice, photo points of selected glaciers.	SWAN	Satellite imagery; photos	SWAN central data	10 years.					X	X											X	X					TBD	
	3 Landscape Processes	X	X	X	X	X	Lake freezeup and ice breakup dates, duration of snow cover, length of growing season onset, duration and relative biomass of vegetation productivity.	SWAN	MODIS imagery	SWAN central data and Alaska Geographic Data Clearinghouse (AGDC)	Nearly continuous; Document annually. Trend analysis 10 years.						X										X	X		X				TBD (Product Code, Dates, Process Code)	
	4 Land Cover and Land Use	X	X	X	X	X	Land use patterns	SWAN	Various multispectral satellite imagery; aerial photos	SWAN central data & Alaska Geographic Data Clearinghouse (AGDC)	5-10 year intervals						X									X	X		X				TBD		
	5 Landscape Patterns, Vegetation Composition and Structure		X	X	X	X	Distribution, extent, species composition, cover, age class/regeneration, physiognomy	SWAN & USFS	Various multispectral satellite imagery; aerial photos; ground sampling	SWAN central data & Alaska Geographic Data Clearinghouse (AGDC)	5-10 year intervals; Trend analysis 3-5 years for 10 years; 7-10 thereafter.		X	X		X	X									X	X		X	X				TBD	
	6 Sensitive Vegetation Communities		X	X	X	X	Distribution, extent, species composition, cover, age class/regeneration, physiognomy	SWAN	field sampling	SWAN central data	3-5 years for 10 years; 7-10 thereafter		X	X		X										X	X	X	X					TBD	
	7 Volcanic & Earthquake Activity	E	E	E	E	E	Frequency, intensity	AVO & WCATWC	AVO & WCATWC Reports &??	AVO & WCATWC	Continuous data collected at AVO; Event driven for Network	X														X								Volcano name; lat/long	
	8 Insect Outbreaks	X	X	X	X	X	Species, timing and areal extent	USFS & ADNR-DOF	ADNR-DOF Forest Health Reports and GIS data.	ADNR-DOF	Annual documentation ; Trend analysis 1, 5, 10 year intervals.	X														X	X							GIS auto ID, USGS quadrangles	

Table 6-1. Vital signs data collection information at a glance.

9/7/2006

Project	Vital Sign	Parks					Measures	Partner Agencies	Primary Data Source	Primary Repository	Frequency	Collection Tools & Mechanisms													Products							Key Identifier(s)			
		ALAG	ANIA	KATM	KEFJ	LACL						Data Harvest	Data Sheet	GPS	Data Logger	Camera	Imagery	Vouchers	Genetics	Lab Work	Audio	Video	Radio telemetry	Other	Reports	GIS	Tabular	Photos	Geo-Imagery	Vouchers	Audio		Video	Other	
Marine Nearshore	9 Geomorphic Coastal Change			X	X	X	Position of the mean water line, position of top and toe of the bluff, position of foreshore and backshore vegetation, sediment type and grain size.	SWAN, GEM, KBERR, CIRCAC, & USGS-ACS	Imagery - LIDAR ; ground survey	TBD	10-12 years	X	X	X		X					X					X	X	X					X		GIS auto ID (?)
	10 Kelp & Eelgrass			X	X	X	Canopy coverage and distribution	SWAN, GEM, KBERR, CIRCAC, & USGS-ACS	aerial-flight survey, video	TBD	Annual. Trend analysis 5-10 years.	X	X	X						X						X	X					X		TBD (Sample block)	
	11 Marine Intertidal Invertebrates			X	X	X	Composition, relative abundance, and contaminants in mussels	SWAN, GEM, KBERR, CIRCAC, & USGS-ACS	field sampling	TBD	Annual. Trend analysis 5-10 years.	X	X	X					X						X	X								TBD (Sample block)	
	12 Seabirds			X	X	X	Seasonal abundance	SWAN, GEM, KBERR, CIRCAC, & USGS-ACS	observations from boat	TBD	Annual. Trend analysis 5-10 years.	X	X	X												X								TBD (Sample block)	
	13 Black Oystercatcher			X	X		Relative density and diet	SWAN, GEM, KBERR, CIRCAC, & USGS-ACS	aerial-flight survey, ground survey	TBD	Annual. Trend analysis 5-10 years.		X	X			X								X	X								TBD (Sample block, Shore segment)	
	14 Sea Otter			X	X	X	Relative abundance, age-specific survival, and diet	SWAN, GEM, KBERR, CIRCAC, & USGS-ACS	aerial-flight survey	TBD	Annual. Trend analysis 5-10 years.		X	X	X	X									X	X									TBD (Sample block)
	15 Marine Water Quality		X	X	X	X	Temperature, conductivity, dissolved oxygen, pH, salinity	SWAN, GEM, KBERR, CIRCAC, & USGS-ACS	Satellites; field sampling	SWAN central data & STORET GEM?	Seasonal and annual	X	X	X	X		X					X	X		X									TBD (Sample block, site)	
	16 Harbor Seal			X	X	X	Occurrence and abundance at haulouts	NMFS	aerial-flight survey, photos	Polar Ecosystems Program (PEP) at NMFS-NMML	Rotates every 5 years.	X				X						X		X	X	X								TBD	

Table 6-1. Vital signs data collection information at a glance.

9/7/2006

Project	Vital Sign	Parks					Measures	Partner Agencies	Primary Data Source or Method	Primary Repository	Frequency	Collection Tools & Mechanisms														Products								Key Identifier(s)
		ALAG	ANIA	KATM	KEFJ	LACL						Data Harvest	Data Sheet	GPS	Data Logger	Camera	Imagery	Vouchers	Genetics	Lab Work	Audio	Video	Radio telemetry	Other	Reports	GIS	Tabular	Photos	Geo-Imagery	Vouchers	Audio	Video	Other	
Lakes, Rivers & Fish	17 Surface hydrology	X	X	X	X	X	River discharge and lake level	SWAN	gauging stations, data loggers, ground work; aerial photos, modeling	SWAN central data	Annual. Trend analysis 5 years.	X	X	X	X										X	X								TBD (watershed)
	18 Water Quality	X	X	X	X	X	DO, Temperature, pH, and specific conductance	SWAN	Field sampling	SWAN central data & STORET	5 years	X	X	X	X					X					X	X	X							TBD (watershed)
	19 Resident Lake Fish	X	X	X	X	X	Species richness, species turnover, exotic species	SWAN	gill net, beach seines	SWAN central data	3-5 years		X	X	X?				X?							X	X	X						TBD (watershed, lake name)
	20 Salmon	X		X		X	Abundance of adult spawners, growth rates and distribution, timing of spawning runs, freshwater residence time, body condition	SWAN & ADFG	counting towers, aerial-flight survey	ADFG & FWS	Annual. Trend analysis 5 years.	X	X	X												X	X	X						TBD (watershed, ADFG Anad. Stream number)
Terrestrial Animals	21 Brown Bear	X	X	X		X	Distribution and abundance	SWAN & ADFG	aerial-flight transects, incident reports	TBD	5-10 years)	X	X	X	X	X?											X	X	X					TBD
	22 Wolf	X	X	X	X	X	Distribution, abundance	SWAN & ADFG	aerial-flight survey	TBD	3-5 years	X	X	X	X	X?											X	X						TBD
	23 Wolverine	X	X	X	X	X	Distribution, abundance	SWAN & ADFG	aerial-flight survey	TBD	3-5 years	X	X	X	X	X?											X	X						TBD
	24 Moose			X		X	Distribution, abundance, age-sex combination	SWAN & ADFG	aerial-flight survey; sightability model	TBD	3-5 years	X	X	X	X?											X	X	X						TBD (game management units)
	25 Caribou			X		X	Distribution and abundance	ADFG	aerial-flight survey, radio	TBD	Annual. Trend analysis 5 years.	X	X	X	X?						X												TBD (game management units)	
	26 River Otter (coastal)			X	X	X	Distribution, abundance	SWAN & GEM	latrine survey	TBD	Annual. Trend analysis 5-10 years.		X	X	X?				X							X	X						TBD	
	27 Bald Eagle	X	X	X	X	X	Nest occupancy and productivity	SWAN & USFWS	aerial-flight survey	TBD	1-5 years	X		X?	X?												X	X						USGS Quadrangle, GIS auto ID, (maybe nest)

Table 6-1. Vital signs data collection information at a glance.

9/7/2006

Project	Vital Sign	Parks					Measures	Partner Agencies	Primary Data Source	Primary Repository	Frequency	Collection Tools & Mechanisms													Products							Key Identifier(s)		
		ALAG	ANIA	KATM	KEFJ	LACL						Data Harvest	Data Sheet	GPS	Data Logger	Camera	Imagery	Vouchers	Genetics	Lab Work	Audio	Video	Radio telemetry	Other	Reports	GIS	Tabular	Photos	Geo-Imagery	Vouchers	Audio		Video	Other
Human Activities	28 Resource Harvest	X	X	X		X	Community-based surveys, species harvested, volume harvested, permits	SWAN & USFWS	ADFG-Sport Fish, Comm fish; NPS-Subsistence Div.	ADFG & NPS-Subsistence Div.	Annual. Trend analysis 3-5 years.	X													X	X	X							TBD
	29 Visitor Use	X	X	X	X	X	Type, level, distribution, impacts	SWAN & Parks	staff observations; business permits	SWAN	Annual. Trend analysis 3-5 years.	X	X												X		X							TBD (Park code)
	30 Exotic Plants	X	X	X	X	X	Occurrence, distribution of non-indigenous plants	SWAN & AKRO	Alien Plant Control and Management Database (APCAM); Alaska Exotic Plants Information Clearinghouse (AKEPIC)	APCAM; AKEPIC	Annual review	X													X	X	X							TBD
	31 Visibility and Particulate Matter		X			X	IMPROVE suite for visibility and fine particulates	USFWS & Colorado State University	Annual downloads of data summaries and reports from IMPROVE website; Possible NADP site(?)	Interagency Monitoring of Protected Visual Environments (IMPROVE)	Continuous data collected. Trend analysis 5 years.	X			X?					X?					X		X						Site code (TUXE1, SIME1)	

Source: Information extracted from the SWAN protocol development statements (PDS), 2006 and the WASO I&M Vital Signs database, 2006. Please see the respective PDS for more current and complete information.

6.5. Data Harvesting

Harvesting data from other sources should be as standardized as possible.

Considerations for harvesting are as follows:

- Does the data source organization know of the I&M Program and its data needs?
- Does a Memorandum of Understanding (MOU) need to be in place?
- What is the contingency if this data source is no longer available?
- Can downloads or requests for data be consistently exported/imported and scheduled?
- How will the downloaded data be stored and integrated into the I&M Program?
- How should errors be addressed?
- Is the data source organization interested in integrating I&M Program data, where appropriate?
- Is the documentation adequate?

Vital signs monitored through data harvesting should address these questions in their data management protocols.

Data harvested from another agency may not match the Network's needs exactly. Boundaries used for summary data often do not coincide with park boundaries or specific parameters may not be currently collected. In addition, data may not be meet SWAN specifications to reach the vital sign objectives. Project leaders are encouraged to work with these agencies to incorporate the Network's needs as much as possible. Money can be saved and more efficient research can be done by improving existing systems.

6.6. Data Collection for Field Studies

Biological inventories and monitoring projects are the most common examples of field studies conducted by the SWAN I&M Program. The project leader is responsible for ensuring that data collection, data entry, verification, storage, and archiving for all field projects are consistent with the SWAN I&M Program standards. In addition to general standard operating procedures (SOPs) that define network-wide requirements, protocol specific SOPs are developed detailing procedures and/or methodologies specific to each protocol. The data manager will work closely with the principal investigator and Network staff to develop these guidelines and methodologies connected with data collection, storage, and maintenance of project data. This may range from detailing the proper usage of data entry forms or databases to outlining calibration procedures for automated data loggers. Refer to the individual protocols for protocol-specific SOPs.

Listed below are some of the tools available for field data collection. This list serves as a guide for consideration as the vital signs protocols are developed.

Refer to the protocol-specific SOPs for details on how these tools may apply to individual projects.

- **Field Forms** – are the most common method of recording field data. Use a formatted, project-specific data sheet as opposed to a field notebook. Field notebooks are important for entering additional notes and observations. It is recommended to use acid-free paper to prevent fading and subsequent data loss. Some circumstances may warrant the use of paper and writing implements that can withstand moisture, dust, and other extreme environmental conditions (e.g., “Rite in the rain” paper). Standardized data sheets that identify the information to be recorded and forms that reflect the design of the computer data entry interface will help ensure that all relevant information is recorded and subsequent data entry errors are minimized. Data sheets should contain as much basic preprinted project information as possible and sufficient space for recording relevant metadata such as date, collectors, weather conditions, etc. They should clearly specify all required information, using examples where needed to ensure that the proper data are recorded. Data recorders should adhere to the following guidelines:
 - o All information added to the data sheet must be printed and clearly legible.
 - o If alterations to the information are necessary, the original information should be crossed out with a single line and the new information written next to the original entry. Information should never be erased and old information should not be overwritten.

Upon return from the field, copies of all original data sheets should be made and checked for legibility and completeness (i.e., no data cut off at the edges). The original datasheets should be used for data entry after which they will be stored in the fire proof cabinet. Copies of original datasheets should be archived offsite.

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

- **Cameras** – Photographs provide an excellent visual record of field visits. Cameras are useful for capturing photo point records of long-term study sites. They also serve well for automated data collection by remotely recording information using web cams or trip cameras.
- **Field Computers** – increase data collection and data entry efficiency by eliminating the need for paper field forms. Data can be downloaded directly from the field computers to the office desktops thereby eliminating the data entry step. Fewer chances for error because QA/QC checks can be built into the field database. Could be inefficient if copious amounts of notes or comments need to be recorded in the field.

Field computers, however, are subject to environmental constraints such as heat, dust, and moisture. When handheld computers are used for data entry in the field, the data should be downloaded daily to avoid potential loss of information. No field computer should be used unless it is equipped with a removable flash memory card to store field data in case of a catastrophic failure of the field unit. Batteries should be checked prior to a data collection trip, and they should be charged at the end of every field day. Also, in case the unit becomes inoperable in the field, printed data sheets should always accompany field teams on data collection trips.

- o **Palm-top computers (PDAs)** – the small size and relative low cost of these devices make them attractive options for entering field data. Good for small field projects but not powerful enough for large data intensive field projects. PDAs can be ruggedized fairly easily and at a relatively low cost. Most run either Windows CE or Palm operating systems that may require additional processing/programming to transfer/create the database structure in the field units.
- o **Tablet PCs** – same properties as most laptops and provide the user with the convenience of a touch screen interface. They are bulkier, more expensive and harder to ruggedize than PDAs but are more powerful as well. Good for field projects that are very data intensive. Because these units run Windows XP (Tablet Edition) the project database can be directly transferred from desktop units to field units without additional programming steps.
- **Automated Data Loggers** – mainly used to collect ambient information such as air or water quality information. These devices can be calibrated and programmed to automatically record data and store them for later download directly to a computer, thereby eliminating the possibility for manual data entry errors. Must be properly calibrated and maintained so field crews must receive proper training and SOPs outlining the calibration/unit maintenance

procedures. SOPs are currently in development and will be distributed to all field crews using these devices.

- o **Permanently deployed devices** – often cost prohibitive. Data from these devices must be retrieved and batteries changed on a regular basis. These intervals should be defined in the protocol.
- o **Portable hand-held devices** – deployed for sampling only during site visits. Generally less expensive than units that are permanently deployed in the field. Prior to and following field visits device components should be inventoried to ensure that all necessary equipment is accounted for.
- **Geographic Positioning System (GPS) Units** – Two types of GPS units are often used during field work in SWAN parks to collect location information.
 - o **Handheld Units (e.g., Garmin)** – good for collecting general position information. Not recommended for obtaining high accuracy location information.
 - o **Trimble GPS Receivers** – good for collecting highly accurate (sub-meter) location information.

6.7. Data Collection for Remote Sensing

Because of the remoteness and low accessibility of many areas within SWAN parks, remote sensing will be a primary means of collecting landscape data. Considerations for selecting remote sensing imagery are as follows:

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

Each remote sensing product is unique. It is imperative the user fully understands the product he or she is using. Products should be accompanied with well documented metadata. Imagery not rectified cannot be used for measuring distance or area, but may provide a low cost and timely overview of the landscape. Any protocols involving remote sensing should involve consultation with a professional remote-sensing specialist. Table 6-2 can be used as a guide to understanding the trade-off between accuracy and costs among different imagery sources.

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

Table 6-2. Satellite resolution, swath width, area, cost and history.

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

* cost uses the multi-spectral unit specifications

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

T requires tasking

Base Cartography

AKRO is currently developing rectified ortho-imagery base products that will qualify as topographic maps. Resolution is expected to be 1 meter or better. This project is likely to be three or more years from completion (from 2005).

Reoccurrence of this high resolution and quality of data is likely to be 20 or more years. These data will provide the network parks with comprehensive cartographic base as well as high resolution imagery for interpretation.

Moderate Resolution Imaging Spectroradiometer (MODIS)

MODIS provides large landscape and frequent views. It has a viewing swath of 2,330 km and views the entire surface of the Earth every one to two days.

Resolution is approximately 250 meters. MODIS imagery will be used as a means to collect data for several vital signs (currently under development).

Landsat Thematic Mapper (TM)

Legacy satellite data, such as data collected from the Landsat TM, will be used as a needed. Imagery varies from one park to the other in classification. Resolution is 30 meters. A significant archive of such imagery is available through the University of Alaska in Fairbanks.

Aerial Photographs

Legacy aerial photos will be used on an as needed basis. Photos available for SWAN include black and white photos from 1950 and color infrared (CIR) photos from the 1980s. Additional coastal photographs are taken in a few SWAN park units. SWAN anticipates these photos will provide critical information over time and is acquiring copies of existing photos of key areas. A protocol for any new photos taken should be developed prior to acquisition. Appendix 6-1 provides a sample scope of work and specifications for AKRO acquisitions.

6.8. Changes to Data Collection Procedures

Changes to data collection procedures are made based on valid reasons for altering the methodologies. Most issues will be identified during the design and testing stages of the protocols and changes will be implemented prior to the collection of field data. The protocol development process includes attempting to identify and address any foreseeable issues that might occur with data acquisition and processing. Unforeseen issues may arise after data collection has begun that require revision of procedures/protocols. Improvements in technology may also require changes to procedures. Significant changes to the protocols must be approved by the project leader, the Network coordinator, and the data manager. The Network coordinator must evaluate the proposed changes and determine if additional peer review is required before accepting the modifications.

Changes to protocols and associated data collection procedures may also occur as a result of scheduled program reviews. During the review, data may be analyzed to determine if the current protocol is meeting stated objectives. If it is determined that the protocol has not achieved the desired results then changes should be recommended. All changes must be carefully documented within the SOPs and any associated databases

6.9. Data Compilation, Processing and Integration

All data, whether collected in the field by the I&M Program or downloaded from another organization, will need to be compiled and processed. This compilation should follow the data management guidelines developed for the specific data type (see Appendix 3-1). For example, tabular data will need to be entered or imported into a pre-approved MS Access database. QA/QC procedures are more fully describe in Chapter 7.

After a single season or milestone of data is compiled and QA/QC is completed, these data should be integrated with master datasets. Some extraction of information may be required. For example, the exotic species vital signs may have their own datasets. A subset of data may need to be extracted to integrate and update NPSpecies.

6.10. Chapter Credits

Geoffrey Sanders (National Capital Region Network), and Jess Grundblatt (Alaska Regional Office).

7. Quality Assurance /Quality Control

The success of I&M networks is dependent on the quality of the data they collect and manage. Analyses performed to detect ecological trends or patterns require data that are recorded properly and have acceptable precision and minimal bias. Poor quality data can limit detection of subtle changes in ecosystem patterns and processes, can lead to incorrect interpretations and conclusions, and could greatly compromise the credibility of the I&M Program. To ensure that SWAN produces and maintains data of the highest possible quality, procedures have been established to identify and minimize errors at each stage of the data lifecycle (Figure 7-1). Data quality related to sampling design is covered in Chapter 4 of the SWAN Vital Signs Monitoring Plan (Bennett et al. 2006) and the individual protocols and is not discussed here.

7.1. NPS Standards

The NPS Director's Order #11B, "Ensuring Quality of Information Disseminated by the National Park Service," was issued in order to promote data quality (National Park Service 2002). It defines '*quality*' as incorporating three key components: objectivity, utility, and integrity.

Objectivity consists of: 1) *presentation*, which focuses on whether disseminated information is being presented in an accurate, clear, complete, and unbiased manner within a proper context; and 2) *substance*, which focuses on ensuring accurate, usable, and reliable information.

Utility refers to the usefulness of the information to its intended users.

Integrity refers to the security of information; e.g., protection from unauthorized access or revision to ensure that the information is not compromised through corruption or falsification.

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

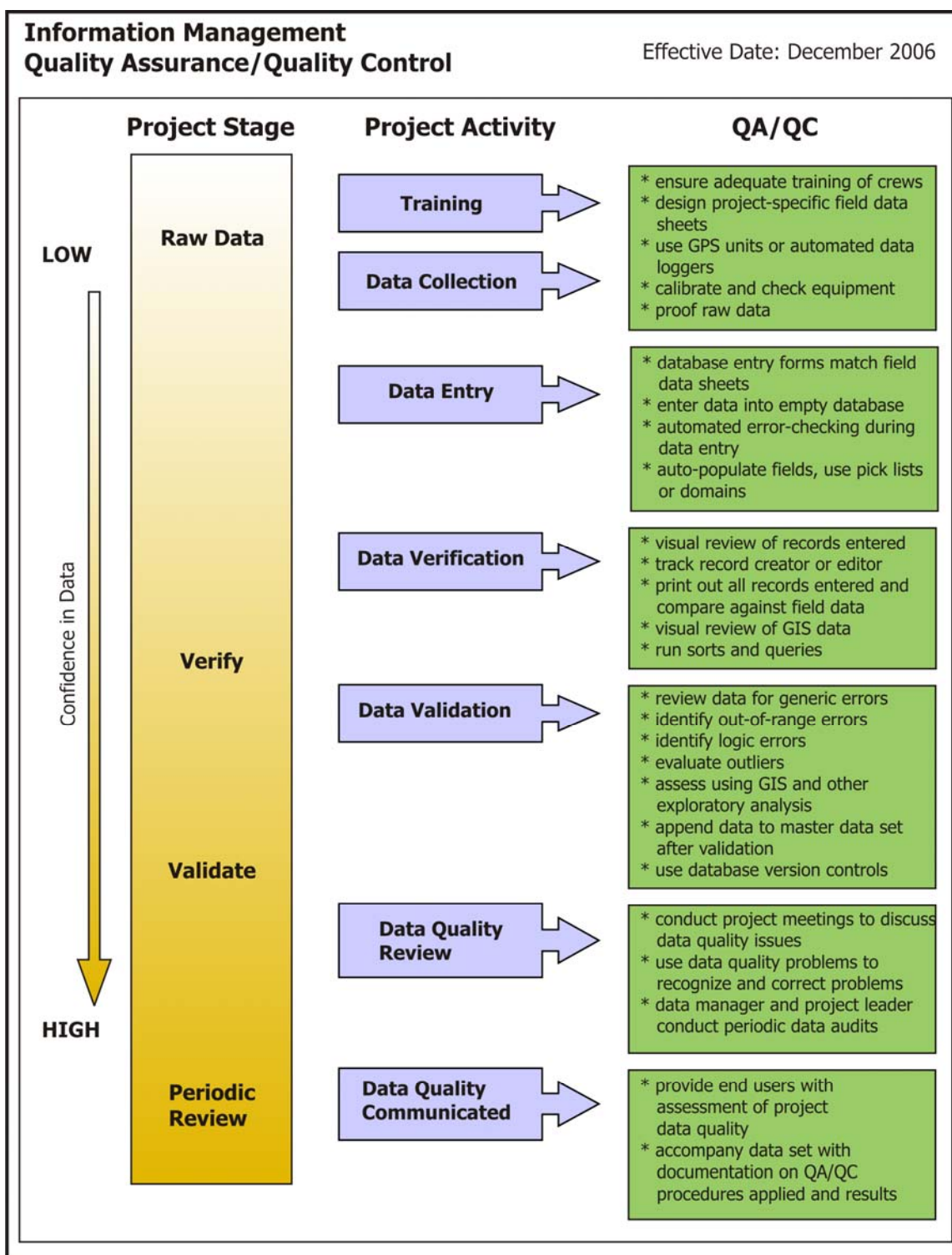


Figure 7-1. Schematic of QA/QC controls during project stages.

7.2. Quality Assurance and Quality Control Mechanisms

Palmer (2003) defines *quality assurance* (QA) as “an integrated system of management activities involving planning, implementation, documentation, assessment, reporting, and quality improvement to ensure that a process, item, or service is of the type and quality needed and expected by the consumer.” He defines *quality control* (QC) as “a system of technical activities to measure the attributes and performance of a process, item, or service relative to defined standards.” Quality assurance procedures maintain quality throughout all stages of data development. Quality control procedures monitor or evaluate the resulting data products.

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

- *Errors of commission* include those caused by data entry or transcription errors, or malfunctioning equipment. They are common, fairly easy to identify, and can be effectively reduced up front with appropriate QA mechanisms built into the data acquisition process, as well as QC procedures applied after the data have been acquired.
- *Errors of omission* often include insufficient documentation of legitimate data values, which could affect the interpretation of those values. These errors may be harder to detect and correct, but many of these errors should be revealed by rigorous QC procedures.

QA/QC procedures applied to ecological data include four procedural areas (or activities), ranging from simple to sophisticated, and inexpensive to costly:

- 1) defining and enforcing standards for electronic formats, locally defined codes, measurement units, and metadata
- 2) checking for unusual or unreasonable patterns in data
- 3) checking for comparability of values among data sets
- 4) assessing overall data quality

Much QA/QC work involves the first activity (defining and enforcing...), which begins with data design and continues through acquisition, entry, metadata development, and archiving. The progression from raw data to verified data to validated data implies increasing confidence in the quality of the data through time.

7.3. Roles and Responsibilities

Quality assurance methods should be in place at the inception of any project and continue through all project stages to final archiving of the data set. It is essential that each member of the team have a stake in data quality, and is responsible for the quality of the results generated from his or her tasks. In particular,

the data manager is responsible for:

- developing protocols and SOPs, in collaboration with the project leader, to ensure data quality
- making project leaders, technicians, etc., aware of the established procedures and enforcing adherence to them
- evaluating the quality of all data and information against NPS standards before dissemination outside the network
- performing periodic data audits and quality control checks to monitor and improve the data quality program.

Project leaders must:

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

Project technicians must:

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

7.4. Goals and Objectives

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

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

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

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

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

monitoring vital sign, some general concepts apply to all network projects. The general QA/QC procedures presented in this plan were primarily adapted from the Draft Data Management Protocol (Tessler & Gregson 1997) and the ideas contained in Michener and Brunt (2000). These general guidelines will ensure that all data collected are checked for integrity before being integrated into the monitoring program databases.

7.5. Data Collection

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

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

7.5.1. Methods for Reducing Collection Errors

Ensure that field crews receive proper training.

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

Data collection in SWAN parks is expensive and logistically challenging. A pre-season test of the protocols in the field at a local area is encouraged. This should include NPS staff and contractors. This allows for adjustments to be made and priorities to be determined.

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

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

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

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

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

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

Use an electronic device for data collection whenever possible.

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

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

Use automated data loggers where appropriate.

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

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

Be organized and keep a log.

Organization is the key to good data collection methods. Maintaining a log of important decisions and events will help clarify information and contribute to an accurate report.

Perform quantitative assessments of data quality.

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

7.6. Data Entry

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

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

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

7.6.1 Methods for Reducing Data Entry Errors

Enter or download data in a timely manner.

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

Design efficient data entry forms.

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

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

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

Track record creation and edit details in the database.

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

Build automated error checking features into the database.

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

- *Auto-filled fields.* Whenever possible, the data in a field should be auto-filled by the computer. For example, if a location ID is comprised of a park code, project code, and a unique number, those elements are automatically inserted into the location ID field, ensuring that the record always contains a unique identifier.
- *Range limits.* Where the appropriate values for a particular field span a finite range, the data entry program can check the entered value against the specified minimum and maximum values for that parameter. When a value is outside the accepted range, a warning

message appears and asks the user to reenter a valid value. For some fields, values outside a specified 'normal' range may be acceptable. In this case, the warning message asks the user to verify the entry before continuing.

- *Pick lists.* The data entry application may also use pop-up pick lists for standardized text items where spelling errors can occur. For example, rather than typing in a species code or name (where a misspelling generates a new species in the database), the code or name is selected from a list of valid species codes or scientific names and automatically entered into the species field. A pick list may also be used when only certain entries are acceptable. Lists are not appropriate for all written fields but should be used when appropriate.
- *Unique constraints.* Duplicate and incorrect data entry can often be caught with the application of unique constraints on data entry fields. These constraints are particularly useful when importing data from other applications.

Provide a clean, organized work environment.

Desktop space near the computer should be free of clutter and distractions. There should be enough space for two stacks of paper documents, one from which data are being entered and one from which data have been entered. A pad or notebook should also be available for making notes.

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

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

7.7. Verification and Validation Procedures

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

The data manager and project leader collaboratively establish SOPs for verification and validation and the project leader or designee will validate the data after verification is complete. The project leader is also responsible for reviewing all data products and reports before they are released outside the network. The data and project leaders will evaluate the results of verification and validation and determine any procedural or data form revisions that may be

indicated by the results. Technicians will follow the SOPs for verification of data, make required changes, and document those changes.

7.7.1. Methods for Data Verification

Data verification immediately follows data entry and involves checking the accuracy of the computerized records against the original source (usually hard copy field records), and identifying and correcting any errors. SWAN procedures are to verify *all* records entered against their original source. When the computerized data are verified as accurately reflecting the original field data, the paper forms are set aside for use later in the data validation process.

The following four verification methods will be used by SWAN for all project data:

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

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

7.7.2. Methods for Data Validation

Although data may have been correctly transcribed from original field notes or forms, they still might be inaccurate or illogical. For example, entries of stream pH of 25.0 or a temperature of 95°C in data files raise doubt about their accuracy, and such entries almost certainly are incorrect, whether or not they were properly transcribed from field forms. This process of reviewing computerized data for range and logic errors is called *validation*, and it can accompany data verification *only* if the reviewer has comprehensive knowledge about the data. Validation is a separate operation carried out *after* verification by a project specialist who can identify generic and specific errors in particular data types.

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

Corrections or deletions of logical or range errors in a data set require notations in the original paper field records about how and why the data were changed. Modifications to the field records should be clear and concise while preserving the original data entries or notes. Validation efforts should also include a check for the completeness of a data set because field sheets or other sources of data could easily be overlooked.

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

- *Data entry application programming.* Certain components of data validation are built into data entry forms. The simplest validation during data entry is range checking, such as ensuring that a user attempting to enter a pH of 20.0 gets a warning and the opportunity to enter a correct value between 1.0 and 14.0 (or within a narrow range appropriate to the study area). Not all fields, however, have appropriate ranges that are

known in advance, so knowledge of what are reasonable data and a separate, interactive validation stage is important.

Edwards (2000) suggests the use of 'illegal data' filters, which check a specified list of variable value constraints on the master data set (or on an update to be added to the master) and create an output data set. This output data set includes an entry for each violation, along with identifying information and an explanation of the violation.

A caveat should be interjected regarding the operative word 'illegal.' Even though a value above or below a given threshold has never before been observed and the possibility that it could occur seems impossible, such an observation is not always an illegal data point. Edwards (2000) points out that one of the more famous data QA/QC blunders to date occurred when NASA's computer programs deleted satellite observations of ozone concentrations that were below a specified level, seriously delaying the discovery of the ozone hole over the South Pole.

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

Data quality assurance procedures should not try to eliminate outliers. Extreme values naturally occur in many ecological phenomena; eliminating these values simply because they are extreme is equivalent to pretending the phenomenon is 'well-behaved' when it is not. Eliminating data contamination is perhaps a better way to explain this quality assurance goal. If contamination is not detected during data collection, it will be detected later only if an outlying data value results. When we detect an outlier, we should try to determine if some contamination is responsible.

GIS, database, graphic, and statistical tools for ad-hoc queries and displays of the data can be used to detect outliers. Some of these outlying values may appear unusual but prove to be quite valid after confirmation. Noting correct but unusual values in the documentation of the data set saves other users from checking the same unusual values.

- *Other exploratory data analyses.* Palmer and Landis (2002) suggest that in some cases, calculations for assessments of precision, bias, representativeness, completeness, and comparability may be applicable and that for certain types of measurements, evaluation of a detection limit may also be warranted. Normal probability plots, Grubb's test, and simple

and multiple linear regression techniques may also be used (Edwards 2000).

7.8. Version Control

Version control is the process of managing copies of changing files over the course of a project. Change includes any alteration to the structure or content of the files, which should not be made without the ability to fully recover a data set as it existed before changes were made. Before making any major changes to a file, a copy of the file should be saved with a unique identification number. This process is particularly important when appending newly-validated data to a master data set, or when upgrading a database to a new version. In addition, proper controls and communication are required to ensure that only the most current version is used for data entry or analysis.

Specific naming conventions and directory structures related to version control are detailed in individual monitoring SOPs. The data manager will determine the version control method that will be used, and other network personnel are responsible for accurately designating versions for any files they have worked with.

General guidelines for version control file naming are as follows:

<project_file>_<QA/QC_status>_<date>.<ext>

Where:

<project_file> = the name of the established base file name

<QA/QC_status> = the status of the information. For example, “_RAW_”; “_VERIF_”; “_VALID_”

<date> = The date of the file, as YYMMDD.

<ext> = the file extension, such as .mdb, .xls

7.9. Data Quality Review and Communication

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

7.9.1. Value of Feedback from QA/QC Procedures

QA procedures may need revision to improve quality levels if verification and validation processes reveal an unacceptable level of data quality. Quality checks should not be performed with the sole objective of eliminating errors, as the results may also prove useful in improving the overall process. For example, if the month and day are repeatedly reversed in a date field, the data entry technicians may require retraining about the month/day entry order. If retraining is

unsuccessful in reducing the error's occurrence, the computer program may need to be rewritten so that month and day are entered separately, field length limits are enforced, or a pick list is created. In this manner, the validation process will serve as a means of improving quality.

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

7.9.2. Monitoring Conformance to Plans and Standards

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

Audits include verification of the following:

- Data collection and reporting requirements are being met
- Data collection and reporting procedures are being followed
- Verification and validation procedures are being followed
- Data file structures and maintenance are clear, accurate and according to plan
- Revision control of program documents and field sheets are adequate
- Calibration and maintenance procedures are being followed
- Seasonal and temporary staff have been trained in data management practice
- Metadata collection and construction for the program proceeds in a timely manner
- Data are being archived and catalogued appropriately for long-term storage

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

7.9.3. Communicating Data Quality

The Network will use data documentation and metadata to notify end users, project managers, and Network management of data quality. A descriptive

document for each data set/database will provide information on the specific QA/QC procedures applied and the results of the review. Descriptive documents or formal FGDC-compliant metadata will document quality for spatial and non-spatial data files posted on the Internet.

7.10. Chapter Credits

Debbie Angell (Sonoran Desert Network) and Margaret Beer (Northern Colorado Plateau Network).

7.11. References

- Bennett, A.B., W.L. Thompson, and D.C. Mortenson. 2006. Vitals signs monitoring plan, Southwest Alaska Network. National Park Service. Anchorage, AK.
- 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. Blackwell Science Ltd. Methods in Ecology Series, Malden, MA.
- Michener, W. K. 2000. Metadata. 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.
- Palmer, C. J. 2003. Approaches to quality assurance and information 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, DC.
- Tessler, S., and J. Gregson. 1997. Draft Data Management Protocol. NPS I&M Program. (<http://www1.nrintra.nps.gov/im/dmproto/joe40001.htm>). Accessed 28 November 2004.

8. Data Documentation

Documenting data is the most important step toward ensuring that data sets are usable well into the future. Data longevity is roughly proportional to the comprehensiveness of their documentation (Michener 2000).

8.1. Federal and NPS Standards

Direction for documentation at the federal and NPS servicewide level are as follows:

- Executive Order 12906, mandates federal agencies to “...document all new geospatial data it collects or produces, either directly or indirectly...” using the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM).
- FGDC CSDGM extensions, such as the Biological Data Profile, Remote Sensing Extension, and Shoreline Data Profile are not required, but recommended where appropriate.
- NPS Geographic Information System (GIS) Committee requires all GIS data layers be described with FGDC standards and the NPS Metadata Profile.

8.2. Network Standards

SWAN standards for documentation are as follows:

- Project document will be completed using the following tools:
 - o Project Organizer
 - o Project Tracking Database
- Spatial data documentation will be completed using the full standards set by the FGDC CSDGM.
- Tabular database documentation will be completed using
 - o Minimum standards set by the FGDC CSDGM
 - o “Database Documentation Template” for more detailed documentation.
- Programming code will contain specific header information as well as annotation throughout the program as to purpose and function of the specific code.
- Vital signs protocol documentation will be completed using
 - o “Guidelines for long-term monitoring protocols” (Oakley et al. 2003)
 - o Master Versioning Table, and additional vital signs documentation described in this chapter.

8.3. Project Documentation

8.3.1. Project Organizer

Projects should be well organized in a project directory structure or an existing repository. A Project organizer file, called index.doc, should be created to document primary contacts, project objectives, organization and purpose of the files, naming standards, and any other related documentation. The Project organizer helps to organize and document the project files, and enables others to understand and access the project information, as needed. The Project organizer is simplified method that stays within the project directory, providing direct access to documentation (see Appendix 8-1).

8.3.2. Project Tracking Database

Projects should be logged into the SWAN project tracking database to provide basic documentation on each project. The project tracking database is designed to be flexible for I&M Program operations. This database tracks the following information:

- Contact information
- Dates
- Links to Program goals, associated parks, vital signs, permits, investigators annual reports
- Abstracts
- Status, repository, and notes on deliverables.

Reports can be produced from this database to:

- Document the project
- List the status of deliverables

Transfer necessary project information to park curators for archival purposes.

Appendix 7-1, page 4- illustrates an example of a report.

8.4. Spatial Data Documentation

Spatial data will use the FGDC CSDGM standard. Metadata records created by the networks are integrated into NPS, NR-GIS Metadata database. Associated data sets will be stored in the NR-GIS Data Store. The locations of the associated data sets are directly linked in the metadata record. Metadata records will be coordinated through the AKRO GIS Team to ensure quality and consistency.

8.5. Tabular Database Documentation

Tabular databases will use the minimum FGDC CSDGM standard for documentation. This allows data discovery through the use of FGDC clearinghouses. It is difficult to document a relational database in adequate detail using this standard alone. Therefore, in addition to the metadata, a database document, following the template for database documentation (Appendix 8-3), should be written and should include:

- Revisions to the database
- Overview of the use and purpose of the data
- Illustration of the entity diagrams
- Data dictionary of tables, attributes, and relationships
- Explanation of queries, forms and reports

8.6. Programming Documentation

Programming, such as Visual Basic (VB) or SAS, will use a header to document the program. In addition, annotation will be made throughout the program to document the purpose and functions. The header documentation should include:

- File name of the program
- Original author and title
- Description
- Usage
- Parameters
- Requirements
- Datasets used
- Revision history including revision date, revision author and title, reason for revision.

8.7. Vital Signs Protocol Documentation

8.7.1. Master Version Table

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

The Northern Colorado Plateau Network (NCPN) has suggested using a database to track the narrative and SOP version numbers in a Master Version Table (MVT). The MVT contains a Version Key Number that designates the narrative and SOP versions that are in use at a specific time (Table 8-1). Every protocol contains an SOP entitled “Revising the Protocol Narrative and SOPs.” The MVT is contained in this SOP and is required to be updated when any protocol revisions are made. The protocol narrative, SOPs, and data will not be distributed independently of the MVT. The MVT is in development by NCPN and is further described in the NCPN Data Management Plan.

Table 8-1. Example of the Master Version Table used to track changes in vital signs protocol narrative and SOPs in the project protocol.

Version Key #	Version Key Date	Narrative	SOP #1	SOP #2	SOP #3	SOP #4	SOP #5	SOP #6	SOP #7
VK1	12/15/2004	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
VK2	5/10/2005	1.00	1.01	1.01	1.00	1.00	1.00	1.00	1.01
VK3	11/18/2005	1.00	1.01	1.01	1.10	1.01	1.00	1.00	1.01

8.7.2. Additional Vital Signs Metadata

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

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

8.8. Chapter Credits

Margaret Beer (Northern Colorado Plateau Network) and Bill Eichenlaub (Glacier Bay National Park).

8.9. References

Executive Order 12906, "Federal Register," Volume 59, Number 71, (1994) pp. 17671-17674.

- <http://www.fgdc.gov/publications/documents/geninfo/execord.html>
- Pertaining to the legislation for the FGDC CSDGM standard

Federal Geographic Data Committee. 2000. Content standard for digital geospatial metadata workbook. Version 2.0. 122 pp.

- http://www.fgdc.gov/publications/documents/metadata/workbook_0501_bmk.pdf
- Excellent reference workbook for completing metadata.

Oakley, Karen L., Lisa P. Thomas, and Steven G. Fancy. "Guidelines for long-term monitoring protocols." *Wildlife Society Bulletin* 31, no. 4 (2003): 1000-1003.

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. Chapter 7 of the SWAN Vital Signs Monitoring Plan (Bennett et al. 2006) contains the background and overall approach to data analysis and reporting by the Network. The associated data management objective is to provide valid data in formats that support scheduled and ad hoc display, query, analysis, summary, and reporting of data. Routine and scheduled data summary and analysis requirements and procedures are identified in each vital sign monitoring protocol. The following sections discuss data management activities related to using GIS and database application software for data summary and analysis, and to prepare data for analysis using statistical software applications.

9.1. Periodic and Annual Reporting

The data manager will work with the biometrician and others involved in data analysis to specify and design or adapt database objects, fields, and values to support the formats and functions necessary for analysis using statistical software applications, e.g., SAS® (commercial; <http://www.sas.com>) and R (freeware; <http://www.r-project.org>). A list of monitoring reports and venues of common vital signs, and their purpose, frequency, and intended audience is shown in Table 7.3 in the SWAN Vital Signs Monitoring Plan (Bennett et al. 2006). Some basic summary and reporting functions required by a vital sign monitoring protocol can be developed within Microsoft Access database applications where data are stored. Examples include descriptive statistics (mean, standard deviation, sample size). The biometrician and others will use existing and custom data conversion and export functions in Microsoft Access to prepare data sets for import into other software applications. Spatial analysis and maps will be produced by Network and/or park affiliate staff. Ad hoc queries and reports will be handled on a case by case basis due to their dynamic nature. SWAN will enhance its web site over time to deliver reports and provide supplemental background data and information.

9.2. Long-term Trends and Analyses

Most long-term data analysis will involve statistical software programs to perform the two descriptive and trend analyses described in Chapter 7 of the SWAN Vital Signs Monitoring Plan (Bennett et al. 2006). Data formats required by statistical software often involve arrays of binary or discrete values that represent one or more parameters. Data analysts and the network data manager

will identify and develop the data conversion routines necessary to generate these formats for analysis.

GIS functions can also contribute to understanding long-term status and trends of vital signs and ecosystems. Methods may be developed to analyze time series data, perform geostatistical functions, and do spatial network analysis with hydrography features. Tabular and spatial results can be shared in reports and made available on the Network's web site and via internet map services.

9.3. Chapter Credits

Bill Thompson (Southwest Alaska Network) and Rob Daley (Greater Yellowstone Network).

9.4. References

Bennett, A.B., W.L. Thompson, and D.C. Mortenson. 2006. Vitals signs monitoring plan, Southwest Alaska Network. National Park Service. Anchorage, AK.

10. Data Dissemination

SWAN data management aims to ensure that:

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

Links to SWAN public data products may be accessed via the SWAN public website, under “Information Discovery” (<http://www.nature.nps.gov/im/units/swan/>). This will be the Network’s portal for the most current data distribution as the monitoring program develops. Distribution instructions for each data set will be provided in the respective metadata.

10.1. Mechanisms for Distribution

SWAN data products (final deliverables or periodic milestones) will be “packaged,” where possible, and made available for distribution as a complete set. This package is similar to the materials consolidated for natural history archiving and curation (see Chapter 12).

Access to SWAN data products will be facilitated via a variety of means that allow users to browse, search and acquire Network data and supporting documents. These means include, but are not limited to (see also Table 10-1):

- Links to SWAN public data products may be accessed by the SWAN public website, under “Information Discovery” (<http://www.nature.nps.gov/im/units/swan/>).
- NR-GIS Metadata and Data Store. Distribution instructions for each dataset will be provided in the respective metadata.
- Service-wide databases, such as NPSTORET, NPSpecies, and NatureBIB.
- Regional, Network, or Park data servers protected with read-only access.
- 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.

SWAN plans to use the internet for dynamic and on-going data that may not yield specific tangible products.

Table 10-1. Primary repositories for SWAN information and associated specimens.

Item		Repository
Reports (public)	digital	NatureBIB, Data Store, SWAN data server
	hard copy	Alaska Resources Library and Information Services (ARLIS), park and network libraries
bibliography		NatureBib
Network-generated digital datasets and data products (public, non-sensitive) • Certified data and data products (including photographs) • Metadata		NR-GIS metadata and data store, SWAN data server, NPSpecies, EPA STORET.
Network-generated digital project data and info (NPS staff, sensitive) • Raw, validated and analyzed data • Metadata • Submitted reports • Digital photographs • Digital presentations		SWAN data server. Selected vital sign data may be housed externally with a Memorandum of Understanding (MOU).
Project product materials • Specimen Vouchers • Photograph film		University of Alaska Museum, park archives, or other curation facility (according to project protocol)
Project administrative records or misc. items. (hard copy)		SWAN office

Information will be made available to two primary audiences: public and NPS employees, as determined by data sensitivity and development status. Only fully documented, certified, non-sensitive data and data products may be uploaded to public distribution repositories or otherwise released to the public.

10.2. Ownership, FOIA, and Sensitive Data

SWAN products are considered property of the NPS (OMB, Circular A-110, Section 36).

The Freedom of Information Act, 5 U.S.C. § 552 stipulates that the United States Government, including NPS, must provide access to data and information of interest to the public, regardless of whether or not the federal government created the records. FOIA is intended to establish a right for any person to access federal agency records that are not protected from disclosure by exemption or by special law enforcement record exclusions. Under the terms of FOIA, agencies must make non-protected records available for inspection and copying in public reading rooms and/or the Internet.

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

- Executive Order No. 13007: Indian Sacred Sites

- National Parks Omnibus Management Act (NPOMA; 16 U.S.C. 5937)
- National Historic Preservation Act (16 U.S.C. 470w-3)
- Federal Cave Resources Protection Act (16 U.S.C. 4304)
- Archaeological Resources Protection Act (16 U.S.C. 470hh)

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

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

The Network will comply with all FOIA restrictions regarding the release of data and information, as instructed in NPS Director’s Order #66 and accompanying Reference Manuals 66A and 66B (currently in development). Managing natural resource information that is sensitive or protected requires the following steps:

- Identification of potentially sensitive resources
- Compilation of all records relating to those resources
- Determination of which data must not be released in a public forum
- Management and archival of those records to avoid their unintentional release

Classification of sensitive data will be the responsibility of Network staff, park superintendents, and project 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.

Following are suggested guidance for determining whether information should be protected:

- Has harm, theft, or destruction occurred to a similar resource on federal, state, or private lands?
- Has harm, theft, or destruction occurred to other types of resources of similar commercial value, cultural importance, rarity, or threatened or endangered status on federal, state, or private lands?
- Is information about locations of the park resource in the park specific enough so that the park resource is likely to be found at these locations at predictable times now or in the future?
- Would information about the nature of the park resource that is otherwise not of concern permit determining locations of the resource if the

information were available in conjunction with other specific types or classes of information?

- Even if relatively outdated, is there information that would reveal locations or characteristics of the park resource such that the information could be used to find the park resource as it exists now or is likely to exist in the future?
- Does NPS have the capacity to protect the park resource if the public knows its specific location?

10.2.1. Access Restrictions on Sensitive Data

Network staff is responsible for managing access to sensitive data handled by the Program. All potentially sensitive park resources will be identified and investigators working on network projects will be informed that:

- All data and associated information must be made available for review by network staff prior to release in any format
- Any information classified as protected should not be released in any format except as approved in advance by NPS

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

When preparing information for any repository, Network staff ensures that all protected information is properly identified and marked. All references to protected information are removed or obscured in any reports, publications, maps, or other public forum.

Network staff will remove any sensitive information from public versions of documents or other media. They will isolate sensitive from non-sensitive data and determine the appropriate measures for withholding sensitive data. The main distribution applications and repositories developed by the I&M Program, are maintained on both secure and public servers and all records marked 'sensitive' during uploading will only become available on the secure servers. Procedures for assigning a sensitivity level to specific records when uploading to both the NPSpecies and NatureBib databases are given on the following websites:

- <http://science.nature.nps.gov/im/apps/npspp/index.htm>
- <http://www.nature.nps.gov/nrbib/index.htm>

Thus, access to data on sensitive park resources can be limited to network staff or research partners. However, limits to how these data are subsequently released must also be clearly defined. It is crucial that the person uploading records to the

online applications (repositories) is familiar with the procedures for identifying and entering protected information.

10.2.2. NPS Only vs. Public

Only data subjected to complete quality control are released, unless necessary in response to a FOIA request. Products of incomplete, poor, or questionable quality (typically legacy data) may not be appropriate for the public and should be managed separately from acceptable quality. These lower quality products may be the only source of information on the natural history of the park, may have been the basis for early management decisions and are still valuable in-house. Provided these do not contain sensitive data as described in this chapter, these data may be released to the public upon specific FOIA request. They must be accompanied with qualifying documentation.

I&M Program applications, such as NatureBIB and NR-GIS metadata, provide a means to flag these data as “NPS Only” or “Public”. Additional notes describing the quality should be added to the records in these applications, as resources permit.

10.3. Feedback Mechanisms

Overall comments and questions concerning SWAN I&M Program are welcome at any time and may be submitted via e-mail or telephone to the Network Coordinator. The SWAN website will provide an opportunity for NPS staff, cooperators and the public to provide feedback on data and information gathered as part of the Network’s I&M Program. The metadata attached to each dataset available to the public will include information on where to send comments or questions regarding the specific data.

10.4. Chapter Credits

Sara Stevens (Northeast Coast and Barrier Network) and Wendy Schumacher (Washington Support Office).

11. Data Maintenance, Storage and Archiving

This chapter describes procedures for the long-term management and maintenance of digital files and data that result from SWAN projects and activities. The overall goals of these procedures are:

- to avert the loss of information over time
- to ensure that network information can be easily obtained, shared, and properly interpreted by a broad range of users.

Effective long-term data maintenance is inseparable from proper data documentation, and an essential part of any archive is accompanying explanatory materials (Olson and McCord 1998). Data documentation is more fully explained in Chapter 8.

11.1. NPS and Network Standards

- SWAN will standardize on Microsoft products as required by all Department of Interior agencies (Department of Interior, Assistant Secretary for Policy, Management, and Budget, Findings and Determination, September 13, 2002).
- SWAN will update and maintain data sets to no more than two versions behind current version of software, or store the data set in American Standard Code for Information Interchange (ASCII) format, complete with data and file documentation.

11.2. Electronic Archiving and Storage

Finite (short-term) projects

Electronic version of the deliverables, as listed in the project tracking database, cooperative agreements, permits, study plans or other administrative records will be electronically archived by three archiving methods:

- 1) Online project archive, where all records are packaged as a stand-alone database. Types of information to be included are listed in Chapter 1, Table 1-1.
- 2) Offline project archive, same information as above, which may include draft files. Stored on CD or DVD.
- 3) Product Library, where like products are stored with other products, such as the final report stored with other reports.

On-going (long-term) projects

Milestone products should be electronically archived similarly to the finite (short-term) projects described above. On-going data sets will be inherently archived and stored on the SWAN servers and backup procedures.

11.3. Electronic Archiving Process and Workflow

Electronic files should be consolidated and packaged for electronic archival when a project is complete or reaches periodic milestones for ongoing projects. SWAN

project leaders are responsible for packaging electronic files and data for the data manager. Figure 11-1 illustrates the general workflow of electronic storage (in green), which is then followed by the hard copy storage (discussed in the next chapter).

The project leader should prepare the electronic files as follows:

- Complete the project organizer documentation
- Comply with file naming standards
- Comply with documentation standards
- Include all related documents listed in Chapter 1, section 5
- Make a clear distinction between draft and final files. Draft files will be included in project backup, but will be deleted for final online project archival
- Make a clear distinction between public and sensitive information
- Ensure the project tracking database is complete (this will require working with the data manager)

The data manager will:

- Archive the project information on the centralized \Archived_Projects directory. Stored indefinitely.
- Complete a project backup on CD or DVD; Media may be stored for 5 years.
- Integrate deliverables, such as final reports, into appropriate catalogs, electronic libraries and clearinghouses (Appendix 4-1), as described in this Plan or in the monitoring protocols. Stored indefinitely.

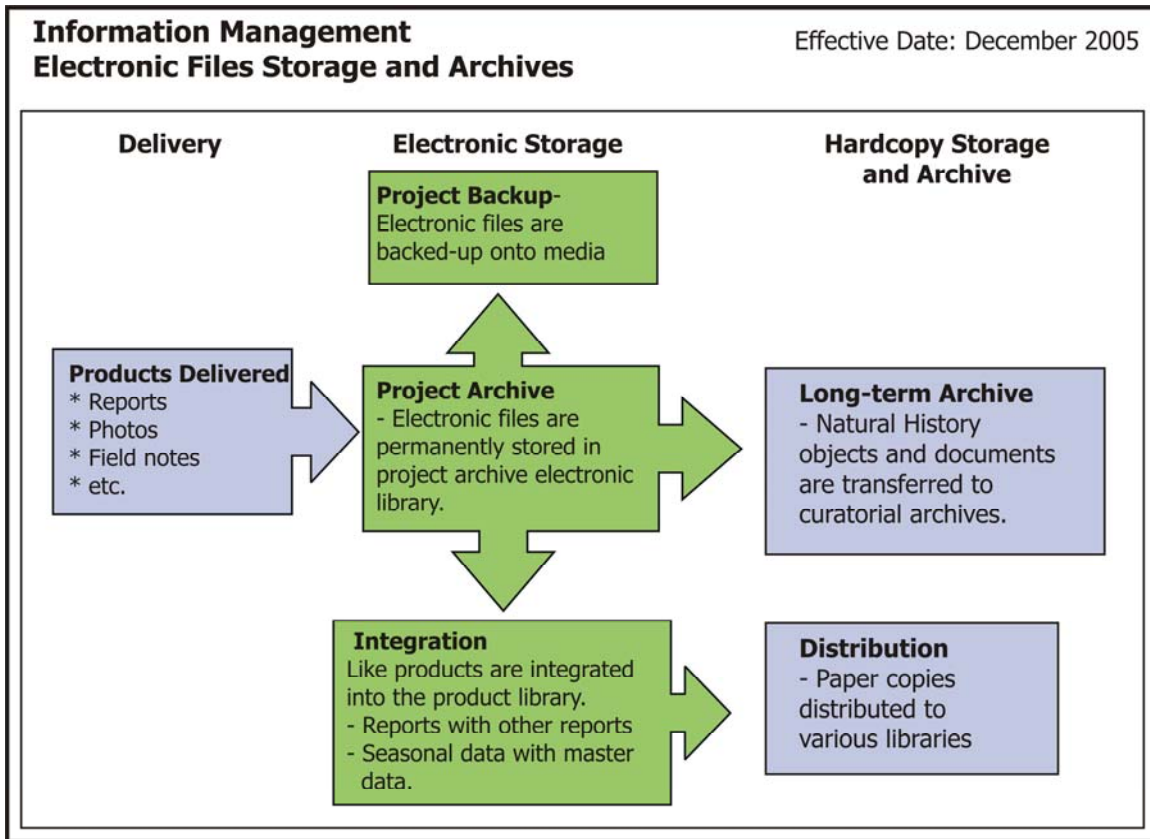


Figure 11-1. Electronic files storage and archives diagram.

11.4. Quality Control for Converted Data

Databases that are converted from one version of database software to an upgraded version will require additional QC, particularly when the database applications are actively used for data entry or analysis. Forms, queries, reports, and data entry will be thoroughly tested during upgrades.

For data sets of limited use, the Network may decide to convert to ASCII format. In this case, the data sets will no longer be upgraded. Complete documentation is necessary and should include table, field and object relationship descriptions. All ASCII files created from databases will undergo quality control activities or functions to ensure that the number of records and fields correspond to the source data set, and that conversion has not created errors or data loss.

11.5. Digital Data Backups

11.5.1 Project Backups

Project backups are done at known milestones of a project. This may be after data entry, just before a project directory clean-up, or at project completion. Interim project backups are encouraged and are the responsibility of the project leader. The data manager will complete a project backup at the time of project completion.

11.5.2. Data Backup for Catastrophic Event Protection

The risk of data loss exists from a variety of sources including catastrophic events (e.g., fire, flood), user error, hardware failure, software failure or corruption, and security breaches or vandalism. Performing regular backups of data and arranging for off-site storage of backup sets are the most important safeguards against data loss.

SWAN is integrated into the AKRO local area network. SWAN data files are stored on disk arrays distributed among multiple AKRO network data servers.

The AKRO network employs a Disk-to-Disk-to-Tape (D2D2T) onsite backup system described as follows: A complete backup copy of pertinent network data is kept on the 7 TB disk array of a separate backup server. (NOTE: "Pertinent data" does not include "*.tmp", "*.exe", or "*.sys" files nor the contents of "recycler", "msdownld.tmp", "TEMP", "I386", "win98", "winnt", "windows" or "cab" folders on any machine.)

Disk-to-Disk portion of the backup system

- Every Tuesday through Friday, new files added to any of the computer network servers on that day are added to the backup copy and network files changed on that day are updated on the backup disk copy.
- On Monday, new files added to any of the network servers on that day are added to the backup copy and network files changed on that day are updated on the backup copy and any file or folder deletions that occurred on the network since the previous Tuesday are deleted from the backup disk copy.

Disk-to-Tape portion of the backup system (runs after the Disk-to-Disk portion

- On Friday evening, a full copy of the backup disk is backed up to a Tape library.
- Every Monday through Thursday evening, new and changed files on the backup disk array are backed up to the tape library.

Offsite data storage

When the tapes in the library are filled, the library is reloaded with fresh tapes and the removed tapes are stored offsite. Two complete sets of the tapes will be

kept offsite. AKRO currently rotates through an inventory of 270 tapes. Statistics kept on each tape are used to determine the appropriate retirement of that tape.

Backup of desktop computers

Backup of data that resides on the desktop computers of networked staff, although ultimately the responsibility of each staff member, is also provided for in the computer network backup schema. Each AKRO workstation is configured with up to 4 shares that are backed up to the network if the machine is online during the “workstation backup” jobs that are run several times each week. If the workstation is detected on the network during the job, the changes since the last backup are duplicated to the copy on the backup server and the workstation data is wrapped into the D2D2T system described above.

General

Random restores are performed periodically to test the system. The verify option is not utilized during tape backups because the doubled backup times required to do so would exceed the available backup windows.

11.6. Chapter Credits

David Oradei (Alaska Regional Office) and Margaret Beer (Northern Colorado Plateau Network)

11.7. References

Olson, R.J. and R.A. McCord. 1998. Data Archival. In: Michener, W.K., J.H. Porter, and S.G. Stafford. Data and information management in the ecological sciences: a resource guide. LTER Network Office, University of New Mexico, Albuquerque, NM. Pp. 53-57.

12. Natural History Archiving, Storage, and Records Management

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

This chapter is not intended to provide a full description of archiving procedures. That information is covered in museum manuals and regulations. It is intended to provide guidance in making the transition from completed products delivered by the project leader to a secure and long-term storage facility managed by park and/or regional repositories.

12.1. NPS Standards

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. 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 NPS-19 Appendix B, which provides guidelines on natural resource-related records (including, specifically, the results of I&M Programs), indicates that all natural resource records are considered “permanent” and need to be retained either in an appropriate park museum facility or at the National Archives. 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.

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

12.2. Park Unit Standards

Direction for managing materials for each park unit is described in each collection facilities “Scope of Collection Statement”. In general, only materials directly “... related to one or more of the park’s themes or site-related materials that the NPS is legally mandated to preserve” will be accepted. All of these materials are required to be cataloged in the NPS cataloging system (ANCS+) (36 CFR, Section 2.5)

All specimens collected as part of the Network projects are property of the respective park. No samples should be tested or destroyed without prior consent of the park curator.

12.3. Network Standards

SWAN will adhere to the standards described above. In addition, SWAN’s strategy is to keep like specimens together, preferably in Alaska, where research can be done across the landscape, within a discipline, or for the benefit of the state. SWAN will limit the number of non-NPS repositories for each type of materials to avoid the scattering of voucher specimens (For example, all small mammals will be stored at the University of Alaska Museum at the University of Alaska - Fairbanks).

12.4. Role of curators in curation and museum collection storage

Curators and archival specialists for SWAN parks and the region are an excellent source of expertise, advice, and guidance on curatorial issues, and they have a role in almost every project undertaken by the Network. Project leaders should involve park curators during the project planning stage to ensure that all aspects of museum curation of documents, specimens, and other objects are considered, and that any associated expenses are included in project budgets.

An accession number is assigned by the park curator at the time of a park research permit, when there will be field collections, or at the completion/milestone of a project, for “desktop” research. After collection, a range of catalog numbers can be assigned by the curator for the principal investigator to use. The accession and catalog number assignments must be requested by the project leader.

The project leaders will package their project’s materials and provide this to the Alaska Regional Curatorial Center (ARCC). The ARCC will ensure the materials are properly stored, cataloged and archived. The project leader is ultimately responsible for ensuring this is completed.

12.5. Workflow

SWAN project leaders are responsible for preparing materials for the curator. The workflow illustrated here and described in this chapter refers to the general procedure for making the transition between the project leaders and the museum staff. The museum staff will follow a much more detailed procedure for actual curation.

Figure 12-1 illustrates the general workflow of archiving natural history materials. The workflow begins with the delivery of the products, walks through the electronic storage of these products, and then illustrates (in green) the storage and archiving of materials.

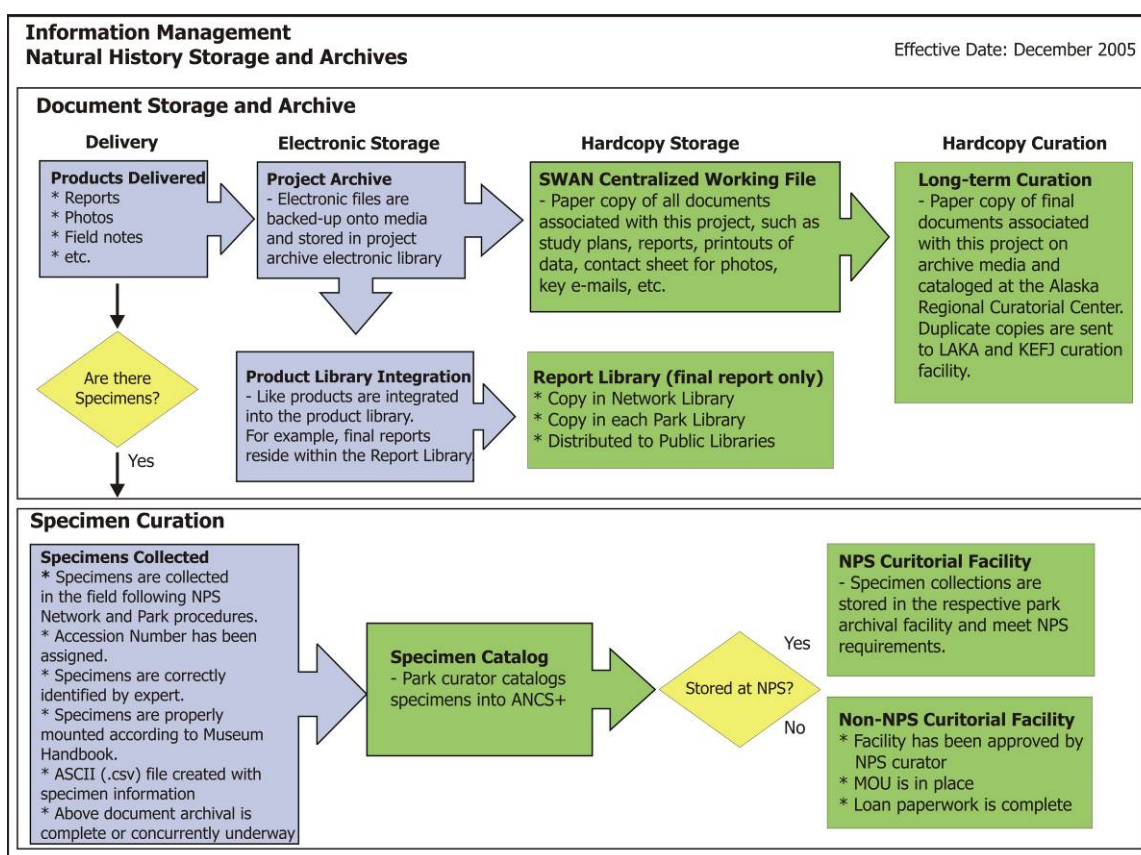


Figure 12-1. Workflow of natural history curation

12.6. Network Archiving Process

Project materials should be consolidated and packaged for curation when a project is complete or reaches periodic milestones for on-going projects. The project leader should acquire an accession number from the respective curator.

The project leader should package the project information as follows:

- A coversheet, memo, or Collection Catalog Worksheet listing contact information, project abstract and purpose, sensitivity and use of materials, and all materials included in the package (see Appendix 12-1 for example).
- All materials are clearly labeled with:
 - Park acronym
 - Date, or range of dates
 - Accession number
 - Project number
- Field notes are on acid free paper (preferred) or copy paper and stored in a 3-ring binder or book box.
- Other paper materials such as reports and data printouts are also stored on acid free paper (preferred) or copy paper and placed individually in an acid free, labeled folder.
- Photos should be printed and documented.
- Voucher specimens are properly labeled according to the NPS Museum Handbook, Part II
- For specimens not residing in park collection facilities, loan paper work is complete and a copy of the form is stored with the project package.
- Specimens not residing in NPS repositories should be stored in a selected facility that meets NPS museum collection standards (see Appendix 12-2 for currently approved facilities).
- Curation of specimens must be addressed in network vital signs protocols and should specify:
 - Name and address of facility
 - Memorandum of Agreement, if applicable
 - List of items to be curated or loaned
- CD or DVD of all electronic materials. Jewel case should be labeled with:
 - Park acronym
 - Date created
 - Range of dates for information
 - Accession Number
 - Project Number
 - List contents

The museum staff will catalog this information into ANCS+ and will properly store the materials according to the NPS Museum Handbook, Part II.

A similar package should be prepared as “working copies” for the Network and each affected park. These will be stored in the respective Network and park project file cabinets and will be available for day to day usage.

12.7. Archive and Storage Locations

12.7.1. Alaska Regional Curatorial Center (ARCC), Anchorage, AK

Many SWAN reports and documents encompass data from multiple parks, which makes it challenging to accession archival copies into a specific network park museum. In these instances, ARCC will accession and catalog these projects into their collection. The ANCS+ record will reference all parks included in a report or document to help potential users locate the materials. Duplicate copies of these reports will be distributed to the Lake Clark/Katmai Study Center (curation facility for ANIA, ALAG, KATM and LACL) and the Kenai Fjords National Park curation facility. Voucher specimens will remain the property of their respective parks curation facility (discussed later in this chapter).

ARCC provides a temperature and humidity-controlled facility, professional museum staff, and meets all museum standards set by NPS. This repository will be used for original documents and associated materials produced by the network (e.g., reports, photographs, field notes, permits).

12.7.2. Lake Clark/Katmai Study Center (LAKA), Anchorage, AK

LAKA is the museum repository for Alagnak Wild River, Aniakchak National Monument and Preserve, Katmai National Park and Preserve, and Lake Clark National Park and Preserve. LAKA provides temperature and humidity-controlled facilities, a professional museum staff, and meets all museum standards set by NPS. Duplicate copies of documents and associated materials produced by the Network (e.g., reports, photographs, field notes, permits) will be stored here, whereas the originals are cataloged and stored at ARCC.

Specimens and other non-document objects related to the LAKA parks, will be curated, cataloged, and managed by LAKA.

12.7.3. Kenai Fjords National Park (KEFJ) Curation Facility, Seward, AK

KEFJ provides its own archival repository for the park. Currently, KEFJ does not meet all museum standards set by NPS. Archival duties are assigned as ancillary duties to natural resource staff. KEFJ may elect to use the ARCC for items. It is anticipated KEFJ will be providing an NPS museum standard repository when a new park facility is built.

Duplicate copies of documents and associated materials produced by the Network (e.g., reports, photographs, field notes, permits) will be stored here, whereas the originals are cataloged and stored at ARCC.

Specimens and other non-document objects related to KEFJ, will be curated, cataloged, and managed by KEFJ.

12.7.4. Non-NPS Facilities

NPS does not have the facility to curate all specimens collected. Moreover, some specimens would be more beneficial to scientists if they resided with other like specimens, such as the mammal collection stored at the University of Fairbanks, DNA samples of a particular taxon, etc. These facilities must meet NPS museum requirements, an MOU must be in place, and proper loan paper work from the park unit must be completed. All specimens collected for the Network are property of the respective park. No samples should be tested or destroyed without prior consent by the park curator.

Table 6-1 lists potential primary repositories that may be used by the I&M Program. Not all of these repositories have been reviewed or an MOU in place. These will be done as needed. Prior consent of the park is required.

12.7.5. Network Working Files and Library

SWAN documents will be centralized, organized, and stored in a common file cabinet. Files may contain project files, administrative documents, key e-mails or correspondence, reports and data printouts. Files should be filtered for final and pertinent information before curating.

A separate file cabinet for the SWAN library will contain annual and final reports and may contain reports related to the inventories or vital signs. All materials in this library will be cataloged into NatureBIB and should be citable. This library should not contain draft materials.

12.7.6. Park Unit Working Files and Library

A similar filing organization as described above is recommended for each park unit. SWAN will not directly manage documents at the park level but will provide each park unit with a project working file folder and completed reports to populate their respective libraries.

12.7.7. Public Resource Libraries

The Alaska Resource Library and Information System (ARLIS) and other resource libraries in the state will receive a copy of final and public reports (Southwest Alaska Network 2004). Although the NPS museum collections and the Network or park libraries provide storage for NPS and cooperators, the libraries provide the hardcopy storage easily available to the public and outside researchers. Their role as an off-site facility and contributor to the public should not be overlooked.

12.8. NPS Automated National Catalog System (ANCS+)

The cataloging system for NPS service-wide is ANCS+ (using ReDiscovery software customized for NPS). Mostly, staff with cataloging duties are the only staff with this software. Park collections are entered into ANCS+ at the local level, then “rolled up” to the service-wide database. The cataloging system uses an accession number to identify a collection, and catalog numbers to identify the items within the collection. For example, the KEFJ 2004 Freshwater Fish Inventory has one accession number, but contains several catalog numbers for each fish specimen collected.

Park staff may access the catalog system by one of three methods:

- o Contact a park curator and review together
- o Request a copy of the program and data from the park curator
- o Access the information online using ReDiscovery Web Catalog (<http://www.nps.gov/curatorial>)

At the present time, none of the SWAN parks have submitted their collections to the Web Catalog. This service is provided with an added cost to the park facilities.

12.9. Photographs

Museum curators have been reluctant to fully embrace digital photography and some have expressed concern that, with the accelerating rate of technological change, documentary heritage is in danger of being lost in the information age (Cox 2000). SWAN will keep most photographs in electronic format for ease of use, but realize hardcopy photographs are equally important as other hardcopy products.

Under the SWAN project leader's direction, digital and hardcopy photos will be screened for quality, redundancy and adequate documentation. SWAN accepts and processes 35mm slides (preferably Kodachrome or Ektachrome) which have a proven long-term stability (Wilhelm and Brower 1993), and 4x6 color prints. Original photographs are a high priority for placing in archival storage conditions.

Slides are labeled using indelible pigment ink, or using laser-printed archival-quality slide labels. Slide labels will include: a unique ID, project name, accession number, photographer name, photo date, a brief identification of contents (e.g., species name, plot ID), and geographic location (latitude, longitude, and datum or a description). All slides are stored in polypropylene slide sleeves. In addition, all slides are scanned and saved digitally as Tagged Image File Format (TIFF) (preferred) or JPEG files, and these are used as the primary means of distributing or reproducing the images.

Photographs are stored in individual polypropylene sleeves and within archival boxes. Each photo is 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 must also be submitted.

Every image, regardless of format, has an entry into a photo database (ThumbsPlus or other approved database) where attributes such as electronic file name, keywords, project, photo description, photographer, date, and location are cataloged.

12.10. Specimens

The Network assists with planning for monitoring projects that includes budgeting for specimen preparation and records processing and cataloging. Specimens collected under the auspices of SWAN will be provided to the respective park in which they were collected for curation, or to a repository approved by a park (where the specimens are considered on loan). SWAN will provide park curators with necessary data for cataloging each specimen. This data will be in comma-delimited format (.csv) for automated uploading into ANCS+. Data provided to non-NPS curators will be in an appropriate format for each institution.

To gain the most benefit to Alaska as a whole, SWAN will seek statewide repositories for specimens of like material. For example, mammal specimens will be stored with the University of Alaska Museum. This will allow future researcher access to not only NPS specimens, but other specimens in their area of interest.

12.10.1. Duplicate Vouchers

If the Network feels adequate collections have been made of a species, the duplicate specimens are returned to the parks.

12.11. Other Materials

All materials related to a project should be provided in the collection. Not all materials, however, are described in this chapter, such as audio or video. These items should be discussed on a case by case basis with the museum staff.

12.12. Chapter Credits

Stephanie Stephens (Alaska Regional Curatorial Center), Hildy Reiser (Chihuahuan Desert Network), Rob Daley (Greater Yellowstone Network), and Margaret Beer (Northern Colorado Plateau).

12.13. References

Cox, R.J. 2000. Searching for authority: archivists and electronic records in the New World at the fin-de-Siecle. *First Monday* 5:1.

(http://firstmonday.org/issues/issue5_1/cox/index.html)

Accessed November 28, 2004.

[Museum Handbook, Part II](#), *Museum Records*. Washington, DC: National Park Service, 1984 (revised).

National Park Service. 2001. NPS Director's Order 19: Records Management.

<http://www.nps.gov/refdesk/DOrders/DOrder19.html>

Southwest Alaska Network. 2004. Specifications for Distribution of Reports for the Inventory and Monitoring Program. National Park Service. Anchorage, AK.

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 Publication Company, Grinnel, IA.

12.14. Additional References

Following are guidelines listed in the Appendices that pertain to natural history archiving process.

Study Plan Agreement Template: Curation of Natural History Specimens

- o Appendix 12-1
- o Provides a template that may be used as an attachment for a study plan. This is particularly helpful when writing contracts. It simply states that any vouchers collected are the property of NPS. During the study the vouchers may be kept with the researcher on loan and should be returned to the park after processing. The researcher has the obligation to properly label the vouchers.
- o Approved by LAKA

Study Plan Handout: Natural History Specimen Collection Curatorial Responsibilities for national Park Service Collectors

- o Appendix 12-2
- o Provides an overview of the responsibilities of PI's with regard to museum collections.
- o Approved by ARCC

NPS Museum Collection Handbook

- o <http://www.cr.nps.gov/museum/publications/MHII/mushbkII.html>

- o Provides overarching guidance for museum collections
- o Approved service-wide

Director's Order #19: Records Management and Disposition Schedule

- o <http://www.nps.gov/refdesk/DOrders/DOrder19.html>
- o <http://data2.itc.nps.gov/wapc/records/nps19app-b.pdf>
- o Provide overarching guidance on which records should be kept indefinitely and which may be destroyed.
- o Approved service-wide

SWAN Park Unit Scope of Collections

- o 36 CFR Part 79: Curation of Federally-Owned and Administered Archeological Collections. 1991.
- o Kenai Fjords National Park Scope of Collection Statement and Collecting Guidelines
- o Aniakchak National Monument and Preserve Scope of Collection Statement and Collecting Guidelines
- o Lake Clark National Park and Preserve Scope of Collection Statement and Collecting Guidelines
- o Alagnak Wild Scope of Collection Statement and Collecting Guidelines
- o Katmai National Park and Preserve Scope of Collection Statement and Collecting Guidelines
- o Approved service-wide, regional, and park

ANCS+ User Manual, Data Dictionary, and ReDiscovery software support

- o <http://www.cr.nps.gov/museum/publications/ancs.html>
- o <http://www.rediscov.com/default.aspx?include=npssupport.htm>
- o Provides the technical information of field descriptions used in ANCS+. ReDiscovery is the software used to implement ANCS+.
- o ANCS+ Catalog Data Fields – Appendix 12-7
- o Approved service-wide

MOU Template: Agreement between the National Park Service and [Repository] on Management of NPS Natural History Collections.

- o Appendix 12-3
- o Provides a template that may be used when establishing an agreement with a long-term repository, such as a University or another agency. MOU details policies, accountability, and agreements from both NPS and the repository.
- o Approved region-wide

Forms:

- o SWAN Transfer of Natural History Files to Curation Excluding Object Collections. Appendix 12-4
- o Catalog Worksheet – Archives Module COLLECTION

- o Catalog Worksheet – Archives Module ITEM (Record)
- o Form 10-127. Conditions for Outgoing Loans Appendix 12-5

Templates:

- o Appendix 12-6
- o A series of Excel spreadsheets are available on the SWAN server that helps to transfer voucher information to ANCS+. These templates and instructions originated from Hildy Reiser from Chihuahuan Desert Network and will be tested by SWAN.
- o In review

Link between I&M and Archives, Inventory Program, Alaska Regional Meeting, January 30, 2004 Minutes

- o Appendix 12-8
- o Minutes from early discussion of curating natural history materials. Much of the results have been captured in this chapter. This provides a good overview of the reasons behind the decisions.
- o Approved statewide, within I&M.

13. Implementation

“If you do what you’ve always done, you’ll get what you’ve always gotten.”
- Anthony Robbins

As a data provider, we may be satisfied with what we have been doing for years. It may meet our immediate needs. The I&M Program, however, focuses attention on our data consumers – managers who need to be able to make decisions based on the data we provide or researchers who will be using these data decades from now. The central goal of vital signs monitoring is to understand the long-term patterns and processes of park ecosystems at multiple spatial scales. Good data management practices are essential to achieving this goal. Data cost less to save than to collect again, and ecological data often cannot be collected again at any cost.

The DMP contains practices that may be new to staff and cooperators. 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. The DMPs for each of the 32 I&M networks are the first comprehensive documents of their kind and will establish a gateway at each network for managers, educators, and the greater scientific community to reach data and information about the park resources.

Good data management practices will take time. Some data collection procedures and data management practices are already in use and may require minimal revisions. Others may involve several iterations of procedures and databases before reaching their acceptable and functional data reporting formats.

13.1. Common Data Management Sins

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.* This can happen for any number of reasons beyond the staff’s control. Unfinished projects, however, are the most difficult to manage. It takes several times longer to document an incomplete project than a finished one. The best data management practice is to finish the project or reach a logical milestone.
- *Lack of header information.* Often, sheets of information do not have a date, name or subject. It requires reading the content of the material in

order to make a determination if the information is relevant. This is extremely time consuming. Often relevant information is discarded because it does not have adequate header information. A good data management practice is to use a memorandum format or a standard header for all files for a project, no matter how insignificant.

- *Lack of documentation.* Most projects and data are inadequately documented. Without documentation, the data are often considered unusable. This DMP describes documentation techniques. Documentation should be on-going throughout the duration of the project.
- *Lack of file naming standard and version control.* In many cases it was impossible to know which file was the final product.
- *No archive or final storage.* Tucking years of data away in an individual's desk or personal computer isn't accessible to anyone else. Often it requires in-house knowledge to know the information even exists. Often when there is a turnover in staff, this information is lost.

13.2. Years 1-5 of Implementation

The first few years of the implementation of this plan will involve learning, testing, and refining. Data management fundamentals training will be provided to network and park staff. The DMP will be implemented for the first few vital sign monitoring protocols. During this period, templates and SOPs will be developed and tested. These will be applied to the other vital sign monitoring protocols. We anticipate that there will be some bottlenecks and their identification and elimination is an important phase in the implementation of this plan.

Goals:

- All Network staff understand the fundamentals of data management, which include:
 - File management
 - Documentation
 - Quality assurance and quality control
 - Electronic storage
 - Archive storage
- Improve data management practices by implementing
 - Conceptual Data Models for selected vital signs
 - Testing of data entry prior to field work
- Develop common SOPs that can be used for multiple protocols
- Identify common analysis and reporting needs
- The first vital signs monitoring protocols will be implemented following the DMP

13.3. Years 5+ of Implementation

In five years, the DMP will be revised. This next revision may become more streamlined and direct. With the experience gained and practices commonplace, generalizations may be eliminated.

Goals:

- The DMP will be revised and streamlined for vital signs monitoring, and more specifically address
 - o data integration
 - o enterprise databases
 - o standard analysis and reporting across vital sign monitoring
- All vitals signs monitoring protocols will be implemented following the DMP
- Databases and custom reporting will continue to be enhanced.
- Framework and gateway for integration of data with other agencies or networks will be established.

Per peer review, the following topics should be reconsidered in the next version of this plan.

13.4. Chapter Credits

Alan Bennett (Southwest Alaska Network)