



Data and Information Management Plan ***Arctic Network Inventory & Monitoring Program***

Natural Resource Report NPS/ARCN/NRR—2008/053



ON THE COVER

(Bull moose in the morning sun, Noatak National Preserve)

Photograph by: Lois Dalle-Molle, NPS

Data and Information Management Plan

Arctic Network Inventory & Monitoring Program

Natural Resource Report NPS/ARCN/NRR—2008/053

National Park Service
4175 Geist Road
Fairbanks, AK 99709

August, 2008

U.S. Department of the Interior
National Park Service
Natural Resource Program Center
Fort Collins, Colorado

The Natural Resource Publication series addresses natural resource topics that are of interest and applicability to a broad readership in the National Park Service and to others in the management of natural resources, including the scientific community, the public, and the NPS conservation and environmental constituencies. Manuscripts are peer-reviewed to ensure that the information is scientifically credible, technically accurate, appropriately written for the intended audience, and is designed and published in a professional manner.

Natural Resource Reports are the designated medium for disseminating high priority, current natural resource management information with managerial application. The series targets a general, diverse audience, and may contain NPS policy considerations or address sensitive issues of management applicability. Examples of the diverse array of reports published in this series include vital signs monitoring plans; monitoring protocols; "how to" resource management papers; proceedings of resource management workshops or conferences; annual reports of resource programs or divisions of the Natural Resource Program Center; resource action plans; fact sheets; and regularly-published newsletters.

Views, statements, findings, conclusions, recommendations and data in this report are solely those of the author(s) and do not necessarily reflect views and policies of the U.S. Department of the Interior, NPS. Mention of trade names or commercial products does not constitute endorsement or recommendation for use by the National Park Service.

Printed copies of reports in these series may be produced in a limited quantity and they are only available as long as the supply lasts. This report is also available from the Natural Resource Publications Management website (<http://www.nature.nps.gov/publications/nrpm>) and the Arctic Network Inventory and Monitoring website (<http://www.nature.nps.gov/im/units/arcn>), or by sending a request to the address on the back cover.

Please cite this publication as:

National Park Service. 2008. Data and information management plan: Arctic Network Inventory and Monitoring Program. Natural Resource Report NPS/ARC/NRR—2008/053. National Park Service, Fort Collins, Colorado.

Change History

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

The following revisions have occurred to this plan since August 29, 2008

<i>Version #</i>	<i>Date</i>	<i>Revised by</i>	<i>Changes</i>	<i>Justification</i>

Contents

Figures and Tables	ix
Acknowledgements	x
Chapter 1. Introduction	1
About the Arctic Network.....	1
Purpose and Scope	2
Data Management and the Inventory and Monitoring mission	3
Goals of the Inventory and Monitoring Program.....	3
Data Management Goals.....	4
Data as a Valued Commodity	4
Data Management Expectations	4
Revisions to this Plan.....	5
Chapter 2. Data Management Infrastructure and System Architecture	7
Architecture.....	7
NPS Infrastructure	9
National Applications	10
Alaska Region Infrastructure	11
Park and Network-Level Infrastructure	11
Chapter 3. Data Management Process and Workflow	13
Project Process	13
Planning and Approval	14
Design and Testing	15
Implementation	15
Product Integration.....	15
Evaluation and Closure	15

Chapter 4. Data Management Roles and Responsibilities	17
Roles and Responsibilities	17
Chapter 5. Data Acquisition and Processing.....	21
Network Data	21
National Park Service Data.....	22
External Data	24
Museums and Herbaria	24
National and State Agencies	24
Non-Governmental and Other Organizations	24
GIS Data.....	24
Data Management Process.....	24
Acquisition of New Spatial Data	25
Maintaining Digital Files	25
Data Discovery.....	26
Chapter 6. Quality Assurance and Quality Control	27
National Park Service Mandate for Quality.....	27
Data Quality Goals and Objectives.....	28
Quality Assurance and Control Duties	28
QA/QC Process Cycle.....	30
Project Planning and Database Design	31
Database Design - Record-level Tracking	31
Lookup Tables	31
Project SOPs	32
Data Acquisition	32
Data Entry	33

Verification and Validation.....	33
Methods for Data Verification	34
Supplementary Methods	34
Validation.....	35
Methods for Data Validation	35
Data Quality Review and Communication	36
Conformance.....	36
Documenting and Communicating Quality	37
Data and File Management Operations.....	37
File Naming Standards.....	37
Version Control.....	38
Laboratory Data	38
Chapter 7. Data Documentation.....	39
Metadata Profile.....	39
Documentation Roles and Responsibilities	40
Metadata Tools.....	40
Metadata Process	40
Metadata Maintenance	42
Protocol Versions.....	42
Non-Program Data Documentation	42
Derived Data Documentation	42
Legacy Data Documentation.....	42
Chapter 8. Data Analysis and Reporting.....	43
Analysis.....	43
Reporting.....	43

Long-term Trends and Analyses	44
Funding and Timeliness	44
Chapter 9. Data Dissemination	45
Introduction.....	45
Data Distribution Mechanisms	45
Data Ownership	46
National Park Service Policy on Data Ownership	46
Establishing Data Ownership: Cooperative or Interagency Agreements	47
Data Classification: protected vs. public	47
Access Restrictions on Sensitive Data.....	49
Dissemination Mechanisms	50
Arctic Network Primary Server	51
Non-monitoring data.....	51
Alaska Resources Library and Information Services (ARLIS)	51
NPS Focus.....	52
Water Quality Data	52
Data Availability.....	52
Data Release Policy	52
Feedback Mechanisms	53
Data Error Feedback Response Procedures	53
Chapter 10. Data Maintenance, Storage and Archiving	55
Data Maintenance	55
Spatial Data.....	55
Files.....	56
File Backup Plan.....	56

Hardcopy Data and Information - Maintenance	56
Finalized Digital Datasets	56
Physical Material Archival	56
Project Related Items	57
Photographs.....	57
Chapter 11. Implementation.....	59
References.....	61

Figures and Tables

Figure 1. Arctic Network Parks	2
Figure 2. Relationship of the national-level data management plan to network-level data management documents adopted from chapter 1.2 of the Data management guidelines for inventory and monitoring networks (National Park Service 2008).	3
Figure 3. An enterprise client-server database will be the heart of the Arctic Network data management system.	8
Figure 4. Arctic Network generalized workflow showing major data repositories.....	9
Figure 5. Principle information technology components.....	10
Figure 6. Information technology connectivity between Network, Park and National operations.	12
Figure 7. Project life cycle.....	14
Figure 8. Some common data management elements affecting degree of need for QA/QC. Planning and training for data collection (QA) and entry is always a premium.....	28
Figure 9. General course of data and associated QA/QC procedures. Quality control with regards to data analysis is specific to each project and addressed in appropriate standard operating procedures.	30
Figure 10. The metadata document contains all information pertaining to project data.	39
Figure 11. Generalized path of data within the Arctic Network. The primary data server is accessed via desktop and web-based applications. Analyzed or derived data products are stored on the primary server, archived and catalogued (integrated) along with certified data.....	41
Figure 12. Water quality data flow.	52
Table 1. The principal Inventory and Monitoring applications and data repositories.	10
Table 2. Data stewardship responsibilities at each phase in the project life cycle.	17
Table 3. Network personnel roles and responsibilities	18
Table 4. Examples of park-based natural resource information that is processed or documented by the network.....	22
Table 5. Data quality roles and responsibilities.....	29
Table 6. Repositories for Arctic Network Program data and information.....	51

Acknowledgements

We the willing, led by the unknowing, are doing the impossible for the ungrateful. We have done so much, with so little, for so long, we are now qualified to do anything, with nothing.

--The Metro Para pledge

Anyone who does much snowshoeing in Alaska knows that the person out in front breaking trail does all the work for those who follow. I derived, adapted, and at times outright copied much of the material in this plan from a group of visionary trailbreakers to whom I am greatly indebted. I would like to thank the following individuals who said what needed to be said much better than I ever could: Doug Wilder (Central Alaska Network), Dorothy Mortenson (Southwest Alaska Network), Rob Daly (Greater Yellowstone Network), John Boetch (North Coast Cascades Network), Brent Frakes (Rocky Mountain Network) and Margaret Beer, (Inventory and Monitoring Washington Support Office). I am deeply indebted to all the data managers who started ahead of me in our phased implementation. Your efforts have cleared an easy path for those of us who follow.

Chapter 1. Introduction

Capt. Kirk: Spock, give me an update on the dark area ahead.

Spock: No analysis due to insufficient information.

Capt. Kirk: No speculation, no information, nothing? I've asked you three times for information on that thing and you've been unable to supply it. Insufficient information is not sufficient, Mr. Spock! You're the science officer.

You're supposed to have sufficient data all the time.

--"Star Trek" (1966)

No resource management enterprise can make objective, rational decisions to meet its objectives without good data. In the absence of sensed and quantified information decisions can only be based on assumptions, emotion and intuition. The National Park Service's Inventory and Monitoring program was developed to address the growing need among park managers for scientific information on which to make management decisions. The Inventory & Monitoring Program has committed to a building a program of effective ecosystem monitoring using state of the art data management practices to supply park managers with sound scientific information. This plan provides the vision and the roadmap for successfully accomplishing the Arctic Network's data management goals.

About the Arctic Network

The Arctic Network is one of thirty-two networks in the National Park Service's Inventory and Monitoring program and one of four networks in Alaska. The network consists of five public parklands, all of which are located in northwestern Alaska (Figure 1):

- Bering Land Bridge National Preserve
- Cape Krusenstern National Monument
- Noatak National Preserve
- Kobuk Valley National Park
- Gates of the Arctic National Park and Preserve

These parks are among the largest and most remote of all the national park lands and cover myriad arctic and subarctic habitats from the western sea coast to the eastern peaks of the Brooks Range. Administratively, the Arctic Network is integrated with and operates out of the Fairbanks Administrative Center in Fairbanks, Alaska and the Western Arctic Parklands offices in Nome and Kotzebue, Alaska. More information about the Arctic Network can be found on the web at <http://science.nature.nps.gov/im/units/arcn/index.cfm>.



Figure 1. Arctic Network Parks

Purpose and Scope

This document is not intended to address national I&M Program guidelines, Directors Orders, or legislative requirements. These items are addressed in Data management guidelines for inventory and monitoring networks, also referred to as the National Data Management Plan (National Park Service 2008), which provides broad guidance in terms of the National I&M Program, NPS Directives, and other legislative requirements.

This network-level data management plan is intended to act as a supplement to the National Data Management Plan, and function as a guide for establishing and maintaining a system that serves the specific data management needs of the Arctic Network. This plan establishes the general concepts and procedures the Arctic Network, its cooperators, partners, and potentially individual park units will use to ensure the quality, interpretability, security, longevity, and availability of program data and related information. Projects initiated by this program and the personnel involved with those projects must follow the guidance provided in this plan.

This plan will reference stand-alone standard operating procedures and task instructions that provide details on specific network operational procedures (i.e., back-up procedures, document and data documentation, document and data tracking, etc.). Figure 2 shows the relationship of the national-level data management plan to network-level data management documents. Standard operating procedures are available on the Arctic Network website and file server.

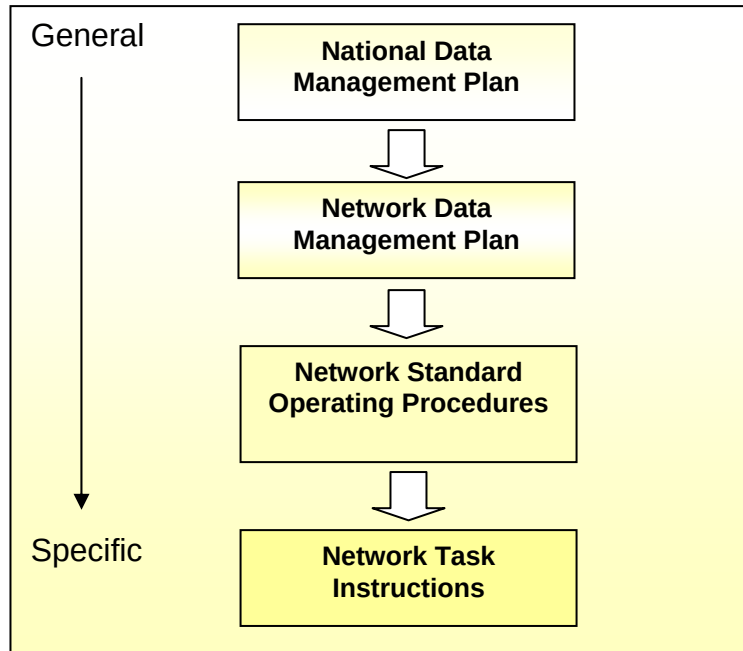


Figure 2. Relationship of the national-level data management plan to network-level data management documents adopted from chapter 1.2 of the Data management guidelines for inventory and monitoring networks (National Park Service 2008).

Data Management and the Inventory and Monitoring mission

Providing sound, quantifiable information to park managers is a laudable goal for the Inventory and Monitoring program. Program directors have acknowledged that while such a goal is worthwhile, it is not easy without technical support. Someone has to keep the hardware spinning and the numbers crunching in order to reliably funnel synthesized information to individuals needing it. Data management fulfills this role by providing expertise in the technical aspects of biological science including client-server databases, application development, metadata generation, web development and geographic information systems. By expressing up-front support for data management the Inventory and Monitoring program is not only publicly committing to a culture of good data management practices but also expressing pride in the program's resulting technical products.

Goals of the Inventory and Monitoring Program

The Inventory and Monitoring program has five broad goals:

1. Inventory the natural resources under National Park Service stewardship to determine their nature and status.
2. Monitor park ecosystems to better understand their dynamic nature and condition and to provide reference points for comparisons with other, altered environments.
3. Establish natural resource inventory and monitoring as a standard practice throughout the National Park system that transcends traditional program, activity, and funding boundaries.
4. Integrate natural resource inventory and monitoring information into National Park Service planning, management, and decision making.

5. Share National Park Service accomplishments and information with other natural resource organizations and form partnerships for attaining common goals and objectives.

Data Management Goals

The Arctic Network's main data management goals are listed below:

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

Data as a Valued Commodity

Data is an expensive asset. Data acquisition is time-consuming, arduous, and typically very costly. Despite the recognition of data's value, data management has, historically taken a backseat to higher priority activities such as publishing results and obtaining funding for new research. The causes for this lack of dedication toward good data management are varied but generally include a traditional lack of:

1. User-friendly software
2. Metadata standards
3. Internet data warehouses and data servers
4. Internet in any consistent and reliable form
5. Expertise in database design and data normalization
6. Requirements from funding agencies to publish data along with results
7. Culture of data stewardship
8. Mandates for data management funding

This situation is changing, however. Good data management practices are becoming more commonplace with time. Organizations are realizing the secondary value of data to large-scale synthetic research works as well as landscape modeling efforts that rely on data from many small, on the ground projects. Well documented and preserved data is also valuable for opportunistic research projects, or background information on which to base estimates of trend or for time-series comparisons. Indeed major funding agencies are now requiring that data be published with results. Government agencies are following suit and the Inventory & Monitoring program is leading the charge. The NPS Inventory and Monitoring program has mandated visionary provisions for data management staffing and funding.

Data Management Expectations

- Inventory and Monitoring data will be held to a high standard of quality
- Data management will be a collaborative effort benefitting everyone in the network. Each stakeholder must commit the time and resources required
- Data management will be a beforethought, rather than an afterthought
- Data management requires support from our collaborators in establishing and reinforcing a culture of good data management practices

Revisions to this Plan

Success seldom emerges fully fledged on the first try. Success is rather the very public culmination of a hidden, cyclical path of trial, failure, learning and revision. As our program progresses we fully expect to make and learn from mistakes. We also will keep aware of new technologies and potential partnerships and attempt to exploit them for the benefit of the program. Most of all, we will not hesitate to modify this document if it will enhance the probability of achieving our goals.

Chapter 2. Data Management Infrastructure and System Architecture

Sattinger's Law: It works better if you plug it in.

Just as no one would expect a carpenter to build a quality home without a good hammer, one cannot expect to build and maintain a data management system without high quality computing resources. As a network we are fortunate that we have a broad array of hardware, software and support services already in place and available to us through the Fairbanks Administrative Center, the Alaska Regional Office and the Washington Support Office. These resources are discussed in chapter two of the National Data Management Plan (National Park Service 2008).

Managing data for an enterprise as large as the Arctic Network requires that information be very centralized and accessible to the level of the least common denominator. Centralized data systems are systems where an enterprise database stores a master copy of all required data. Clients (or data users) use a variety of software applications to log in to a master database. The central database is designed to handle multiple users through user permissions management, record-locking and concurrency management. Such a system ensures that all users have the correct permission to access or edit data, and all users have the latest snapshot of the data. In addition, data dissemination is instantaneous as the data store can update public and private websites, geographic information systems and internet map servers 'on the fly'. Data access issues can be provided via multiple portals to the system Figure 3. In the discussion that follows, systems architecture signifies the applications, database systems, repositories, and software tools that make up the digital framework of our data management enterprise. This chapter will focus on the details of the Arctic Network data management system architecture.

Architecture

The Arctic Network will employ a highly centralized model of data management. At the heart of the system is an enterprise database that will contain not only monitoring and project data, but also project tracking information and a suite of commonly used lookup tables and related information. Our workflow consists of loading high quality information into the database and subsequently extracting data as needed for administration, analysis and reporting. A centralized scheme such as this has numerous advantages over more distributed systems including simplified backups, granular security schemas, minimized dataset versioning problems and efficient common access to data resources to stakeholders. Metadata, and in some cases, data files, will be uploaded to the NPS Data Store as well as the Permanent Data Stack (X:\ Drive) at the Alaska Regional Office in Anchorage.

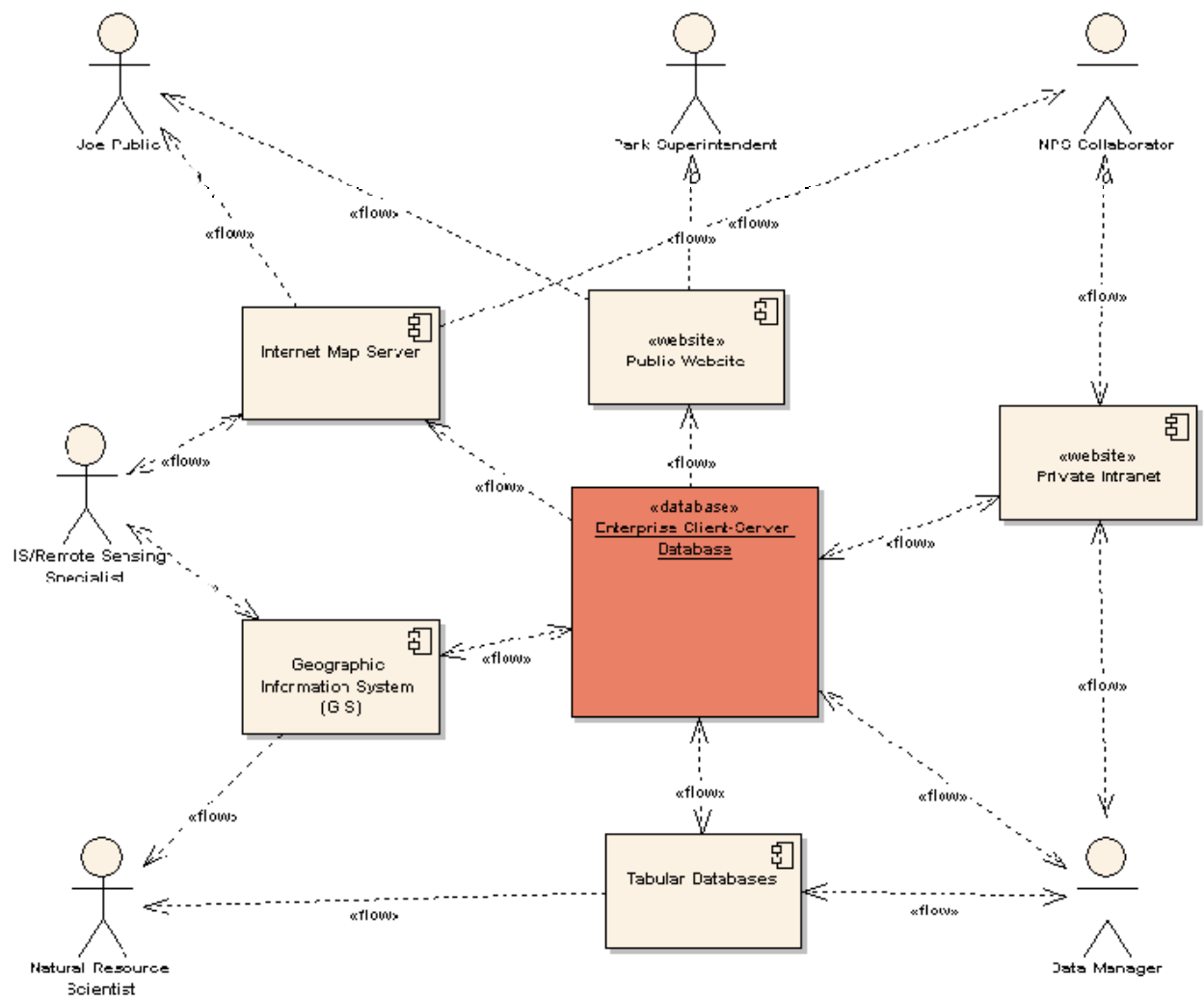


Figure 3. An enterprise client-server database will be the heart of the Arctic Network data management system.

In addition to the primary data server we will employ a common file server with structured directories for each project (NPS 2007c). This shared network file server will store the myriad files generated by projects including everything from administrative details, through draft reports and analyses to working data files. Investigators will be required to use these directories and keep them up to date.

Figure 4 presents a generalized schematic of the Arctic Network data management infrastructure as well as a life cycle for Arctic Network data. While variations to the cycle will occur based on project specifics, the majority of data handled by the Arctic Network will follow this general course.

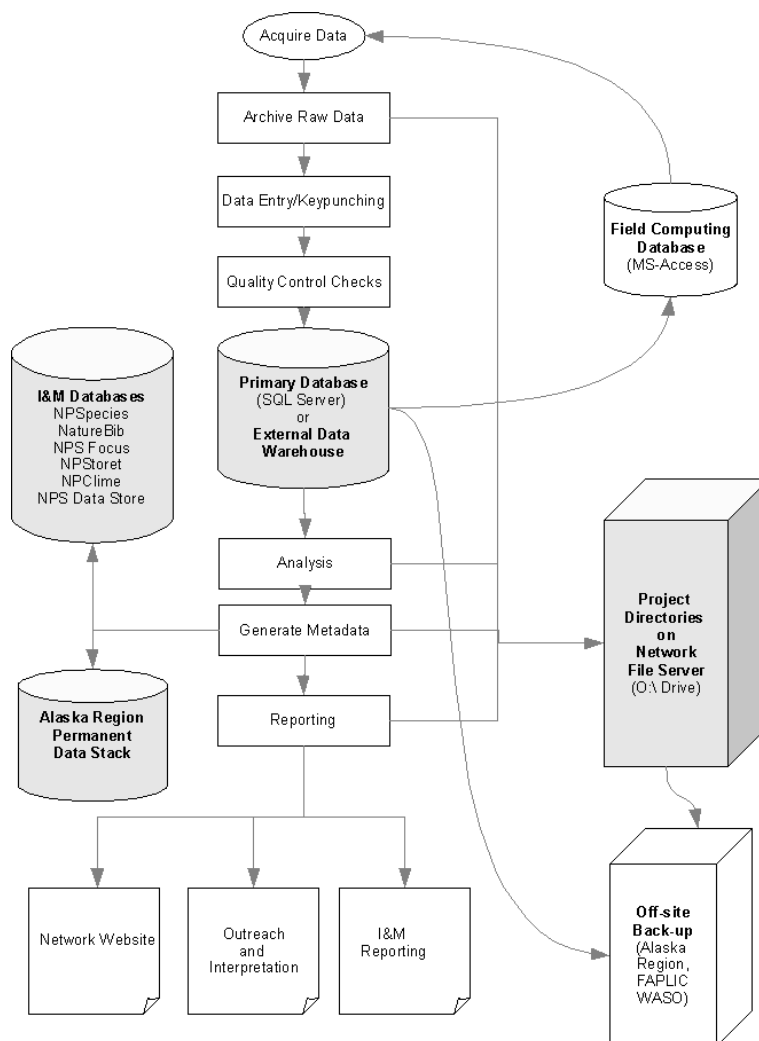


Figure 4. Arctic Network generalized workflow showing major data repositories

NPS Infrastructure

An important element of a data management program is a reliable, secure network of computers and servers (Figure 5). Our digital infrastructure has three main components: park-based local area networks (LAN), network data servers, and servers maintained at the national level. This infrastructure is maintained by park, regional, and national IT specialists, who administer all aspects of system security and backups. Figure 5 illustrates these components.

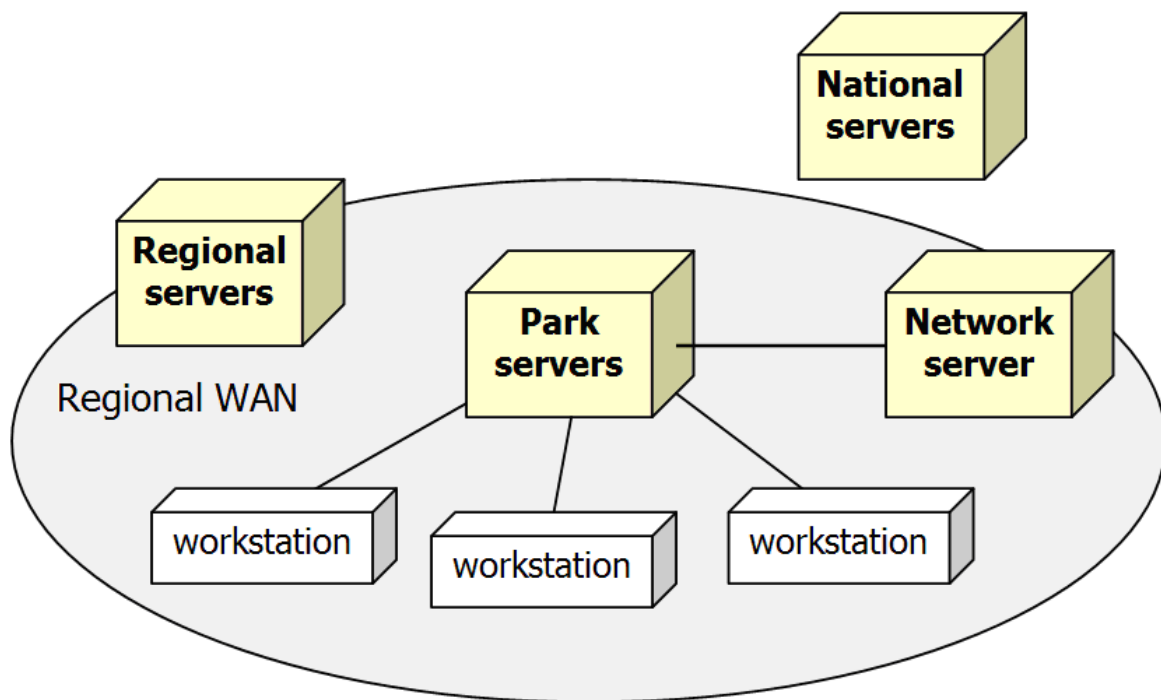


Figure 5. Principle information technology components.

National Applications

The basic tenet of I&M information management is to provide integrated natural resource databases and information systems that provide park staff with relevant, high-quality data for management decisions, resource protection, and interpretation. The NPS Natural Resource Program Center (NRPC) and the I&M Program actively develop and maintain a national-level information management framework and suite of applications that use a combination of desktop applications and Internet-based systems.

The principal applications and data repositories developed for the I&M Program and service-wide programs, and referenced throughout this document, are summarized below (Table 1, NPS 2007i,o). See the National Data Management Plan (National Park Service 2008) for more detail.

Table 1. The principal Inventory and Monitoring applications and data repositories.

Application	Description
Naturebib	Bibliographic Database
Dataset Catalog	Metadata Tool
NPStoret	Water Quality Database
Natural Resource Database Template (NRDT)	Monitoring Application Framework

NPS Data Store	Web-Based Dataset Search Tool
Vital Signs Internet Map Services (VSIMS)	Web Based Spatial Data Dissemination Tool
Natural Resource Monitoring Protocol Clearinghouse	Web-Based Clearinghouse Of Monitoring Protocols
Natural Resources Monitoring Partnerships (NRMP)	Web-Based Clearinghouse Of Monitoring Projects And Protocols
NPS Automated National Catalog System (ANCS+)	Curatorial Cataloging System
Research Permit And Reporting System (RPRS)	Web-Based Research Permit Manager

Alaska Region Infrastructure

The Alaska Regional Office offers a central repository for many types of information and acts as the backbone for base cartographic data in the Alaska region. The regional GIS Team provides support of various spatial data operations via a request process and a liaison assigned to each park. The Alaska Regional Office GIS Team also maintains a server that will act as offsite backup for all Arctic Network data. Implementation of this plan will include arrangement for transfer of Arctic Network data to the Alaska Regional Office GIS server in addition to the NPS Data Store.

To summarize, the Alaska Regional Office provides the following that the Arctic Network will utilize to meet its goals:

- Wide Area Network file server for general file exchange and storage
- GIS and related tabular data accessible via custom applications distributed to the parks as well as the Alaska GIS Data Clearing House (www.nps.gov/akso/gis).
- File server to provide offsite storage for all Arctic Network data

Park and Network-Level Infrastructure

The Arctic Network infrastructure is composed of elements from the national Inventory and Monitoring Program, the NPS Alaska region and Arctic Network-member parks. The Arctic Network relies primarily on the system infrastructure of the Fairbanks Administrative Center and the park headquarters of the Western Arctic Parklands in Nome and Kotzebue. The Arctic Network will manage a primary repository located in the Fairbanks Administrative Center for data and information generated by the Network. These data will be accessible to park personnel via custom applications as well as the network website and shared network drive (O:\ drive, NPS 2007c). The primary Arctic Network repository will be backed up to an offsite server. Applicable portions of the Arctic Network data will be stored in national Inventory and Monitoring Program databases such as NPSpecies, NatureBib and NPS Data Store (NPS 2007i,o). Certain datasets will be maintained by outside organizations, but metadata will be maintained in the primary

Arctic Network repository or the NPS Data Store. Arctic Network operations, where possible, will be designed to function within and augment Park operations (Figure 6).

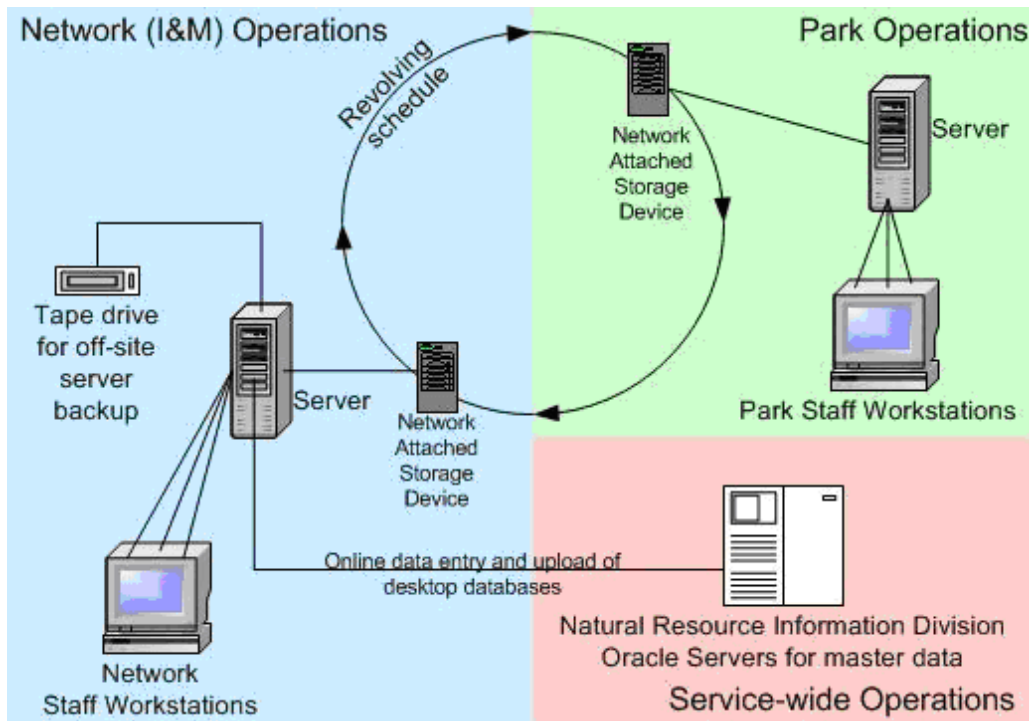


Figure 6. Information technology connectivity between Network, Park and National operations.

Chapter 3. Data Management Process and Workflow

43rd Law of Computing: Anything that can go wr--
root@localhost:>Segmentation violation -- Core dumped

The organizational umbrella under which questions are asked and related data is collected is typically called a study or a project. Inventory and Monitoring projects are of two general forms: 1) Short-term projects, which may include individual park research projects, inventories, or pilot work done in preparation for long-term monitoring and 2) Long-term projects, which are mainly monitoring projects central to the Inventory and Monitoring Program, but may also include multi-year research projects and monitoring performed by other agencies and cooperators. Long-term projects typically require a higher level of documentation, peer review, and adherence to standards to ensure consistency over time. Regardless of the expected duration of a project, each has a life cycle that can be broken into phases. These phases are discussed below.

Project Process

Both short and long-term project types can be divided into five primary stages (Figure 7):

1. Planning and approval;
2. Design and testing;
3. Implementation;
4. Product integration;
5. Evaluation and closure.

Each stage is characterized by a set of activities carried out by staff involved in the project. Primary responsibility for these activities rests with various individuals according to the different phases of a project. Major tasks associated with each phase of the project life cycle are described in Figure 7.

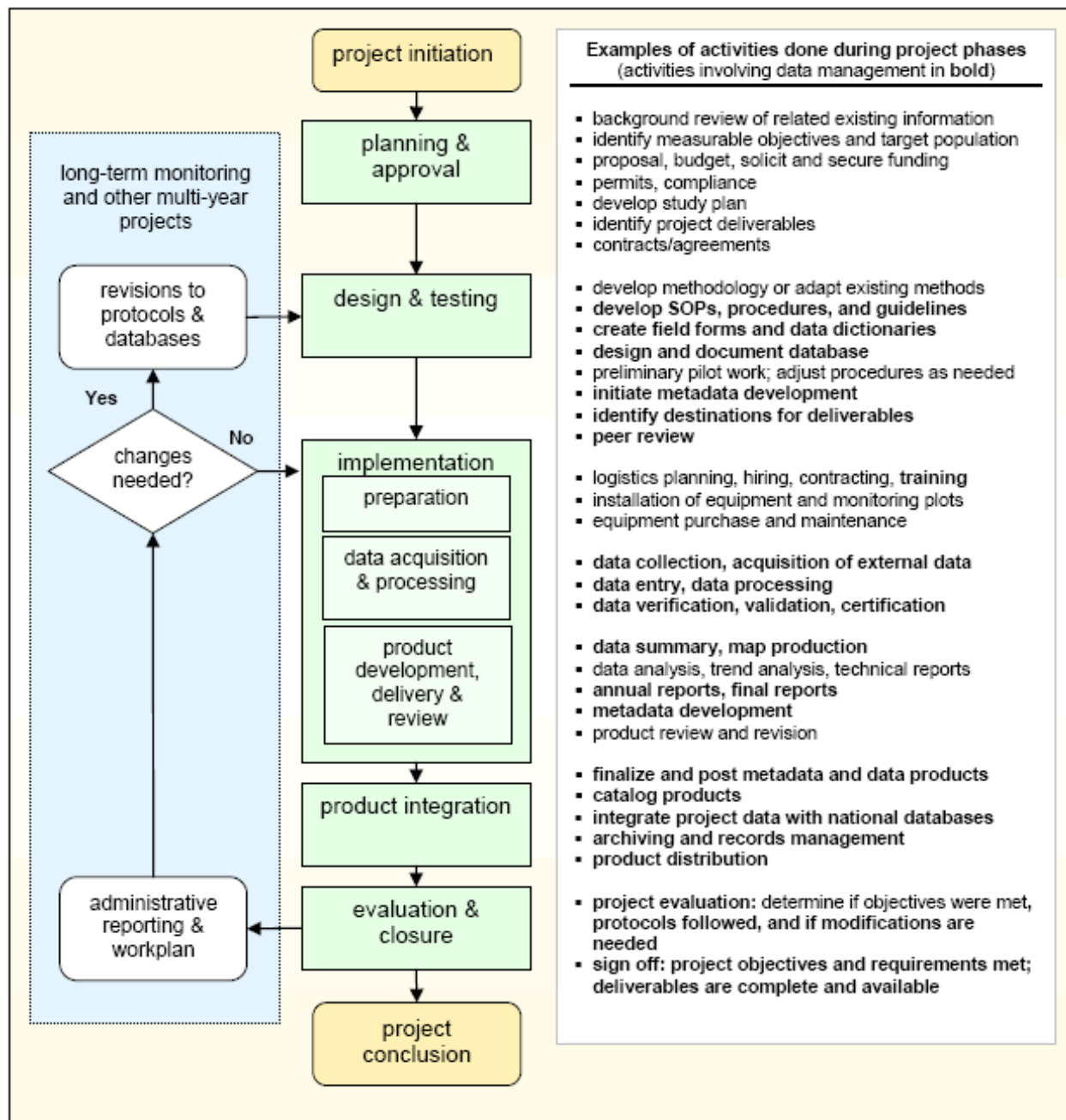


Figure 7. Project life cycle

Planning and Approval

This initial phase is when many of the preliminary decisions are made regarding project scope and objectives. Funding sources, permits, and compliance are also addressed in this phase. Primary responsibility rests with project leaders and program administrators. Data management responsibilities include reviewing contracts, agreements, or permits for language that describes the formats, specifications, and timelines for project deliverables (NPS 2007a,b,d,k,m).

Design and Testing

During this phase, details are worked out regarding how data will be acquired, processed, analyzed, reported, and made available to others. The project leader is responsible for developing and testing project methodology, or for modifying existing methods to meet project objectives. It is essential that the project leader and the data manager work together throughout this phase in order to build the basis for good data management throughout the project. An important part of this collaboration is the development of the data design and data dictionary, where the specifics of data to be collected are defined in detail. Devoting adequate attention to this aspect of project is possibly the single most important part of assuring the quality, integrity and usability of the resulting data. Once the project methods, data design, and data dictionary have been developed and documented, a database can be built to meet project requirements (NPS 2007d,f, Southwood, 2002a,b).

Implementation

During the implementation phase, data are acquired, processed, error-checked and documented. Products such as reports, maps, geographic information system (GIS) themes, and other products are developed and delivered during this phase. The project leader oversees all aspects of implementation, from logistics planning, contracting, training, and equipment procurement, to data acquisition, report preparation and final delivery. Throughout this phase, data management staff function primarily as facilitators: they provide training and support for global positioning systems (GPS), database, and GIS applications, ensure data verification and validation, summarize or format data as needed for analyses, and assist with data documentation and product development (NPS 2007n). The specific roles of data management staff during this phase will vary depending on the project. As much as possible, these roles should be worked out in advance of project implementation. Toward the end of this phase, project staff members work to develop and finalize the deliverables that were identified in the project planning documents (e.g., protocol, study plan, agreement). Raw and derived data products, metadata, reports and other documentation should be delivered to the project leader, who then works with the data manager to move products to their final repositories (NPS 2007k).

Product Integration

During this phase, data products and other deliverables are integrated into national and network databases, metadata records are finalized and posted in clearinghouses, and products are distributed or otherwise made available to their intended audience. Depending on the project, another aspect of integration is merging data from a working database to a master database maintained on the network server. This occurs only after the annual working data set has been certified for quality by the project leader. Certain projects may also have additional integration needs, such as when working jointly with other agencies for a common database. Product integration includes creating records for reports and other project documents in NatureBib, integrating species-related data into NPSpecies, creating or updating metadata files, and posting the resulting products to national clearinghouses such as the NPS Data Store (NPS 2007e,h,i,o).

Evaluation and Closure

For long-term monitoring and other cyclic projects, this phase occurs at the end of each field season, and leads to an annual review of the project. For short-term projects, this phase

represents the completion of the project. After products are catalogued and made available, the program administrator, project leader, and data manager assess how well the project met its objectives, and to determine what might be done to improve various aspects of the methodology, implementation, and formats of the resulting information (NPS 2007a,d).

Chapter 4. Data Management Roles and Responsibilities

Roles and Responsibilities

Data stewardship is the responsibility of all network personnel. All network personnel have a stake in collecting, managing, and analyzing data. It is imperative that everyone's responsibilities be clearly described and that each individual be aware of their duties (NPS 2007a,k). Table 2, below, shows the roles and responsibilities of each person with respect to the project work flow as described in Figure 4.

Table 2. Data stewardship responsibilities at each phase in the project life cycle.

Project Stage	General Description	Principal Responsibility
Design and testing	• Develop methodology or adapt existing methods • Establish data collection protocols • Establish project schedule • Design and test database • Design data sheets	Project Leader Quantitative Ecologist Data Manager GIS Specialist
Implementation	• Record locations, measurements and observations in the field • Digitize source maps • Key in data from a hardcopy source, convert existing data sources • Process images • Prepare summary tables, maps, reports	Project Crew Member Project Crew Leader GIS Specialist Data Manager
Analysis and reporting	• Use data to address monitoring objectives, specifically, status and trends of vital signs • Use data to describe and interpret relationships among vital signs • Synthesize data for different types of end users • Prepare periodic and final reports	Project Leader Quantitative Ecologist
Integration	• Apply procedures that keep data organized and usable; make data available • Document project and data • Archive data and products • Catalog data and products • Publish results • Post on Internet	Project Leader Data Manager GIS Specialist Museum Curator Quantitative Ecologist IT Specialist

Project Stage	General Description	Principal Responsibility
Evaluation	• Evaluate if project meets objectives • Determine changes to methods, procedures, or data design	Project Leader Data Manager GIS Specialist Quantitative Ecologist

Table 3. Network personnel roles and responsibilities

Role	Primary responsibilities related to data management
Project Crew Member	Record and verify measurements and observations based on project objectives and protocols. Document methods, procedures and anomalies.
Project Crew Leader	Supervise crew members to ensure their data collection and management obligations are met, including data verification and documentation.
Data/GIS Specialist or Technician	Perform assigned level of technical data management and/or GIS activities, including data entry, data conversion, and documentation. Work on overall data quality and stewardship with project leaders, resource specialists, and the Network data manager.
Information Technology/Systems Specialist (park IT support staff)	Maintain local area network, establish and maintain system security, and keep software and hardware systems up to date. Additionally, IT staff are responsible for maintaining connections between the LAN and the internet. They work with the Network data manager and GIS liaisons to establish a directory structure (including drive partitions) that provides local access and security for natural resource data. They manage the infrastructure for digital data backups for the local area network.
Network Coordinator	Interface with project leaders to ensure that timelines for data entry, validation, verification, summarization/analysis and reporting are met. Additionally, the Network Coordinator must review and approve proposed changes to project protocols prior to implementation.
Project Leader	Oversee and direct operations, including data management requirements, for one or more Network projects. • Network Data Manager • Maintain communication with project staff • and resource specialist regarding data management
Resource Specialist	Understand the objectives 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 information products.
GIS Manager	Support park management objectives. Coordinate and integrate local GIS and resource information management with Network, Regional, and National standards and guidelines.

Role	Primary responsibilities related to data management
Network Data Manager	and operational support for ,training ,Provide overall Network planning integration of data and information and ,coordination ,the awareness ,data ,information needs ,including people ,management activities software, and hardware Serve as Point of Contact for .National Park Service applications (NPSpecies). Coordinate internal and external data management activities.
Database Manager	Apply particular knowledge and abilities related to database software and associated application(s)
Network Ecologist	Ensure useful data are collected and managed by integrating natural resource science in network activities and products, including objective setting, sample design, data analysis, synthesis, and reporting
Data Manager (National Level)	Provide service-wide database design, support, and services, including receiving and processing to convert, store, and archive data in service-wide databases
Other End Users	These 'information consumers' include park managers and superintendents, researchers, staff from other agencies, and the public. End users are responsible for the appropriate use and application of data and derived products and for providing feedback for improvements.

Chapter 5. Data Acquisition and Processing

Williams and Holland's Law: If enough data is collected, anything may be proven by statistical methods.

Data is dynamic and flows through an organization following a distinct path from acquisition through processing, analysis, reporting and archival. While project process and workflow was discussed earlier the details of data acquisition and processing require more discussion. This chapter describes the general procedures the Arctic Network follows for acquiring and processing natural resource-related data. Procedures will vary depending on the data source, but can be broadly placed into three general categories:

Network Data: Data resulting from projects that are initiated, sponsored, or funded by the Arctic Network.

Other National Park Service (NPS) Data: Data resulting from projects that are initiated, sponsored, or funded by park units, or by regional or national NPS programs.

External data: Data produced or managed by agencies, organizations, or individuals other than the National Park Service.

These categories are addressed in the sections below. In addition we discuss digital file management, software recommendations and the handling of spatial data,

Network Data

Acquisition and processing of biological inventory data is a straightforward process due to the high level of network control over the projects. Most projects are completed by cooperators working under a contract or agreement that has been written specifically by the network for the project. In these agreements Arctic Network itemizes data deliverables and formats that are required from the cooperator. The data manager works closely with the principal investigator and the project leader to develop these data criteria, and to develop a database modeled on the Natural Resource Database Template, where appropriate, that is tailored to the specific data needs of the project. Additional data specifications usually included in agreements are procedures for acquiring and documenting project photographs, collecting specimens, using global positioning systems (GPS), and delivering geospatial data. Data processing typically consists of reviewing the project deliverables, integrating the project into the various Arctic Network cataloging systems, and distributing project results. Principal investigators are responsible for verifying and validating the data they submit; however, Arctic Network performs additional quality checks on all datasets. Data management staff are usually in contact with the principal investigator after the submission of final products in order to resolve discrepancies or answer questions that inevitably arise as staff work with the data.

Data acquisition for network monitoring projects is guided by established protocols that include step-by-step procedures for all aspects of data collection, data entry, quality assessment (QA), quality control (QC), reporting, analysis, distribution, and archiving. Depending on the protocol, data can be acquired by park staff, network staff, cooperators, or other agencies or organizations. Monitoring data need to go through processing and integration steps that are similar to inventory

data; however, many of the steps are repeated annually or on a schedule corresponding to sampling frequency. Each monitoring project will have a processing and integration checklist that is specific to the project to ensure that all steps are completed.

National Park Service Data

A fundamental step in developing and maintaining the information base for Arctic Network is locating, evaluating, and documenting, on an ongoing basis, park-related natural resource information. The term “data mining” refers to this process, which requires regular visits to network parks and establishing data mining procedures specific to each park. The range of materials that require documenting is broad, ranging from historical inventories, to complex databases, to remote sensing data.

As a general rule, the goal of network data mining is not to acquire and store, at the network, copies of all data residing at parks. Instead, cataloging tools are used to describe and document data sources so that potential users can find them. Data that are an exception to this rule are documents or datasets that are of specific need or interest to network staff, or that complement network projects. The highest priority items are past or currently existing monitoring efforts that coincide with our network vital signs.

Parks within the network vary widely in how natural resource information is organized and stored. While some parks have developed efficient and standardized methods of cataloging new information as it arrives, other parks have no formalized approach for organizing their natural resource materials.

Arctic Network works on-site with park staff to locate and catalog their natural resource information. Table 4 summarizes the range of park-based data that the network is cataloging, and the associated processing steps.

Table 4. Examples of park-based natural resource information that is processed or documented by the network.

Data Description	Processing
Gray literature, unpublished reports, research Summaries	Review and document in NatureBib. Evaluate for species data and link to NPSpecies. Obtain copies for network files (preferably electronic) as appropriate; upload full-text document to NatureBib (if not subject to copyright restrictions)
Published reports, journal articles	Review and document in NatureBib. Evaluate for species data and link to NPSpecies. Obtain copies for network files (preferably electronic) as appropriate; upload full-text document to NatureBib (if not subject to copyright restrictions)
Investigator Annual Reports	If IAR is the only reference to a project, document in NatureBib. If a more comprehensive report has resulted from the project, create a NatureBib record for the report and reference the IAR in the NatureBib record.

Data Description	Processing
Photographs	If photos are part of an assemblage accompanied by documentation, enter into NPS Data Store. Evaluate for NPSpecies data (photo vouchers)
Databases, spreadsheets, other electronic data assemblages — undocumented	Review for associated documentation. Locate project leader, if possible, for further explanations. Create NPS Data Store record and document to the extent possible in the Arctic Network data catalog.
Databases, spreadsheets, other electronic data assemblages — documented	Create NPS Data Store record. Evaluate for NPSpecies data. Obtain copies for network files as appropriate
GIS data — undocumented	Retain copy at Arctic Network office as applicable. Document in Arctic Network GIS Inventory database. Create metadata.
GIS data — documented	Retain copy at Arctic Network office as applicable. Document in Arctic Network GIS Inventory database; maintain metadata.
Species lists	Create NatureBib record if author and date are included with list. Evaluate for NPSpecies.
Data collected by NPS regional or national programs (e.g., air quality, sound, night sky)	Create NPS Data Store record
Data collected and managed by other agencies (e.g., stream gauging stations, climate, breeding bird data, other monitoring programs)	Create NPS Data Store record; evaluate for NPSpecies
ANCS+ natural history specimen data	Evaluate for inclusion in voucher table of NPSpecies.
Park library records (ProCite)	Evaluate for inclusion in NatureBib
Park archives	Evaluate for inclusion in NatureBib, NPS Data Store, NPSpecies

Arctic Network data management staff complete most of these cataloging tasks, and have been the primary users of NatureBib, NPSpecies, and NPS Data Store (NPS 2007i,o). Most parks have had no extra staff available to perform this work or to devote to using the databases. Arctic Network places a high priority on maintaining and building these databases, and will continue to perform the work as needed. Data management staff will provide training to park staff interested in learning to use the databases, and will provide reports from the databases in formats that are useful and easily accessible to park staff (e.g., hard-copy printouts, web-based species lists).

Most datasets developed at parks are not maintained in fully-normalized database formats, nor do they need to be. However, datasets designated as being of high priority to the Arctic Network will be converted to current database standards. To avert database problems for new projects, network data management staff are available to help parks as possible or needed with database design, using the Natural Resource Database Template as a model.

External Data

Organizations and agencies outside the National Park Service have been an ongoing source of natural resource data for the network and are included in the network's data mining activities. Processing external data follows the same steps and procedures as processing park data: information is evaluated, entered into one or more cataloging databases, and is then posted or made available as appropriate.

Museums and Herbaria

One of the more challenging network tasks has been acquiring information on specimens that were collected within park units, but that are housed in a variety of repositories across the country (DOI 1993). The Inventory and Monitoring Program at the Washington Office has obtained specimen data from major museums and herbaria that have their holdings in electronic format. Smaller regional and state repositories will be checked as time and budget allows. Frequently the associated and required Automated National Cataloging System (ANCS+) records for these specimens do not exist at the parks. Arctic Network works with park museum curators when receiving and managing specimen data from external sources, and provides them with formatted data for uploading into ANCS+, if appropriate.

National and State Agencies

Occasionally national and state agencies conduct research in national parks. Resulting data may be specific to a park unit, or may encompass a much broader geographic area. The Arctic Network focuses primarily on creating NatureBib records for reports, publications, or other documents resulting from agency projects.

Non-Governmental and Other Organizations

Arctic Network also acquires data from private and non-profit organizations. References to theses and dissertations related to park research are located through on-line database searches, as are journal articles and other publications that reference park-based research.

GIS Data

The Arctic Network will leverage the GIS data store housed by the Alaska Regional Office in Anchorage. Spatial information for the Alaska parks are searchable via NPS Theme Manager software. Spatial data is replicated to the Fairbanks Administrative Center nightly and available to all network personnel via a mapped network drive. Spatial data collected by the Arctic Network will be archived with associated metadata with the Alaska Regional Office and the NPS Data Store.

Data Management Process

The major elements in the process for executing data management within the framework may be summarized as:

- Data acquisition (field work, satellite download, etc.)
- Raw data archival (physical and possibly digital material)
- Data entry/import
- Quality assurance and control)

- Documentation (metadata, etc.)
- Analysis & reporting
- Data cataloging, documentation, and archival (digital)

The elements listed above roughly comprise a data life cycle within the Arctic Network. Figure 4 presents the relative timing of these elements within a project (NPS 2007a). For the life cycle to have actionable meaning, we must also specify how and where these core data management elements will be done. For the most part, these are documented in the standard operating procedures for each vital sign. There are, however, some commonalities to these procedures dictated in this plan.

Figure 7 presents a generalized schematic of the data life cycle for Arctic Network data. While variations to the cycle occur based on project specifics (e.g. in the case of climate or air quality data which will be managed by outside entities), the majority of data handled by the Arctic Network will follow this general course.

Acquisition of New Spatial Data

Spatial data acquired by the network must meet established standards and product specifications (NPS 2007a, NPS 1994). Additional standards or requirements may be identified on a project-by-project basis. Spatial data gathered using global positioning system (GPS) units must also meet established standards (NPS 2007n). The Alaska Regional office maintains a GPS guidance website (<http://www.nps.gov/gis/gps/>) that the Arctic Network will use. This website addresses GPS instruments, instrument settings, field operation, data collection, data processing, and a standardized method for acquiring and managing location data.

Maintaining Digital Files

The Arctic Network maintains hierarchical directory structures for storing project-related digital files (NPS 2007c). The key aspects of this file management strategy are as follows:

- Provides a consistent and logically organized structure that complements the Network's day-to-day operations.
- Minimizes confusion about where information is located.
- Allow for efficient, comprehensive, regular and secure means of backing up information.
- Accommodates various security settings for different types of users;
- System is hierarchical and minimizes the ambiguity of where information should go. Ideally, the organization should be logical enough that one could navigate through the entire structure without requiring a map;
- Separate in-progress/draft/unofficial/in-progress files from final/official information;
- Be broad enough to clearly accommodate all types of information handled by the Network staff;
- Be simple without sacrificing functionality;
- Minimize the number of sub-folder levels and also refrain from having more than 8-10 folders at any given level;

- Information should never be arbitrarily divided in a way that increases the possibility of corruption or loss (e.g., metadata should never be separated from the data)
- When possible and appropriate, the structure should conform to other NPS and national directory structure standards
- Standards such as naming conventions and hierarchical filing are enforced within the Libraries, Database, and GIS sections. Although less stringent in other sections, these conventions are encouraged as good practice.

Our system can be summarized as follows:

- The Arctic Network will manage a primary repository located in the Fairbanks Administrative Center for data and information generated by the Network. These data will be accessible via national databases, custom desktop and web applications as well as the Arctic Network website and will be open to authorized personnel.
- The primary Arctic Network repository will be backed up to an off-site server in the Alaska Region.
- Required portions of the Arctic Network data will be stored in national Inventory and Monitoring Program databases such as NPSpecies, NatureBib and NPS Data Store.
- Certain Arctic Network datasets will be maintained by outside organizations however metadata for these will be maintained in the primary Arctic Network repository or the NPS Data Store.

Data Discovery

It is impossible to predict wholly the demands which will be made on a data interface for the complex body of data and information the Arctic Network intends to manage. We assume, however, the following basic needs:

- Robust data browsing:
 - View summary results from any one vital sign measure
 - Results comparison between vital signs
 - View monitoring results by time and location
- Robust data access (download):
 - Easy data subsetting (data searches that allow data to be selected by data, location, subject, etc.)
 - Useful format choices
 - Dissemination tracking
 - Accompanying metadata
- Robust data upload for Project Leaders and authorized personnel
- Robust data discovery so that our data and information may be easily found by outside stakeholders

Our data management applications, both desktop and web based must reflect the above requirements.

Chapter 6. Quality Assurance and Quality Control

Sturgeon's Law: 90% of everything is crud.

Weinberg's Second Law: If builders built buildings the way programmers wrote programs, then the first woodpecker that came along would destroy civilization.

Scientific analyses and conclusions are only as good as the data from which they are derived. Our ability to detect trends or patterns in ecosystem processes is enhanced by data of documented quality that minimize error and bias. Data of inconsistent or poor quality can result in loss of sensitivity and incorrect interpretations and conclusions. The Arctic Network monitoring program will bring together analyses from an array of widely-varying projects aimed at detecting changes in the environment over time. The inherent complexity in this endeavor demands stringent and consistent quality assurance (QA) and quality control (QC) measures be applied throughout the program.

Quality assurance involves planning quality into the data while quality control consists of monitoring the system or appraising the product after the product is developed. Data management must ensure that our projects produce and maintain data of the highest possible quality. The Network will establish and document protocols for the identification and reduction of error at all stages in the data lifecycle (NPS 2007b,c,d,e,f, h). These stages include project planning and database design, data collection and entry, verification and validation (certification), documentation (including data quality and sensitivity review), processing, and archiving (Figure 7). The final stage in the data life course is dissemination/integration (NPS 2007l). Detailed QA/QC procedures for these stages are included in the protocols for each project initiated by the Arctic Network (Oakley et al. 2003). This chapter of the data management plan presents more broadly based procedures and policy that govern specific operations within a Arctic Network project. Figure 8 illustrates selected QA/QC procedures in context of the amount of control necessary and data confidence.

National Park Service Mandate for Quality

NPS Director's Order #11B: "Ensuring Quality of Information Disseminated by the National Park Service," issued in 2002, promotes information and data quality. 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, from the perspectives of both the Network and the general public.

Integrity refers to the *soundness* of the data or the confidence one has in the data. Integrity is related to objectivity; however it is possible to have subjective data of high integrity. The

integrity of data is also related to data security, 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, which are accurate, timely, and representative of the most current information available. These standards apply not only to NPS-generated information, but also to information provided by other parties to the NPS if the NPS disseminates or relies upon this information. High quality data and information are mandated by directives and orders, and they are vital to the credibility and success of the Inventory and Monitoring program.

Data Quality Goals and Objectives

The amount of quality assurance and control effort needed is dependent on the scope of a project. In an ideal world, no expense would be spared to generate data of immaculate quality. In reality, the best strategy is to maximize the returns of data quality with respect to available resources and priority of results. These trade-offs are summarized in Figure 8.

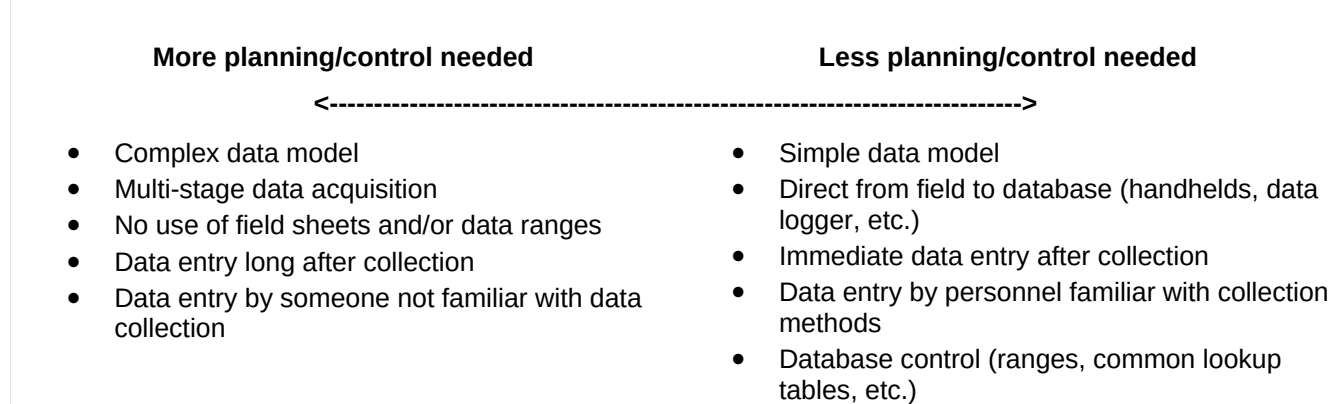


Figure 8. Some common data management elements affecting degree of need for QA/QC. Planning and training for data collection (QA) and entry is always a premium

The most effective mechanism for ensuring that a project produces data of the right type, quality, and quantity is to provide procedures and guidelines to assist the researcher in accurate data collection, entry, and validation. As part of data management operations, the Arctic Network will develop a comprehensive set of SOPs and tools for quality assurance and control in field procedures, data entry/validation/verification and data use (Oakley et al., 2003). This will likely include development of new field-computer-based methods for acquiring data as well as custom programming of database interfaces to both protect the data and control how the data are viewed or summarized.

Quality Assurance and Control Duties

Everyone involved with the handling of project data plays a part in producing and maintaining high quality data. Selected QA/QC roles and responsibilities are shown in Table 5, below.

Table 5. Data quality roles and responsibilities

<i>Actor</i>	<i>Role</i>
Project Manager	• Be aware of quality protocols and convey their importance to technicians and field crews • Ensure compliance with the protocols • Plan for and ensure proper execution of data verification and validation • Review all final reports and information products
Technicians	• Follow established protocols for data collection, data entry, and verification established in the project SOPs
Data Manager	• Develop Network-wide protocols and SOPs to ensure data quality • Make project managers, 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 data quality control operations

QA/QC Process Cycle

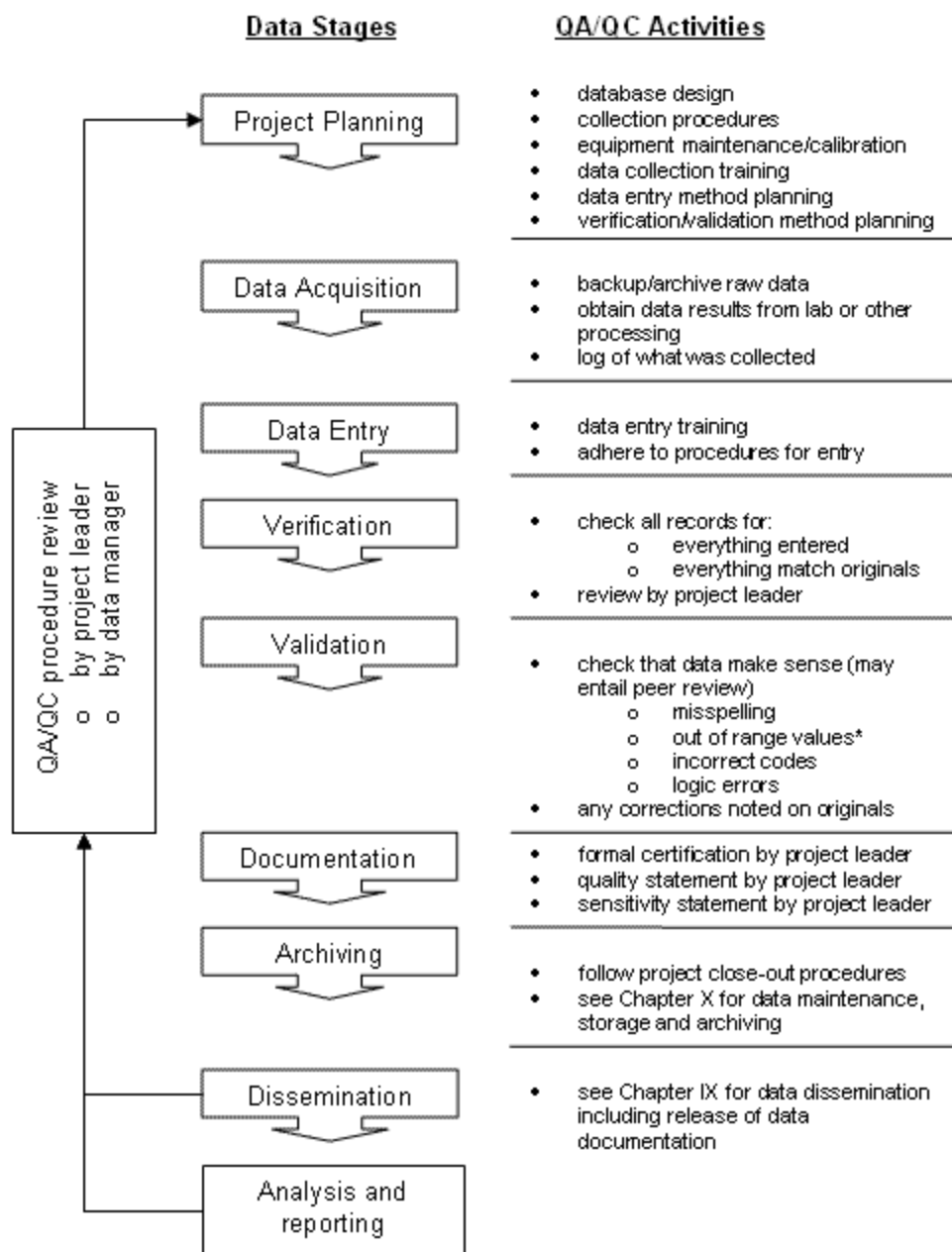


Figure 9. General course of data and associated QA/QC procedures. Quality control with regards to data analysis is specific to each project and addressed in appropriate standard operating procedures.

Data quality assurance and control is a process that expands on the project life cycle described earlier in this document. Figure 9 shows the general course of Arctic Network data and associated QA/QC procedures. Quality control with regard to data analysis is specific to each

project and addressed in appropriate standard operating procedures. Each step in the cycle shepherds the data through collection, verification, documentation and archival.

Project Planning and Database Design

Project planning is of the utmost importance for ensuring success and this is especially true for data management and database design (Southwood 2002a,b, NPS 2004, NPS 2006b,f). We hope to get beyond traditional methods of data management whereby thought is only given to database design long after the field season is over. Experience has taught us that the amount of planning that went into a data product is clearly communicated through the consistency, logical integrity and accuracy of the raw data. Poor data management should not be disguised by misleading data summaries, maps, graphs and statistics. We hold that raw data invariably confesses data management habits, good or poor, during quality assurance and control efforts. Good habits produce good data. To avoid problems at the outset we have instituted a policy of only dealing with structured data. Our rule of thumb is that if the chosen data management software does not require you to format the data it will contain *before* data entry (i.e. databases, not spreadsheets) it is not acceptable. To even the load of data management tasks throughout the year, we will develop and provide each investigator with an annual calendar of tasks, milestones and deadlines that should be accomplished or met during the project life cycle (NPS 2007a). In addition we will institute good practices that will maximize data quality including:

- Adherence to principles of data normalization, business rules and programmatic quality control mechanisms
- Use of appropriate software (generally relational databases)
- Pre-planning for data management early in the project development cycle.
- Common lookup tables for values of parameters recorded in the field by more than one project (such as common weather metrics)
- Each project has SOPs for core data management (collection, entry, verification, etc.)
- Each project uses, at a minimum, standard, pre-printed field sheets for data collection.
- Data entry computer applications resemble field sheets
- Automated error checking features be included in database applications.
- Database application design will maximize the use of auto-fill, auto-correct, value range limits, pick lists, and other constraints specific to projects
- Database applications will include a means to track errors reported on the data after dissemination.
- Database maintenance logs will be maintained for each Arctic Network database and housed in association with database files.

Database Design - Record-level Tracking

As a standard part of database design, the Arctic Network will build into database tables fields that track at the record-level who entered the data, precise entry time and the protocol version under which the data were collected. The benefits to overall data integrity outweigh any inconveniences this overhead data may cause.

Lookup Tables

The Arctic Network will utilize to the fullest extent possible common lookup tables for variables recorded by multiple projects. Examples include weather variables (such as precipitation

intensity, wind speed, etc.), standard equipment and settings (e.g., GPS models and datums) and field personnel.

Project SOPs

Each vital sign protocol will include SOPs that address core data management practices with quality control in mind. These may include:

- Field crew training (addressing both data collection and entry)
- Standardized data sheets
- Use of handheld computers
- Equipment maintenance and calibration
- Procedures for handling data (including specimen) in the field
- Data backup, entry, verification and validation

Data Acquisition

Attention to detail during data collection is crucial to overall data quality. The Arctic Network adopts the following precepts regarding data collection that affect data quality.

- At a minimum, data will be collected on formatted, project-specific data sheets that reflect the overall data design for the project and maximize limitations on values that may be recorded for different parameters. Sheets will be designed to minimize the amount of writing necessary to effectively record observations. Data sheets will be printed on a stock appropriate to field conditions.
- Any project using field equipment will include an SOP for calibration (including the timing of calibration) and maintenance. Such SOPs will specify establishment of an equipment maintenance log.
- Edits on field sheets while in the field will be done by drawing a single line through the information to be changed and adding any replacement information in clear writing next to the original entry.
- Completed field forms will be proofed for errors each day in the field.
- Wherever possible and appropriate, data loggers or field-based computers will be used to collect data. The use of handheld computers requires separate SOPs to describe their use and will include direction for daily review and back up of digital data.

Standardized data sheets that identify the pieces of 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:

- All information added to the data sheet must be printed and clearly legible.
- 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 copies of the data

sheets will be stored as specified in the protocol SOP, and the original data sheets will be used for data entry.

Data Entry

‘Data entry’ is the initial set of operations where we transfer raw data into a computerized form linked to database tables. 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 in an office computer where they can be further manipulated. The goal of data entry is to transcribe field observations into a computer database with 100% accuracy. That is, exactly what was recorded in the field should be entered into the database. Subsequent data verification is conducted to ensure that raw data matches entered data. Following verification, data validation may result in changes *to the entered data*. Data entry is a separate operation from data validation and care must be taken to not impose validation (beyond that automatically imposed by programming rules in a database) during data entry.

The Arctic Network adopts the following precepts affecting data quality with regards to data entry:

- Data will be entered as soon as possible after collection.
- Data entry will be done by someone familiar with data collection. The project leader (with assistance from the data manager if needed) must ensure that data entry staff is familiar with the database software, database structure, and any standard codes for data entry used by the Network. At minimum, data entry technicians should know how to open a data entry form, create a new record, and exit the database properly. They must learn how to commit both a ‘field’ entry and a “complete record” entry and to correct mistakes made while typing.
- If possible, data will be entered by two qualified persons; one to read the observations and one to enter.
- Data will be entered into pre-designed database applications that resemble field sheets and maximize error control. Data will not be entered into spread sheets.
- To the extent possible, data entry will be automated. This may simply entail downloading of data from field-based computers but may include the application of new technology to allow for machine-driven data entry.

Verification and Validation

We appraise data quality by applying verification and validation procedures as part of the quality control process. These procedures are more successful when preceded by effective quality assurance practices (planning). *Data verification* checks that the digitized data match the source data, while *data validation* checks that the data make sense. It is essential that we validate all data as truthful and do not misrepresent the circumstances and limitations of their collection. Failure to follow SOPs for data entry, validation, and verification will render a data set suspect. Although data entry and data verification can be handled by personnel who are less familiar with the data, validation requires in-depth knowledge about the data.

The Arctic Network adopts the following precepts regarding data verification:

- Options for verification methods are given below. Project leaders are responsible for specifying in the project protocol one or more of these methods. At the discretion of the project leader, additional verification methods may be applied.
- The project leader is responsible for proper execution of data verification.
- Data verification is carried out by staff sanctioned by the project leader who are ideally familiar with data collection and entry.
- 100% of records will be verified against original source data.
- 10% of records will be reviewed after initial verification by the project manager. If errors are found, the entire data set verified again.
- A record of each dataset's verification process including number of verification iterations and results will be prepared by the project leader as part of formal metadata generation.
- Certain crucial datasets may require duplicate entry, in which two people enter the same data and then the results are compared for disparity and corrected.

Methods for Data Verification

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

1) *Visual review at data entry.* The data entry technician verifies each record after input. Recorded values in the database are compared with the original values from the hard copy. Identified errors are immediately corrected. This method is the least complicated since it requires no additional personnel or software. Its reliability depends entirely upon the person keying data and thus, is probably the least reliable data verification method.

2) *Visual review after data entry.* All records are printed upon completion of data entry. The values on the printout are compared with the original values from the hard copy. Errors are marked and corrected in a timely manner. When one technician performs this review, the method's reliability increases if someone other than the person keying data performs the review. Alternatively, two technicians can perform this review. One technician reads the original data sheets (the reader), and the second reads the same data on the printout (the checker).

3) *Duplicate data entry.* The data entry technician completes all data entry, 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 database query is then used to automatically compare the duplicate records from the two datasets and report any mismatches. Disparities are manually reviewed and correction applied if necessary. This method adds the overhead of retyping the selected records, as well as the creation of a comparison query. However, it becomes increasingly successful as the value of n decreases. Professional data entry services frequently use this method.

Supplementary Methods

Calculate simple summary statistics and range checks with statistical software to identify duplicate or omitted records. For example, we can view the number of known constant elements, such as the number of sampling sites, plots per site, or dates per sample. We can pose the same question in different ways; differences in the answer provide clues to errors. The more checks we

devise to test the completeness of the data, the greater our confidence that we have completely verified the data.

Validation

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

General step-by-step instructions are not possible for data validation because each data set has unique measurement ranges, sampling precision, and accuracy. Nevertheless, validation is a critically important step in the certification of the data and a required component of any Arctic Network project protocol. 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., pH). 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). We call these types of erroneous data *logic errors* because using them produces illogical (and incorrect) results. The discovery of logic errors has direct, positive consequences for data quality and provides important feedback to the methods and data forms used in the field. Histograms, line plots, and basic statistics can reveal possible logic and range errors.

The Arctic Network adopts the following precepts regarding data validation:

- Each Arctic Network project protocol will address a process for data validation that includes at least one of the methods outlined in the Network data management plan. That process will adhere to the precepts given in the plan.
- Corrections or deletions as a result of data validation require notations in the original paper field records about how and why the data were changed.
- Modifications of the field data should be clear and concise while preserving the original data entries or notes (i.e., no erasing!).
- Validation efforts should also include a check for the completeness of a data set since field sheets or other sources of data could easily be overlooked.
- The Arctic Network will maximize the use of automated routines and/or data summary/visualization such as histograms, line plots, and basic statistics to reveal possible logic and range errors.
- The Arctic Network will maximize the use of database programming to control data entry. In general, this will be achieved via the use of lookup tables but may be accomplished by field-type design in a data base (such as “yes/no” field-types).

Methods for Data Validation

The following general methods can be used to validate data. Specific procedures for data validation depend upon the vital sign being monitored (project subject) and will be included in the project protocols.

1) *Data entry application programming*. Certain components of data validation are built into data entry forms. This method is essentially part of database design. 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 are important.

Caution must be exercised when using lookup tables to constrain variable values. Values occurring outside the range set by a lookup table (established during database design) may not always be invalid. As part of data validation procedures, the project leader is responsible for correct use of lookup tables or other automated value range control.

2) *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 a better way to explain this quality assurance goal. If contamination is not detected during data collection, it is usually only detected later if an outlying data value results. When we detect an outlier, we should try to determine if some contamination is responsible. We can use database, graphic, and statistical tools for ad-hoc queries and displays of the data 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 documentation of the data set saves other users from checking the same unusual values.

3) *Other exploratory data analyses*. 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 (the authors provide examples of procedures that may be applicable). Normal probability plots, Grubb’s test, and simple and multiple linear regression techniques may also be used ([Edwards (2000)]; the author provides SAS and S-plus code for constructing normal probability plots and examples of output showing normal and non-normal distributions).

Data Quality Review and Communication

The National Park Service requires QA/QC review prior to communicating/disseminating data and information. Only data and information that adhere to NPS quality standards may be released (NPS 2007l). Data and Information disseminated to the public must be approved by the appropriate reviewing officials and programs. Documentation of the QA/QC standards used in producing the information and that substantiate the quality of the information must be formally certified and distributed with related data and information. Also, mechanisms must be in place for receiving and addressing comments/complaints pertaining to the quality of data.

Conformance

As part of the close out and evaluation stage of each Arctic Network project, QA/QC procedures will be reviewed by the project leader and any recommendations for change made in the project annual report. Additionally, the Arctic Network data manager will review the QA/QC procedures included in the data management plan and revise as needed. While this review will be part of the regularly scheduled plan review, QA/QC procedures are subject to revision as needed by the Network data manager.

The data manager will conduct periodic “spot checks” of random Arctic Network monitoring projects to ensure compliance with data management plan and project protocol QA/QC procedures. The data manager will track and facilitate the correction of any deficiencies found during this process. The data manager will submit a report of findings to the project leader and the network coordinator within a month of completing any review of QA/QC procedures. The project manager and Network data manager are responsible for ensuring that non-conformities in data management practices are corrected.

Periodic checks by the data manager to see if network staff are adhering to the data quality procedures established in the Data Management Plan and protocols SOPS may include verification of the following:

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

Documenting and Communicating Quality

The final step in the QA/QC for a given dataset 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 as well as the Arctic Network primary data repository. Metadata for each dataset/database will also provide information on the specific QA/QC procedures applied and the results of the review. Typically, data quality information will be conveyed as part of FGDC-compliant metadata. Metadata and data will be available via both the Arctic Network website and the NPS Data Store.

Data and File Management Operations

File Naming Standards

Because of the scope and quantity of files being consolidated into one place, it is critical that computer files be given specific and unique names that will uniquely identify them indefinitely. The following conventions apply to all Arctic Network file names:

- No spaces or special characters within the name
- Use date for version control, as described below
- Use underscore as delimiters

File names for final products will begin with 'ARCN' followed by a the project identification (if project-related), a brief title and a version date in YYYYMMDD format. Each of these file name elements (excepting the extension) is preceded by an underbar (“_”). Additional underbars may be added for clarity. As an example “ARCN_draftData_Management_Plan_20060822.odt” indicates a draft version of the data management plan dated August 22, 2006.

“ARCN00022_StudySiteLocations_20051112.csv” indicates a file related to project 22 that contains study site locations in comma separated values format.

Version Control

Prior to any major changes to a file, a copy of the file with the appropriate version control (in this case the file name is the version control) is made. As indicated above, files are stored with the appropriate eight digit date which serves as version control. This allows the tracking of changes over time and facilitates collaboration between multiple personnel working on common files. With proper controls and communication, versioning ensures that only the most current version is used in any analysis. Using the date, formatted as YYYYMMDD, provides a logical version control and also allows files to easily be sorted by date within a file manager. Working directories should have an '/old' directory. Major revisions should be accomplished by copying the latest version, amending the date portion of the file name and archiving the old version in the /old directory. When the final version is complete the contents of the /old directory should be deleted or compressed and archived (i.e. 'zipped') off the network.

Laboratory Data

All data for a given vital sign, including lab results, will be housed in a distinct relational database accessed via custom applications. Laboratories that will be entering analysis results for a given vital sign will be supplied with a copy of the application so that data may be entered in the manner and format matching that of the rest of the data for a given monitoring parameter. While most professional laboratories exercise their own QA/QC procedures, results received by a project leader are subject to the same QA/QC measures exacted on other project data.

Chapter 7. Data Documentation

Arnold's Laws of Documentation:

1. *If it should exist, it doesn't.*
2. *If it does exist, it's out of date.*
3. *Only documentation for useless programs transcends the first two laws.*

Formal, standard metadata serves as a unifying document for all sources of information about project data (Figure 10). While metadata are addressed in context throughout this plan, this chapter outlines the Arctic Network strategy for generating and handling metadata.

The information needed to understand and use a project's data is embedded in project documents such as study plans, reports, and protocols. Data design documents such as database descriptions and field sampling protocols are often critical to effective analysis of project data. Project tracking applications used by managers may also contain indispensable information about project data. Further, standard data handling such as backup, entry, verification and validation may produce information that affects the way in which the data may be used.

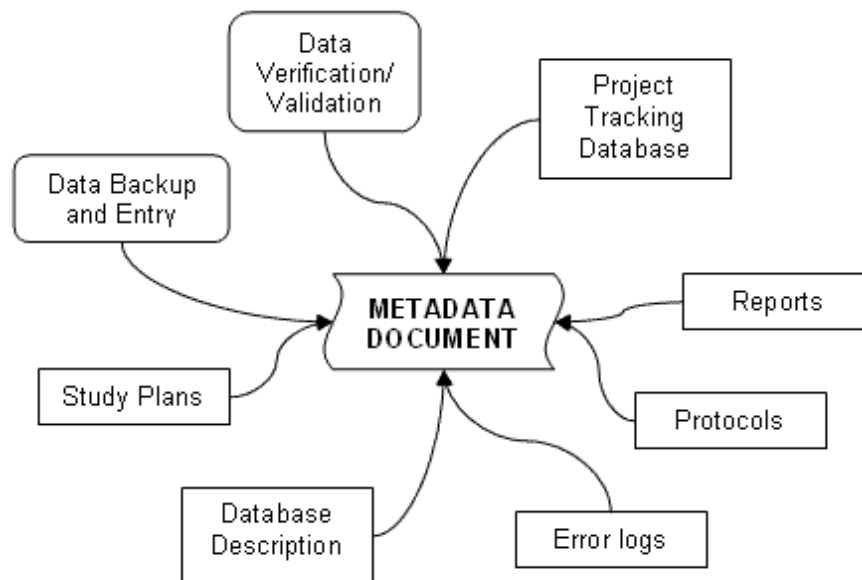


Figure 10. The metadata document contains all information pertaining to project data.

The metadata document will be critical to project organization. The FGDC standard is broadly used and will allow for easy and rapid transfer and sharing of project details, both within the NPS and without.

Metadata Profile

Executive Order 12906, mandates that federal agencies “...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). In addition, EO 12906 directs agencies to plan for legacy data documentation and provide metadata and data to the public.

The FGDC Biological Data Profile contains all the elements of the CSDGM and includes additional elements for describing biological datasets. Metadata created in compliance with the Biological Data Profile can be added to the National Biological Information Infrastructure Clearinghouse. Although not a requirement, completion of the Biological Data Profile for appropriate datasets is recommended.

Documentation Roles and Responsibilities

Documentation duties may be summarized as follows:

Project Leader	Data Manager
• Document project planning, execution and progress in reports, correspondence, project-tracking database, etc. • Generate formal dataset metadata	• Facilitate metadata generation by providing tools and assistance in their use • Establish metadata dissemination protocols (cataloguing of metadata on accessible servers) • Format metadata for dissemination

The duties for the project leader represent a broad scope of tasks, including direction of project personnel and general project management; all aimed at producing a comprehensive metadata document.

Metadata Tools

All GIS data layers must be documented with applicable FGDC and NPS metadata standards. The NPS GIS Committee requires all GIS data layers be described with FGDC standards and the NPS Metadata Profile. Numerous tools exist for developing metadata, but the most appropriate tool is the NPS Metadata Tool and Editor since this tool integrates with the NPS Data Store and provides a robust metadata editing environment. The end goal, however, is a formal, comprehensive metadata document that meets federal (FGDC) and NPS standards. The tools and means of achieving this end product are numerous and changing. Project leaders may use any tool deemed suitable for the project that meets program documentation requirements.

Metadata Process

Figure 7 and Figure 9 indicate the points within the standard project workflow for metadata development. Metadata development begins with project design and planning (Figure 4). Personnel generating metadata should follow the excellent standard operating procedures provided by the NPS Data Store (NPS 2005a-f).

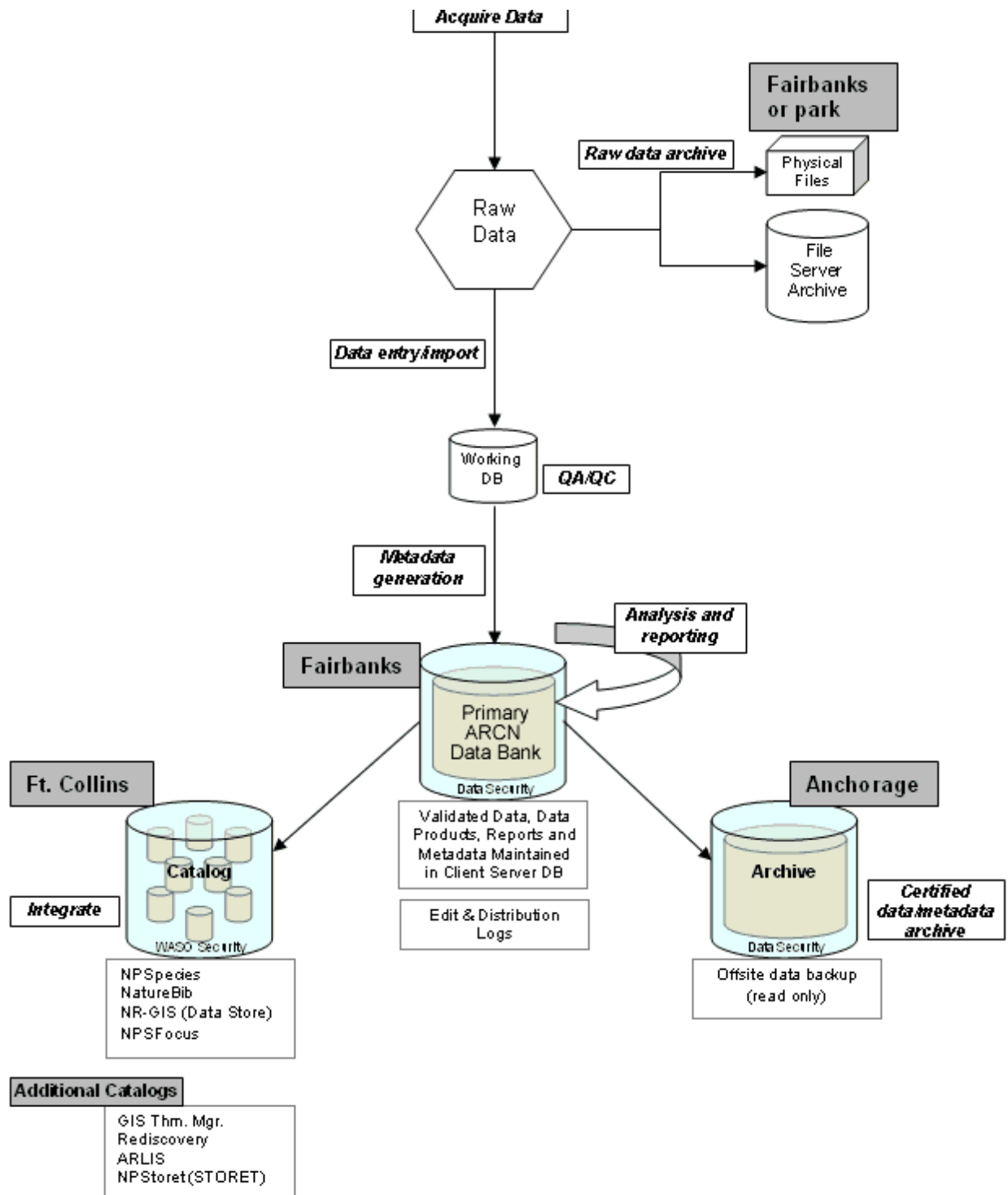


Figure 11. Generalized path of data within the Arctic Network. The primary data server is accessed via desktop and web-based applications. Analyzed or derived data products are stored on the primary server, archived and catalogued (integrated) along with certified data.

In general, a metadata document will be initiated in the design and planning stage of a project. As the project progresses, this document will be augmented by the project leader to include relevant project details. Full development of metadata for project data will be completed after the dataset is certified by the project leader. Metadata are saved on the primary Arctic Network server along with data and copied to the Inventory and Monitoring Program NPS Data Store (Figure 10). Hypertext links within the metadata document will point to accompanying datasets served on the primary Arctic Network server. In general, a single metadata document will apply to both raw and certified versions of the data. Derived data (typically data resulting from the analysis of certified data) may require a separate metadata document. Metadata records will be stored with both hard copy and digital archive data.

Metadata Maintenance

The Arctic Network SOP for metadata generation and maintenance will provide details on the mechanics of handling metadata. Generally, metadata will be created by the project leader, with assistance from the data manager. The Network data manager ensures that updated metadata are copied to the NPS Data Store.

Protocol Versions

Arctic Network data will include a means for recording the protocol version under which each piece of data was collected. The metadata document will contain information about protocol versions used to collect the data.

Non-Program Data Documentation

Data which are generated and/or managed outside the Inventory and Monitoring Program but used in analysis with Arctic Network data or distributed in any manner by the NPS require the same level of documentation produced for Arctic Network-generated data. This includes data produced under contract with the NPS. Metadata for non-program data will be requested from the originating entity by the project leader. Any contracts entered into by the Arctic Network with data producers will stipulate the submission of FGDC-compliant metadata. The Network data manager will assist the project leader in metadata acquisition by providing tools, format protocols and file transfer services.

Derived Data Documentation

Like regular project data, derived data products may not be released without proper documentation. Metadata for any such derived data will include a data lineage pointing to parent datasets.

Legacy Data Documentation

Data acquired via data mining or other means does not enjoy a level of stewardship afforded project data. When older data that have no current steward (typically a project leader) are found to be useful, metadata for those data must be established. No data may be released without accompanying metadata. If ancillary data are used in the creation of derived data products (used in analytical comparison to project data), metadata for those data are necessary to fully document the derived data product.

Chapter 8. Data Analysis and Reporting

Velilind's Laws of Experimentation:

1. *If reproducibility may be a problem, conduct the test only once.*
2. *If a straight line fit is required, obtain only two data points.*

Our network's mission is to provide sound science for park management decision-making. Generating meaningful results is a cornerstone of the Inventory and Monitoring Program. The associated data management role is to provide valid data in formats that support the statistical design of our monitoring program as well as ad hoc querying, analysis and reporting of data. Data managers have an integral role in providing quality data to scientists, but the analysis itself is the purview of project leaders and the network biometrician. In depth information regarding data analysis can be found in the Arctic Network monitoring plan as well as the monitoring protocols and standard operating procedures for each vital sign (NPS 2007a,d).

Analysis

Each Arctic Network vital sign is required to have a series of standard operating procedures (Oakley et al., 2003). A subset of these procedures defines data handling, analysis and reporting including sections on generating data summaries and statistical analyses, recommended reporting schedules, report formats with examples of summary tables and figures and methods for long-term trend analysis. The data manager will work closely with the network biometrician in the development of these standard operating procedures. Data management staff will also aid in adapting database objects, fields, and values to support the formats and functions necessary for analysis using statistical software applications.

Reporting

Reporting is critical to the long-term success of the Arctic Network. Results must be credible and delivered in a timely fashion to the appropriate audiences in a manner that is understandable to them (NPS 2006a, NPS 2007a). There are multiple audiences for monitoring data produced by the Arctic Network program, and each requires information formatted and presented in specific ways. The main audience for monitoring data is the resource managers of each network park and other managers in the National Park Service system, who will use the information to assist with their management decisions.

Some basic summary and reporting functions required by a vital sign monitoring protocol can be developed within 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 to prepare datasets 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. Arctic Network will enhance its web site over time to deliver reports and provide supplemental background data and information.

Long-term Trends and Analyses

Most long-term data analysis will involve statistical software applications. 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.

Geographic Information System functions can also contribute to understanding long term status and trends of vital signs and ecosystems. Methods may be developed to visualize 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.

Funding and Timeliness

The Arctic Network strategy towards data analysis and reporting rests upon providing sufficient funding for these activities so that they occur promptly—that is, to report on the previous phenological year (Oct-Sept) by the following March. The Arctic Network will also focus on producing an annual integrated “State of the Parks” report that effectively communicates the changes and trends observed in each Vital Sign to our primary audience—the natural resource managers of each park.

Chapter 9. Data Dissemination

Too often we forget that genius. . . depends upon the data within its reach, that Archimedes could not have devised Edison's inventions.

--Ernest Dimnet

Introduction

Data management within the Arctic Network Inventory and Monitoring Program aims to ensure that:

- Data are easily discoverable and obtainable
- Only data that have been subjected to full quality control are released
- Distributed data are accompanied by complete metadata which clearly establishes the data as a product of the NPS Inventory and Monitoring Program
- Sensitive data are identified and protected from unauthorized access and inappropriate use
- A complete record of data distribution/dissemination is maintained

Data Distribution Mechanisms

The primary distribution of the network's inventory and monitoring data will occur on the internet. This approach makes data and information available to a broad community of users and does not require a local system to receive and process multiple data requests. The following web-based applications and repositories have been developed to store park natural resource information (NPS 2007i,o):

- **NatureBib**—a master web database housing natural resource bibliographic citations for Inventory and Monitoring Program parks
- **NPSpecies**—a master web-based database to store, manage and disseminate scientific information on the biodiversity of all organisms in all National Park units
- **Biodiversity Data Store**—a digital archive of document, GIS data set and non-GIS data set files that document the presence/absence, distribution and/or abundance of any taxa in National Park Service units
- **NPS Data Store**—online repository for metadata and associated data products.
- **Arctic Network Website**—provides detailed information about the Inventory and Monitoring network and its activities. Metadata, datasets, and related products such as summary reports for all network-funded inventory and monitoring products will be available via the web site. Park-funded and non-NPS-sourced materials related to the vital signs monitoring program will be linked via the network website where possible.

Storing network metadata and datasets in the data stores listed above ensures dissemination to a wide audience.

Data Ownership

National Park Service Policy on Data Ownership

The National Park Service defines conditions for the ownership and sharing of collections, data, and results based on research funded by the United States government. All cooperative and interagency agreements, as well as contracts, should include clear provisions for data ownership and sharing as defined by the National Park Service: All data and materials collected or generated using National Park Service personnel and funds become the property of the National Park Service.

Any important findings from research and educational activities should be promptly submitted for publication. Authorship must accurately reflect the contributions of those involved.

Investigators must share collections, data, results, and supporting materials with other researchers whenever possible. In exceptional cases, where collections or data are sensitive or fragile, access may be limited.

The Office of Management and Budget (OMB) ensures that grants and cooperative agreements are managed properly. Federal funding must be disbursed in accordance with applicable laws and regulations. OMB circulars establish some degree of standardization government-wide to achieve consistency and uniformity in the development and administration of grants and cooperative agreements. Specifically, OMB Circular A-110 establishes property standards within cooperative agreements with higher institutions and non-profit organizations. Section 36 of Circular A-110, "Intangible Property" describes the following administrative requirements pertinent to data and ownership:

(a) The recipient (higher institution or non-profit organization receiving federal monies for natural resource inventory and/or monitoring) may copyright any work that is subject to copyright and was developed, or for which ownership was purchased, under an award. The Federal awarding agency(ies) (in this case the National Park Service) reserve a royalty-free, nonexclusive and irrevocable right to reproduce, publish, or otherwise use the work for Federal purposes, and to authorize others to do so.

Section 36 also states:

(c) The Federal Government has the right to:

- (1) obtain, reproduce, publish or otherwise use the data first produced under an award
- (2) authorize others to receive, reproduce, publish, or otherwise use such data for Federal purposes

(d) (1) In addition, in response to a Freedom of Information Act (FOIA) request for research data relating to published research findings produced under an award that were used by the Federal Government in developing an agency action that has the force and effect of law, the Federal awarding agency shall request, and the recipient shall provide, within a reasonable time, the research data so that they can be made available to the public through the procedures established under the FOIA (5 U.S.C. 552(a)(4)(A)).

(2) The following definitions apply for purposes of paragraph (d) of this section:

- (i) Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This "recorded" material excludes physical objects (e.g., laboratory samples)...
- (ii) Published is defined as either when:
 - (A) Research findings are published in a peer-reviewed scientific or technical journal; or
 - (B) A Federal agency publicly and officially cites the research findings in support of an agency action that has the force and effect of law.
- (iii) Used by the Federal Government in developing an agency action that has the force and effect of law is defined as when an agency publicly and officially cites the research findings in support of an agency action that has the force and effect of law.

Establishing Data Ownership: Cooperative or Interagency Agreements

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

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

Cooperative or interagency agreements or contracts must include a clearly defined list of deliverables and products. Details on formatting and media types that will be required for final submission must be included. Typical products include, but are not limited to, field notebooks, photographs (hardcopy and digital), specimens, raw data, and reports.

All reports generated by or for the Arctic Network must follow current format guidance set by the Alaska Region Inventory and Monitoring Program. Other products resulting from cooperative work must follow Alaska Region Inventory and Monitoring formatting specifications.

The following statement must be included in the Reports and Deliverables section of all Arctic Network agreements and contracts:

“All reports and deliverables must follow the most recent versions of the Inventory and Monitoring product specifications.”

Cooperative agreements and contracts should also provide a schedule of deliverables that includes sufficient time for NPS review of draft deliverables before scheduled final submissions.

Data Classification: protected vs. public

All data and associated information from Inventory and Monitoring activities must be assessed to determine their sensitivity. This includes, but is not limited to, reports, metadata, raw and manipulated spatial and non-spatial data, maps, etc. Network staff must carefully identify and

manage any information that is considered sensitive. The Network must clearly identify and define those data needing access restrictions and those to make public.

The Freedom of Information Act, 5 U.S.C. § 552, referred to as FOIA, stipulates that the United States Government, including the National Park Service, must provide access to data and information of interest to the public. FOIA, as amended in 1996 to provide guidance for electronic information distribution, applies to records that are owned or controlled by a federal agency, 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 exemptions. Under the terms of FOIA, agencies must make non-protected records available for inspection and copying in public reading rooms and/or the Internet. Other records however, are provided in response to specific requests through a specified process. The Department of the Interior's revised FOIA regulations and the Department's Freedom of Information Act Handbook can be accessed at www.doi.gov/foia for further information.

In some cases, public access to data can be restricted. Under one Executive Order, Director's Order #66 and four resource confidentiality laws, the National Parks Omnibus Management Act (16 U.S.C. 5937), the National Historic Preservation Act (16 U.S.C. 470w-3), the Federal Cave Resources Protection Act (16 U.S.C. 4304) and the Archaeological Resources Protection Act (16 U.S.C. 470hh), the National Park Service is directed to protect information about the nature and location of sensitive park resources. Through these regulations, information that could result in harm to natural resources can be classified as 'protected' or 'sensitive' and withheld from public release (National Parks Omnibus Management Act).

The following guidance for determining whether information should be protected is suggested in the draft Director's Order #66 (the final guidance may be contained in the Reference Manual 66):

- 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 where relatively out-dated, 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?

Natural Resource information that is sensitive or protected requires the:

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

Classification of sensitive Inventory and Monitoring data will be the responsibility of the Arctic Network staff, the park superintendents, and investigators working on individual projects. Network staff will classify sensitive data on a case-by-case, project-by-project basis. They will work closely with investigators for each project to ensure that potentially sensitive park resources are identified, and that information about these resources is tracked throughout the project.

The Network staff is also responsible for identifying all potentially sensitive resources to project leaders working on each project. The project leaders, whether network staff or partners, will develop procedures to flag all potentially sensitive resources in any products that come from the project, including documents, maps, databases, and metadata. When submitting any products or results, investigators should specifically identify all records and other references to potentially sensitive resources. Note that partners should not release any information in a public forum before consulting with network staff to ensure that the information is not classified as sensitive or protected.

For example, information may be withheld regarding the nature and/or specific locations of the following resources recognized as ‘sensitive’ by the National Park Service. According to National Parks Omnibus Management Act, if the NPS determines that disclosure of information would be harmful, information may be withheld concerning the nature and specific location of:

- Endangered, threatened, rare or commercially valuable National Park System Resources (species and habitats)
- Mineral or paleontological objects
- Objects of cultural patrimony
- Significant caves

Note that information already in the public domain can, in general, be released to the public domain. For example, the media has reported in detail the return of condors to the Grand Canyon. If an individual requests site-specific information about where the condors have been seen, this information, in general, can be released. However, the locations of specific nest sites cannot be released.

Access Restrictions on Sensitive Data

Arctic Network staff are 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 the National Park Service
- The Network Coordinator, NPS project liaison, or Data Manager identifies 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 into any Arctic Network repository, Arctic 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 Inventory and Monitoring Program, are maintained on both secure and public servers and all records marked 'sensitive' during uploading will only become available on the secure servers.

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.

Dissemination Mechanisms

Data and information will be made available to two primary audiences: public and NPS employees. Table 7 lists the data and product repositories the Arctic Network will use. Data and related products will also be available by request via contacts on the Arctic Network website.

Item	Repository
Reports (public) • Digital	NPS Focus, Arctic Network data server (Fairbanks)
• Hard copy	Alaska Resources Library and Information Services (ARLIS), park libraries
• Bibliography	NatureBib
Arctic Network-generated digital datasets and data products (public, non-sensitive) • Certified data and data products (including photographs) • Metadata	NPS Data Store, Arctic Network data server, NPSpecies, EPA STORET.

Arctic Network-generated digital project data and info (NPS staff, sensitive) • Raw, validated and analyzed data • Metadata • Submitted reports • Digital photographs • Digital presentations	Arctic Network data server, backup server (Anchorage).
Project product materials • Vouchers • Specimen	University of Alaska Museum, Arctic Network or park office (according to project protocol)
Non-product project items (hard copy)	Arctic Network office

Table 6. Repositories for Arctic Network Program data and information.

Arctic Network Primary Server

The Arctic Network will maintain and use file, database and web servers housing electronic data and information managed by the program. These tools will allow staff to browse, evaluate, export, analyze and integrate vital sign monitoring data and information for research, management and reporting purposes. Databases for each vital sign will be designed in MS Access and with back-end tables stored in an enterprise database at the national office in Fort Collins, Colorado. Spatial data will be stored in a shapefile or geodatabase format. The data dissemination system will be developed to include a web-interface for data browsing, query and download.

Non-monitoring data

The distinction between “monitoring” and “non-monitoring” data may be construed as the difference between “active” and “finalized” data. The brunt of Arctic Network data management planning focuses on monitoring data, however, in general the Arctic Network will make every effort to upgrade non-monitoring datasets, such as legacy and inventory data, such that they may be served via the client/server database on the Arctic Network primary data server. Data formatted in this manner will be more easily browsed, manipulated and discovered by end-users.

Alaska Resources Library and Information Services (ARLIS)

Alaska Resources Library and Information Services (ARLIS) provides universal access to natural and cultural resources information and serves as a central library for local, state and federal agencies in Alaska. The park service in Alaska has historically used ARLIS for a variety of information needs and the Arctic Network will build on this relationship by ensuring that all appropriate program output are catalogued with ARLIS for broad distribution.

NPS Focus

Arctic Network data and products residing in NatureBib and NPS Data Store will automatically be searchable via NPS Focus (<http://npsfocus.nps.gov>); a digital library. As NPS Focus develops, additional databases and repositories utilized by the network are expected to be searchable through this portal.

Water Quality Data

Water quality data collected to meet regulatory requirements is managed according to guidelines from the NPS Water Resources Division. This includes using the NPSTORET desktop database application at the parks to help manage data entry, documentation, and transfer. The Network oversees the use of NPSTORET according to the Network's integrated water quality monitoring protocol and ensures the content is transferred at least annually to NPS Water Resource Division for upload to the STORET database (Figure 12).

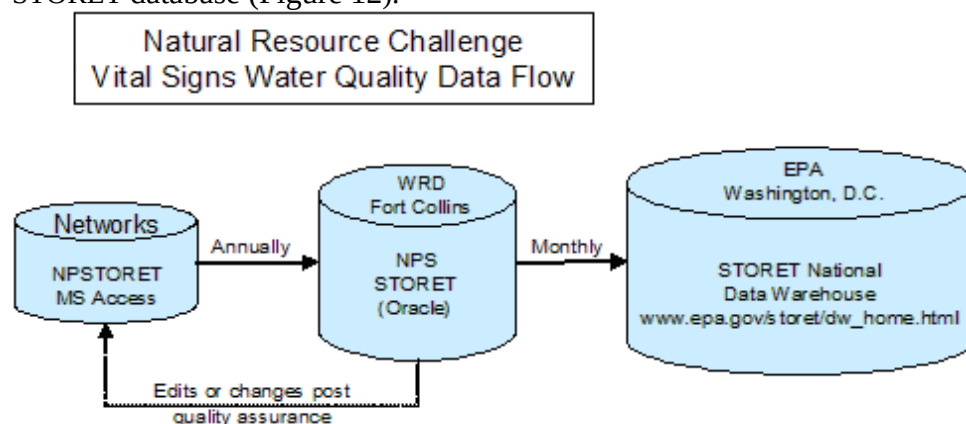


Figure 12. Water quality data flow.

Data Availability

In general, data will be available upon completion of analysis and reporting. Data for which analysis and reporting has not been completed but are otherwise certified (verified and validated) will be released no later than one year after certification.

Data release or integration into dissemination mechanisms depends on both the project protocol and project leader. No data may be released without proper certification.

Arctic Network staff will notify project leaders prior to making datasets available to the public. This will allow each project leader the opportunity to specify in writing any access restrictions that should apply. Any such specifications must indicate how long the data will remain restricted.

Data Release Policy

The Arctic Network will maintain the following policy regarding the release of data and related products:

1. Only fully documented, certified, non-sensitive data and data products may be uploaded to public distribution repositories or otherwise released to the public or other non-NPS recipient.
2. Any released data or data products must be accompanied by full metadata and any additional supporting documentation as determined by the project leader.
3. Arctic Network monitoring program data will be made available according to regular reporting and data integration schedules.
4. Arctic Network non-monitoring data will be released upon completion of data certification (verification and validation) and on condition of project leader approval. However, data for which analysis and reporting has not been completed but are otherwise certified (verified and validated) will be released no later than one year after certification.
5. Distribution instructions for each dataset will be included in respective metadata.
6. Distribution logs specifying recipient name and contact information, intended use of data, export file format, delivery date and method, data content description noting range (by date and geography) of data delivered and description of distributed items (noting that metadata and possibly other supporting documentation were distributed with data), will be kept for each distinct dataset or product managed by the Arctic Network.
7. At the discretion of the project leader, any reports utilizing the data may also accompany distributed data .

Feedback Mechanisms

Comments and questions concerning Inventory and Monitoring project data are welcome at any time and may be submitted via e-mail or telephone to the primary contact of a project or to the Network Coordinator. The Arctic Network website will also provide an opportunity for NPS staff, cooperators and the public to provide feedback on data and information distributed as part of Arctic Network operations. A “comments and questions” link will be provided on the main page of the site for general questions and comments about the Network’s program and projects. A more specific “data error feedback” link will direct comments to Arctic Network staff pertaining to errors found in website accessible data. Data feedback and response activity will be reported to the Board of Directors and to the Technical Committee on a yearly basis.

Data Error Feedback Response Procedures

The Arctic Network will use the following procedure to respond to data feedback:

- Arctic Network staff (or automated process) will immediately acknowledge receipt of any feedback to the sender.
- Data error reports will be recorded in an error log associated with each database .
- Data errors will be investigated within one week of submission. If appropriate and feasible, confirmed errors are immediately corrected in all data repositories. If immediate correction is not possible, the data in question are immediately removed from all dissemination mechanisms.
- A report detailing the reported error and response will be prepared by the network data manager and submitted to the project leader, network coordinator and the error reporter. If a data error is not immediately correctable, the error response report will include recommendations for correcting the error.

- An appropriate level of communication to stake holders will be maintained during all data error investigations.

Chapter 10. Data Maintenance, Storage and Archiving

I have not lost my mind - it's backed up on disk somewhere.
--Unknown

Data is a valuable commodity for a research organization. Data is also highly complex, taking a number of different forms, both dynamic and static, that must be dealt with in unique ways. The maintenance, storage and archiving of data must be carefully thought out and executed if its value is to be preserved. Arctic Network data maintenance, storage and archiving procedures aim to ensure that data and related documents and materials (digital and physical) 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
- Archived in a manner that expedites recovery if needed

Data Maintenance

Data maintenance procedures in the Arctic Network apply primarily to network-generated data. We anticipate most maintenance activity will involve active monitoring datasets. Finalized Arctic Network project data will also be maintained along with active data in a common relational database system. Related metadata as well as database application interface tags will discern between active and finalized data.

Spatial Data

Most spatial data (monitoring data in particular) will reside in a management system comprised of file and web servers and an enterprise database. The benefits to this arrangement are numerous and include:

- Secure, high volume storage space,
- Concurrency
- Multiple user environment,
- Availability of centralized lookup tables and communal information
- Universal access for NPS personnel.

Ongoing spatial datasets will reside in our enterprise database with snapshots archived with metadata on the NPS Data Store and in the Permanent Data Stack at the Alaska Regional Office. Static datasets will be processed and archived with the data store. Dataset metadata and error logs will record data maintenance activities. Database maintenance logs will record any changes to project databases.

A data catalog will be maintained on the Arctic Network website and reflect changes or updates to datasets. These changes or update records will typically reflect data collection, entry, and certification dates but may also record actions to correct errors in the dataset or to migrate data to new digital formats. National and regional repositories for Arctic Network data and information will be updated to reflect current stores on the Arctic Network server.

Files

Files will be housed on the primary Arctic Network shared directory (O:/ drive) in Fairbanks. We will use a standard directory structure (NPS 2007c). Non-sensitive files will be replicated as appropriate to the Arctic Network website to provide access to outside collaborators and to publish the details of our projects.

File Backup Plan

The primary Arctic Network server will be housed in the Fairbanks Administrative Center as part of the local area network. These data will be backed up off-site to either or both the regional office in Anchorage and/or the Fairbanks Alaska Public Lands Information Center at least weekly. Both of these backup sources will be read-only and accessible for data restoration purposes only. Project leaders should additionally employ their own data backups as project work progresses.

Hardcopy Data and Information - Maintenance

Hardcopy materials, including paper data sheets, photo negatives, field notebooks and other items will be scanned where appropriate and stored in on the Arctic Network shared directory. Originals will be tagged with the project ID and filed at the Fairbanks Administrative Center. Copies will be made for off-site storage as these materials are received. Catalogues of archived material will be searchable on the Arctic Network website and maintained to reflect basic accessioning information for each item. Collections will be maintained according to NPS archiving protocols (NPS 2000).

Finalized Digital Datasets

Non-spatial datasets that are considered complete and inactive will be archived with the NPS Data Store in ASCII (text) format unless a proprietary format must be used. “Native format” is defined as the format in which data for a given project are generated. The “native” format may be different for raw, checked and analyzed versions of the data. Note that “specimen” or “sample” data, i.e. physical objects taken from the field do not fall under this definition of “native” format. Data are considered complete and inactive when accompanying metadata, as generated by the project leader or other authorized personnel, list the project status as “complete”. Associated digital content such as submitted project reports, photographs, presentations, etc., will be archived along with project data.

Physical Material Archival

Project Products: Table 6 specifies locations for physical items considered project products such as reports. Additional hard copy project product items such as maps, posters, slides, photographs (including aerial), tape recordings, etc., will be stored in the Arctic Network main office. These items (including project reports) will be archived according to NPS standards and follow the procedures outlined below under “Non-Product Items”.

Specimen and Samples: Field samples and specimens will be archived according to NPS standards following procedures provided in project-specific protocols. Unless otherwise specified in the project protocol, specimen and samples will be provided to the network park in

which they were collected for curation according to park policy. The Arctic Network will provide park curators with associated data and material required for curation.

Project Related Items

Items such as field notebooks and other materials that may be considered raw data or important project information such as correspondence, permits, agreements, etc., will be catalogued and filed in the Network's main office. Such items will be archived along with related project material noted above under "Project Products".

Project materials will be consolidated and "packaged" for archival on an annual basis. The project leader should package the annual project information as follows:

- A coversheet or memo listing all materials included in the package (see Appendix I for project checklist form).
- All materials are clearly labeled with:
 - Park code
 - Date, or range of dates
 - Field notes on acid free paper and stored in a 3-ring binder or book box
 - Other paper materials such as reports and data printouts also on acid free paper and placed individually in a acid free, labeled folder
 - Specimens, such as plants or fish, are properly labeled
 - For specimens not residing in archives, loan paper work is complete and a copy of the form is stored with the project package.
- CD or DVD of all electronic materials. Case should be labeled with:
 - o Park Code
 - o Inventory and Monitoring Network Code
 - o Date created
 - o Range of dates for information
 - o Accession Number
 - o List contents

The archive specialist will catalog this information into NPS Rediscovery and arrange for proper storage of the materials.

Photographs

The Arctic Network is developing a comprehensive photographic database designed to track and manage images of park resources. Photographs from each Arctic Network project will be entered into this database where attributes such as electronic file name, keywords, project, description, photographer, date and location will be catalogued. Digital photo management will generally follow guidelines established in a draft management strategy prepared by the Southwest Alaska Inventory and Monitoring Network.

Chapter 11. Implementation

The Arctic Network will follow the guidelines in Chapter 13 of the National Data Management Plan (National Park Service, 2008) for implementing this data management plan.

References

- DOI. 1993. Departmental manual museum property handbook (411 DM, Volume 1). U.S. Department of Interior, Unpubl. Manual. Washington D.C. .
- 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.
- NPS. 1994. "Executive Order 12906: Coordinating geographic data acquisition and access: The National Spatial Data Infrastructure. ." Retrieved 12-04, 2007, from <http://govinfo.library.unt.edu/npr/library/direct/orders/20fa.html>.
- NPS. 1999. "OMB Circular A-110: Uniform administrative requirements for grants and agreements with institutions of higher education, hospitals, and other non-profit organizations." Retrieved 12/5/2007, 2007, from <http://www.whitehouse.gov/omb/circulars/index.html>.
- NPS. 2000. DIRECTOR'S ORDER #24:NPS museum collections management, National Park Service.
- NPS. 2002. DIRECTOR'S ORDER #11B: Ensuring quality of information disseminated by the national park service. National Park Service.
- NPS. 2002. DIRECTOR'S ORDER #66: FOIA and protected resource information, National Park Service.
- NPS. 2004. Recommended naming standards for inventory and monitoring databases, National Park Service Washington Support Office, Ft. Collins, Colorado.
- NPS. 2005a. NR-GIS Data Store: Metadata parsing guidance. National Park Service. Ft. Collins, Colorado.
- NPS. 2005b. NR-GIS Data Store: Creating metadata. National Park Service. Ft. Collins, Colorado.
- NPS. 2005c. NR-GIS Data Store: Creating non-spatial metadata. National Park Service. Ft. Collins, Colorado.
- NPS. 2005d. NR-GIS Data Store: Creating simple geospatial metadata. National Park Service. Ft. Collins, Colorado.
- NPS. 2005e. NR-GIS Data Store: Metadata and data store instructions. National Park Service. Ft. Collins, Colorado.

- NPS. 2005f. NR-GIS Data Store: Metadata authoring guidance. National Park Service. Ft. Collins, Colorado.
- NPS. 2006a. Instruction to authors—Natural Resource Report and Natural Resource Technical Report. Natural Resource Report NPS/NRPC/NRR—2006/001, National Park Service Ft. Collins, Colorado.
- NPS. 2006b. Natural Resources Database Template specifications, National Park Service Ft. Collins, Colorado.
- NPS. 2007a. Checklist for initiating a new arctic network project: Guidelines and tasks checklist for integrating new projects with the arctic network data catalog. Arctic Network Data Management Standard Operating Procedure NPS/ARCNDMSOP-2007-01. National Park Service, Fairbanks, Alaska
- NPS. 2007b. Specifications for study plans submitted to the inventory and monitoring program, alaska region: Guidelines and template for writing effective inventory and monitoring program study plans. Arctic Network Data Management Standard Operating Procedure NPS/ARCNDMSOP-2007-02. National Park Service, Fairbanks, Alaska
- NPS. 2007c. Project file management: Guidelines for keeping arctic network project files organized, secure and accessible. Arctic Network Data Management Standard Operating Procedure NPS/ARCNDMSOP-2007-03. National Park Service, Fairbanks, Alaska
- NPS. 2007d. Project deliverables specifications: Guidelines for delivering high quality data products. Arctic Network Data Management Standard Operating Procedure NPS/ARCNDMSOP-2007-04. National Park Service, Fairbanks, Alaska
- NPS. 2007e. Data quality assurance and control: Quality assurance/quality control (qa/qc) guidance. Arctic Network Data Management Standard Operating Procedure NPS/ARCNDMSOP-2007-05. National Park Service, Fairbanks, Alaska
- NPS. 2007f. Database planning toolkit: Task checklist and guidance for inventory and monitoring data application development. Arctic Network Data Management Standard Operating Procedure NPS/ARCNDMSOP-2007-07. National Park Service, Fairbanks, Alaska
- NPS. 2007h. Certifying data products: Dataset certification benchmarks and certification form. Arctic Network Data Management Standard Operating Procedure NPS/ARCNDMSOP-2007-11. National Park Service, Fairbanks, Alaska
- NPS. 2007i. Overview of national inventory and monitoring program databases: A quick introduction to national level databases and data management resources. Arctic Network Data Management Standard Operating Procedure NPS/ARCNDMSOP-2007-14. National Park Service, Fairbanks, Alaska

- NPS. 2007k. Monitoring project roles and responsibilities: Coordinating personnel and tasks for efficient project collaboration. Arctic Network Data Management Standard Operating Procedure NPS/ARCN/DMSOP-2007-16. National Park Service, Fairbanks, Alaska
- NPS. 2007l. Approving information for distribution: Process for determining sensitivity and classifying information for distribution. Arctic Network Data Management Standard Operating Procedure NPS/ARCN/DMSOP-2007-17. National Park Service, Fairbanks, Alaska
- NPS. 2007m. Compliance review process: Shepherding Inventory and Monitoring projects through the compliance process in Gates of the Arctic National Park and Preserve. Arctic Network Data Management Standard Operating Procedure NPS/ARCN/DMSOP-2007-18. National Park Service, Fairbanks, Alaska
- NPS. 2007o. Data stores for Inventory and Monitoring program informational products: overview of nps information warehouses. Arctic Network Data Management Standard Operating Procedure NPS/ARCN/DMSOP-2007-20. National Park Service, Fairbanks, Alaska
- NPS. 2007a. Specifications for geospatial and other data deliverables of gis and resource mapping, inventories, and studies., National Park Service. Fort Collins, CO. .
- National Park Service. 2008. Data management guidelines for inventory and monitoring networks. Natural Resource Report NPS/NRPC/NRR—2008/035. National Park Service, Fort Collins, Colorado.
- Oakley, K. L., Lisa P. Thomas and S. G. Fancy 2003. "Guidelines for long-term monitoring protocols." Wildlife Society Bulletin 31(4): 1000-1003.
- Selleck, J., K. KellerLynn 2006. Editorial style guide for park science and natural resource year in review, National Park Service Ft. Collins, Colorado.
- Southwould, A. 2002a. Database specifications for inventory and monitoring studies. , National Park Service, Anchorage, Alaska.
- Southwould, A. 2002b. Recommended database strategies including I&M database template, National Park Service, Anchorage, Alaska.

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

NPS D-169, August, 2008

National Park Service
U.S. Department of the Interior



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

www.nature.nps.gov

EXPERIENCE YOUR AMERICA™