National Park Service
U.S. Department of the Interior

Natural Resource Program Center

Sierra Nevada Network
Data Management Plan

Natural Resource Report NPS/NRPC/NRR—2008/070

ON THE COVER
Upper Rock Creek, Sequoia National Park
NPS Photo

Sierra Nevada Network
Data Management Plan

Natural Resource Report NPS/NRPC/NRR—2008/070

National Park Service
Natural Resource Program Center
1201 Oak Ridge Drive, Suite 150
Fort Collins, CO 80525

November 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; "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, National Park Service. 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 Resources Publications Management Web site (http://www.nature.nps.gov/publications/NRPM) on the Internet, or by sending a request to the address on the back cover.

Please cite this publication as:

Cook, R. R., and P. Lineback. 2008. Sierra Nevada Network data management plan. Natural Resource Report NPS/NRPC/NRR—2008/070. National Park Service, Fort Collins, Colorado

NPS D-45, November 2008

Contents

List of Figures ix

List of Tables xi

Acknowledgements xxiii

Change History xvii

Acronyms used in this document xix

Executive Summary xxi

Chapter 1. Introduction 1
1.1. The Sierra Nevada Network 1
1.2. Purpose and Scope 1
1.3. Goals and Priorities 3
1.4. Organization 3
1.5. Key Terms and Definitions 3
1.6. Making it Work 5
1.7. References 6
1.8. Appendixes 6
1.9. Credits 6

Chapter 2. Infrastructure and Systems Architecture 7
2.1. Objectives 7
2.2. Laws and Policies 7
2.3. General Standards and Guidelines 8
2.3.1. Network Computer Resources 8
2.3.2. Network Security 9
2.3.3. National Applications and Tools 9
2.3.4. SIEN I&M Web Site 10
2.3.5. Collaboration Technologies 10
2.3.5.1. Web Meetings 11
2.3.5.2. Content Management 11
2.3.5.3. Personal Communications 12
2.3.6. Geographic Information Systems 12
2.3.6.1. Devils Postpile National Monument 12
2.3.6.2. Sequoia and Kings Canyon National Parks 12
2.3.6.3. Yosemite National Park 13
2.3.6.4. Recommendations 13
2.3.7. Database Applications Development 14
2.3.9. Digital Data Formats 14
2.3.9.1. Databases 14
2.3.9.2. Spatial data 15

2.4. References 15
2.5. Appendixes 16
2.6. Credits 16

Chapter 3. Project Development and Data Management Workflow 17
3.1. Five Stages of Project Development 17
3.2. The Data Life Cycle 20
3.3. References 23
3.4. Appendixes 23
3.5. Credits 23

Chapter 4. Data Management Roles and Responsibilities 25
4.1. General Standards and Guidelines 25
4.2. Collaboration 27
4.3. References 28
4.4. Appendixes 28
4.5. Credits 28

Chapter 5. Databases 29
5.1. Objectives 29
5.2. Laws and Policies 29
5.3. General Standards and Guidelines 29
5.3.1. The Natural Resource Database Template 29
5.3.2. Database Development 29
5.3.2.1. Conceptual Data Models 29
5.3.2.2. Logical Data Models 30
5.3.2.3. Physical Data Models 33
5.4. Quality Assurance/Quality Control 34
5.5. Database Documentation 35
5.5.1. Basic Requirements 35
5.5.2 Revisions 36
5.6. References 36
5.7. Appendixes 36
5.8. Credits 36

Chapter 6. Acquisition, Processing and Reporting 37
6.1. Objectives 37
6.2. Laws and Policies 37
6.3. General Standards and Guidelines 37
6.3.1. Programmatic Data 37
6.3.1.1. Data Collection 38
6.3.1.2. Data Processing 41
6.3.1.2.1. Spatial data 41
6.3.1.2.2. Photographs 41
6.3.1.2.3. Remote Sensing Data 41
6.3.1.2.4. Vouchers 42

6.3.2. Non-programmatic Data 42
6.3.2.1. Data Collection 44
6.3.2.2. Data Processing 45
6.3.3. Analysis and Reporting 45
6.3.4. Changes to Procedures 45
6.3.5. Maintaining an Edit Log 46
6.4. References 46
6.5. Appendixes 46
6.6. Credits 46

Chapter 7. Quality Assurance and Quality Control (QA/QC) 47
7.1. Objectives 47
7.2. Laws and Policies 47
7.3. General Standards and Guidelines 48
7.3.1. Project Planning and Data Design 50
7.3.2. Data Collection 50
7.3.3. Data Entry 51
7.3.4. Data Verification 52
7.3.4.1. Methods for Data Verification 52
7.3.5. Data Validation 53
7.3.6. Methods for Data Validation 54
7.3.7. Review, Conformance and Communication 54
7.3.8. Roles and Responsibilities 55
7.4. References 56
7.5. Appendixes 56
7.6. Credits 56

Chapter 8. Dataset Documentation 57
8.1. Objectives 57
8.2. Laws and Policies 57
8.3. General Standards and Guidelines 58
8.3.1. Content Standards 58
8.3.2. Format and Storage Standards 60
8.3.3. Metadata Tools 61
8.3.4. Work Flow Process 62
8.3.5. Maintenance and Access 62
8.4. References 63
8.5. Appendixes 63
8.6. Credits 63

Chapter 9. Data Ownership and Sharing 65
9.1. Objectives 65
9.2. Laws and Policies 65
9.3. General Standards and Guidelines 66
9.3.1. Collaborative Agreements 66
9.3.2. Due Notification 67

9.3.3. Access Restrictions on Sensitive Data 67
9.3.4. Answering Requests under the Freedom of Information Act 68
9.4. References 68
9.5. Appendixes 68
9.6. Credits 69

Chapter 10. Data Dissemination 71
10.1. Objectives 71
10.2. Laws and Policies 71
10.3. General Standards and Guidelines 72
10.3.1. Data and Information Distribution Mechanisms 72
10.3.1.1. The SIEN Website 73
10.3.1.2. National Databases 73
10.3.2. Sensitive Data 75
10.3.4. Feedback Mechanisms 75
10.4. References 75
10.5. Appendixes 76
10.6. Credits 76

Chapter 11. Records Management and Object Curation 77
11.1. Objectives 78
11.2. Laws and Policies 79
11.2.1. Records 79
11.2.2. Collections 80
11.2.3. Local Network Policies 80
11.3. General Standards and Guidelines 80
11.3.1. Records Management 80
11.3.1.1. Electronic File Naming Guidelines and Standards 80
11.3.1.2. Directory Structure 81
11.3.2. Archival, Backup, and Storage 82
11.3.2.1. Project Process and Workflow 82
11.3.2.2. Electronic Files 83
11.3.2.2.1. Individual Employee Responsibility 83
11.3.2.2.2. IT Staff Responsibility 84
11.3.2.2.3. Short-term Projects 85
11.3.2.2.4. Long-term Projects 85
11.3.2.2.5. Spatial Data 86
11.3.2.3. Physical Objects and Records 86
11.3.2.3.1. Documents 86
11.3.2.3.2. Natural History Specimens 87
11.3.2.3.3. Photographs 87
11.3.2.3.4. Role of Curators 88
11.4. References 88
11.5. Appendixes 89
11.6. Credits 90

Chapter 12. Project Tracking and Documentation 91
12.1. Objectives 91
12.2. General Standards and Guidelines 91
12.2.1. Project Organization 91
12.2.2. Project Tracking 91
12.2.3. Project Documentation 92
12.2.3.1. Protocol and SOP Version Changes 92
12.2.3.2. Additional Project Metadata 92
12.3. References 92
12.4. Appendixes 92
12.5. Credits 92

Chapter 13. Implementation 93
13.1. Education and Training 93
13.2. Milestone Goals 93
13.3. Revisions 94
13.4. Credits 94
Glossary 95

Figures

Figure 1. Sierra Nevada Network Parks and Ecoregion. 2

Figure 2. Data management steps during the five stages of project development with data management activities involved in each phase. Core activities are in bold. 18

Figure 3. The data life cycle. 22

Figure 4. Schematic of a conceptual data model for monitoring the status of mountain yellow-legged frog populations following removal of non-native fish in high mountain lakes in Sequoia and Kings Canyon National Parks. 31

Figure 5. Schematic of a logical data model for monitoring the status of mountain yellow-legged frog populations following removal of non-native fish in high mountain lakes in Sequoia and Kings Canyon National Parks. 32

Figure 6. Schematic of the physical database used for monitoring the status of mountain yellow-legged frog populations following removal of non-native fish in high mountain lakes in Sequoia and Kings Canyon National Parks. Note: this diagram does not include data type and length information, or domain specifications, which should be declared as part of a physical database model. 34

Figure 7. General course of data and associated Quality Assurance/Quality Control procedures. Quality Control with regards to data analysis is specific to each project and addressed in appropriate standard operating procedures. 49

Figure 8. Some common information management elements influencing the amount of QA/QC needed. 50

Figure 9. The NPS integrated metadata management system. 62

Figure 10. Flow diagram for water quality data from I&M Networks to the National Data Warehouse. 74

Figure 11. General workflow for storage and archival of electronic (in blue) and hardcopy (in green) scientific research products. 84

Tables

Table 1. Categories of data stewardship involving Sierra Nevada Network personnel..... 25

Table 2. Summary of Roles and Responsibilities for a Network Inventory and Monitoring Program..... 26

Table 3. Satellite resolution, swath width, area, cost, and history of various remote-sensing imagery sources (sorted by cost)..... 40

Table 4. Summary of possible data sources for different types of information and repositories where they are maintained. 44

Table 5. Repositories of SIEN public and sensitive data and information. 73

Appendixes available on the SIEN Intranet Site http://www1.nrintra.nps.gov/im/units/sien/)

1A) 1998 National Parks Omnibus Act

2A) Computer Network Framework for SIEN Inventory and Monitoring

2B) National Repositories for Digital Data and Information

2C) Collaboration Technologies – Evaluated Applications

2D) Sequoia and Kings Canyon GIS History and Current Issues

2E) Yosemite GIS Development and History Before 1992

2F) Director's Order #70: Internet and Intranet Publishing

2H) SIEN Directory Structure

3A) Project Data Certification Form

4A) Sierra Nevada Network Inventory and Monitoring Program, Data Management Responsibilities

4B) Sierra Nevada Network Inventory and Monitoring Program, Coordination of Data Management Roles and Responsibilities

5A) Sierra Nevada Network Database Documentation Template

6A) Sierra Nevada Network I&M Program Checklist of Project Specifications

6B) Field Data Collection with Global Positioning Systems

6C) Spatial Data Standards

6D) Sierra Nevada Digital Photo Management Strategy

6E) Sierra Nevada Network Inventory and Monitoring Program Voucher Collection and Documentation Standards

6F) Data Dictionary and Formatting Specifications for ANCS+ Voucher Records

6G) Field Definitions for NPSpecies Voucher and Observation Records

6H) Sierra Nevada Network Guidance for Data Mining and Discovery

Appendixes (continued)

7A) Director's Order #11B: Ensuring Quality of Information Disseminated by the National Park Service

8A) Executive Order 12906: Coordinating Geographic Data Acquisition and Access, the National Spatial Data Infrastructure

8B) NR GIS Data Store: Metadata Authoring Guidance

8C) Migrating Metadata from SMMS to XML

8D) NPS Metadata Tools and Editor: A Comprehensive Editing Tool for National Park Service Metadata

8E) Tools for Creation of Formal Metadata: Metadata in Plain Language

8F) Sierra Nevada Network Guidelines for developing FGDC compliant metadata for both spatial and biological data

9A) The Freedom of Information Act 5 U.S.C. § 552, As Amended By Public Law No. 104-231, 110 Stat. 3048

9B) Sierra Nevada Network FOIA and Sensitive Data Guidelines

9C) Synopsis of Laws and Policies Governing Public Dissemination of Natural Resources Data

10A) NPSpecies Data Standards Manual

10B) NatureBib Data Standards Manual

10C) Transferring Vital Signs Water Quality Data to NPSTORET

11A) Electronic File Naming Guidelines and Standards

11B) Sierra Nevada Network Inventory and Monitoring Program Guidance and Rules for File Directory Structure

11C) Sierra Nevada Network Project Template Directory Structure

11D) Coversheet for Non-Digital Data listing Key Information

11E) National Park Service Standards and Guidelines for Backup, Storage, and Recovery

Appendixes (continued)

11F) Sequoia and Kings Canyon and Devils Postpile Network General Support Services Backup Activity Plan

11G) Yosemite Network General Support Services Backup Activity Plan

11H) Sierra Nevada Network Archive and Storage Locations

12A) NPS Service-wide Project Tracking Systems

Change History

Titles of updated or revised sections of the plan will be followed by brackets containing the date of the latest version of that section and a brief update history. For example, the title of Section 2.2 may be followed by "[06/02/2006; Update History: 1/15/2005, 10/23/2004]". This indicates the current wording of that section was last updated on June 2, 2006 and it was previously revised on January 15, 2005 and October 23, 2004.

The following revisions have occurred:

Original Version #	Date of Revision	Revised By	Changes	Justification	New Version #
1.0					

Acronyms used in this document

ANCS+	Automated National Catalog System
CDM	Conceptual Data Model
CESU	Cooperative Ecosystems Studies Unit
CSDGM	Concept Standards for Digital Geospatial Metadata
DEPO	Devils Postpile National Monument
DMP	Data Management Plan
EPA	U.S. Environmental Protection Agency
ERD	Entity Relationship Diagram (for databases)
ESRI	Environmental Systems Research institute
FGDC	Federal Geographic Data Committee
FOIA	Freedom of Information Act
FTP	File Transfer Protocol
GIS	Geographic Information System
GPS	Geographic Positioning System
HTML	Hypertext Markup Language
I&M	Inventory & Monitoring (Program)
IMS	Internet Mapping Server
IT	Information Technology
ITIS	Integrated Taxonomic Information System
JPEG	Joint Photographic Experts Group
LAN	Local Area Network
LDM	Logical Data Model
LIDAR	Light Detection and Ranging
MOU	Memorandum of Understanding
mp	Metadata Parser
MS	Microsoft – change in text to Microsoft
PWR	Pacific West Region(al Office)
NBII	National Biological Inventory Infrastructure
NOAA	National Oceanic and Atmospheric Administration
NPS	National Park Service
NRDT	Natural Resource Database Template
NRTS	Natural Resource Table Standards
OMB	Office of Management and Budget
PC	Personal Computer
PDA	Portable Digital Assistant
PDF	Adobe Portable Document Format
PDM	Physical Data Model
QA/QC	Quality Assurance/Quality Control
RDMS/RDBMS	Relational Database Management System
RFID	Radio Frequency Identification
SDE	Spatial Database Engine
SEKI	Sequoia and Kings Canyon National Parks
SIEN	Sierra Nevada Network
SOP	Standard Operating Procedure
SQL	Structured Query Language
TIFF	Tagged Image File Format

Acronyms (continued)

USGS	U.S. Geological Survey
WAN	Wide Area Network
WASO	Washington Support Office
XML	Extensible Markup Language
YOSE	Yosemite National Park

Executive Summary

The central mission of the NPS Inventory and Monitoring Program is to provide timely and usable scientific information about the status and trends of park resources to park managers. To meet this challenge, we need an information management system that can effectively produce, maintain and distribute the products of scientific investigation conducted in our parks.

Good data management is the means by which a thorough understanding of the value of scientific information about our natural resources can become a part of our National Park Service heritage. Data management refers to the framework by which data are acquired, maintained, and made available. Data management is not an end unto itself, but a means of maximizing the quality and utility of our natural resource information. A robust data management system is particularly important for long-term programs where the lifespan of a data set will span the careers of several scientists. Viewed in this way, it becomes obvious that data management is vital to the success of any long-term research program.

The purpose of the Sierra Nevada Data Management Plan is to provide I&M and other park staff with a conceptual framework for a system of data management that will ensure the production and dissemination of timely and usable scientific information about the status and trends of park resources to park managers. Our strategy for achieving this goal can be summarized as follows: Ensure the quality, interpretability, security, longevity and availability of our natural resource data. Our objectives include:

- Confidence in the security and availability of natural resource data and related information
- Easy access to most information, and appropriate safeguards for sensitive information
- Awareness of the intended use and limitations of each data set
- Infrastructure and documentation that encourages data exploration
- Compatibility of data sets for exploration and analysis at larger scales and across disciplines
- Implementation of standards and procedures that facilitate information management, and that reinforce good habits among staff at all levels of project implementation – project leaders, technicians, and volunteer data collectors
- A proper balance between the standards needed to ensure quality and usability, and the flexibility to meet specific needs and encourage innovation
- A team of natural resource professionals who view data not as a commodity but as the lifeblood of our work

This plan describes how our Network will:

- Support Inventory and Monitoring Program objectives
- Acquire and process data
- Assure data quality
- Document, analyze, summarize and disseminate data and information
- Maintain nationally developed data management systems
- Maintain, store and archive natural resources data and information

Acknowledgements

The Sierra Nevada Network Data Management Plan is truly a product of many efforts. The community of data managers in the I&M Program and parks within the 32 I&M Networks fosters a collaborative environment with the free sharing of information. Credit is due to all who have preceded SIEN in the completion of network data management plans, and for making these plans available to the rest of us. Special credit and gratitude are extended to the original Data Management Planning Workgroup, coordinated by John Boetsch, who laid the groundwork for much of this work, and to the "Tier 2" and "Tier 3" data managers who deftly extended these early efforts under the leadership of Margaret Beer and Dorothy Mortenson. Specific credits are included at the end of each chapter. We sincerely apologize for any omissions that may have occurred.

We would also like to thank the members of the SIEN Data Management Steering Team for their enthusiastic support of, and contributions to, this plan. We thank Bob Basham, Anne Birkholz, Ward Eldredge, Paul Gallez, Anne Pfaff, Dan Sohn who contributed written material to the first draft; Leona Svancara who coauthored Chapters 6 and 7; and Ginger Bradshaw for research that contributed to the development of Chapters 8 and 10. We thank also: Margaret Beer and Chris Dietrich who graciously offered enthusiastic support and assistance throughout the development of this plan; Leona Svancara, Andi Heard, Meryl Rose, Patrick Flaherty and Ward Eldredge who provided comments on all or portions of the draft; and to Steve Fancy, Lisa Nelson, Joe Gregson, Alison Loar, Mark Wotawa, and all of the other remarkable data management professionals who have contributed so much to the advancement of data and information management within the National Park Service.

Chapter 1. Introduction

The National Park Service (NPS) Inventory and Monitoring (I&M) Program was the catalyst for developing the Sierra Nevada Network (SIEN) Data Management Plan (DMP). The I&M Program represents a long-term commitment by the NPS to assess and document the status and trends of park resources (Appendix 1A – National Park Omnibus Act 1998). To effectively assess and document park resource status and trends, the Network must develop and implement a plan that outlines a sound long-term road map to guide the development, management, and dissemination of data and information.

1.1. The Sierra Nevada Network

The Sierra Nevada Network includes four National Park units in the southern and central Sierra Nevada range of California, including:

- Devils Postpile National Monument
- Kings Canyon National Park
- Sequoia National Park
- Yosemite National Park
-

Collectively, these parks contain about 1.57 million acres of awe-inspiring grandeur and diversity (Figure 1).

1.2. Purpose and Scope

The targeted audience of this plan is the primarily the SIEN I&M program, but the document has been developed to more broadly meet the needs of other park programs. In addition to I&M, this plan is intended to provide guidance to all SIEN natural resource management divisions and United States Geological Survey (USGS) field stations located in Yosemite National Park (YOSE) and Sequoia and Kings Canyon National Parks (SEKI). These programs work closely with the parks and I&M, and have similar lines of business. We use the term Network throughout this document to refer collectively to these programs. The practices and procedures outlined will also apply to cooperators who have fiscal or formal agreements with Network programs.

This plan offers a guide for Network programs and outlines how we intend to implement and maintain data management systems and best practices that optimize the data and information needs of selected programs. It reflects a commitment to ensure the quality, interpretability, security, longevity, and long-term availability of high-quality natural resource data and information. This is accomplished through standards and guidelines outlined for:

- Proper work flow and management of data through a projects life cycle
- The data management responsibilities of each person involved with a project
- Quality assurance and quality control measures that should become standard practice
- Documentation of projects (project development summaries and SOPs) and data sets (formal metadata)
- Handling and protection of sensitive data and information
- Dissemination to the public of non-sensitive data and information

- Proper management, archiving, and storage of all records and objects associated with projects
- Project organization and tracking

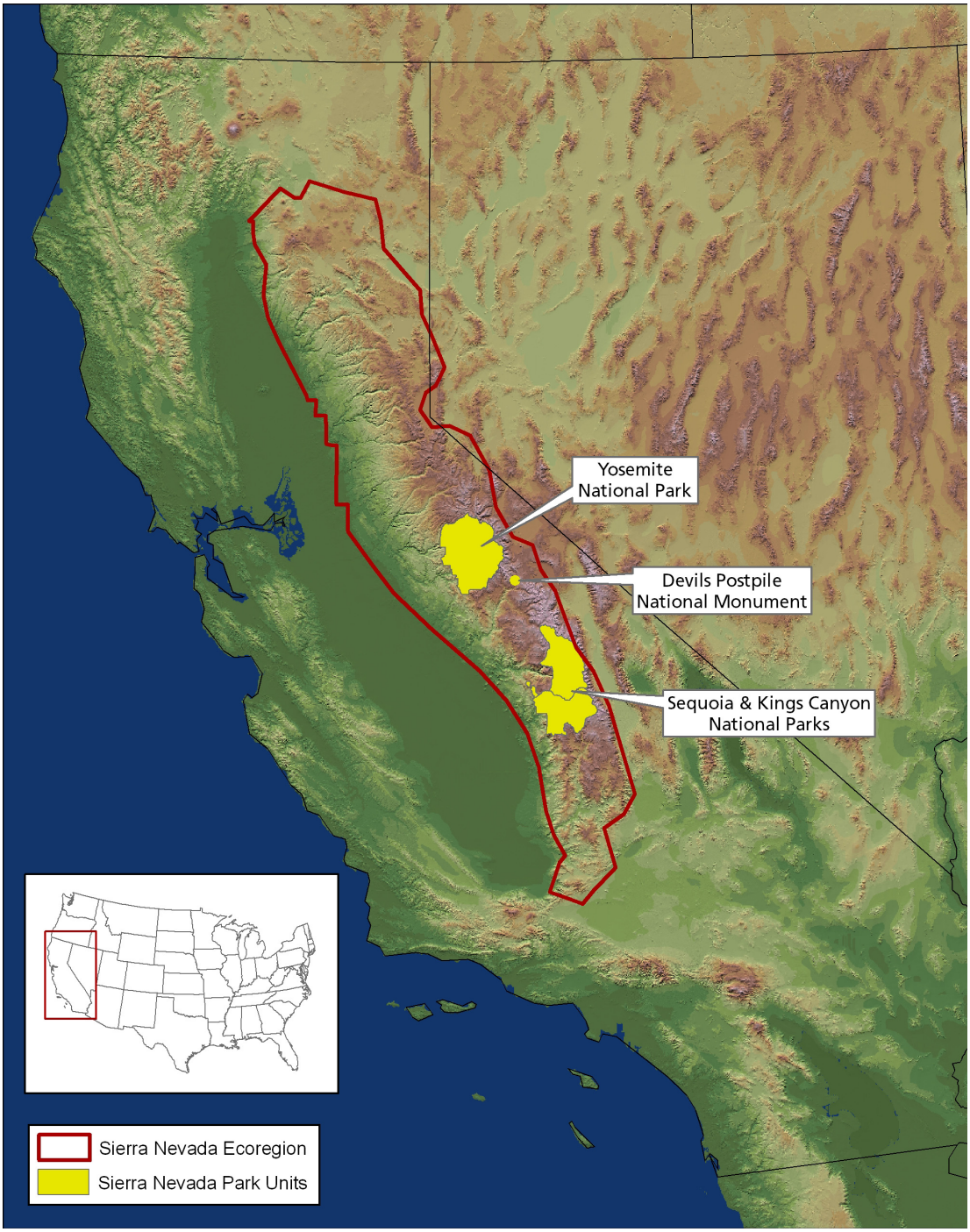

Figure 1. Sierra Nevada Network Parks and Ecoregion.

1.3. Goals and Priorities

The data management goals of the Sierra Nevada Network have developed and evolved over several years and are intended to apply broadly to all Network programs. Adoption of the standards and guidelines contained in this plan can be used as general indicators of program success. I&M staff will be responsible for application to I&M projects; park and USGS staff to their own research and management programs. Park IT staff will be responsible for implementing many of the infrastructure and systems-architecture requirements, including data backup and archiving.

Priority must be given to new projects as these will most easily be able to incorporate the practices we describe. Priority for legacy data should be given to:

1. Data sets needed for current project development
2. Data sets used frequently by park staff or cooperators
3. Historic data sets archived for possible future use

1.4. Organization

The main body of this document addresses broadly these principal subject areas, but directs most of the details, including: descriptions of laws and policies; detailed standards and guidelines; descriptions of Network data management tools and applications, and standard operating procedures into chapter appendices. It has been designed with a consistent modular format to maximize ease of locating specific information on each general topic. The SIEN Intranet site (http://www1.nrintra.nps.gov/im/units/sien/) will contain appendices, guidance documents and standard operating procedures to all Network staff and interested NPS users.

Most chapters include the following sections:

Objectives – Specific statements describing the results to be achieved by the programs subscribing to this plan.

Laws and Policies – Requirements and sideboards established through laws, policies, and mandates at different levels of governance, ranging from broad federal law to NPS and USGS policies, to policies developed specifically for the Sierra Nevada Network.

General Standards and Guidelines – Relevant Network standards and guidelines applicable to specific chapter topic(s). These reference more detailed information in chapter appendices.

At the end of each chapter is a list of references to literature cited, a list of appendices, and chapter credits.

1.5. Key Terms and Definitions

Since the principal purpose of this plan is to provide standards and guidelines for the development, management, and dissemination of *data,* we strive to apply consistent terminology when discussing key concepts and terms.

Data are distinct pieces of information, usually formatted in a special way. This includes symbols or representations of facts or ideas that can be communicated, interpreted, or processed by manual or automated means.

Digital natural resources data generally occur either in tabular form or as files specially formatted for use in Geographic Information Systems (GIS). GIS-formatted data include, but are not limited to, shapefiles, coverages, or georeferenced Tagged Image File Formats (TIFFS).

GIS data contain information about the location and shape of, and relationships among, features on the surface of the earth and are usually stored as geographic coordinates and topology. Topology is used to compare the geographic locations of features relative to one another (e.g., roads connected to a highway, two vegetation polygons adjacent to one another).

Tabular data are usually organized into logical tables of records and fields, arranged in a matrix of rows and columns. Tabular data can be displayed, manipulated, and stored as simple text files or in applications software (e.g., spreadsheets, relational databases). Tabular data can contain reference to a geographic coordinate system and when they do, can also be displayed with GIS software.

Spatial data are any data that reference geographic coordinates. GIS data always contain these references. Tabular data that contain spatial references are also considered spatial. For example, a table of wildlife observations might include the x and y coordinates of the location of observation, along with other information, such as species identity, observer name, and date of observation. These data can be imported and displayed graphically with other spatial data layers, such as roads and highways, or vegetation types. For this reason, the terms *GIS data* and *spatial data* are often used interchangeably. We use *spatial data* to refer to any data set containing reference to a geographic coordinate system.

Tabular data (spatial or not) can be organized within a *Relational Database Management System* (*RDBMS*). RDBMS allow data to be structurally organized for maximum efficiency of storage and retrieval. If properly designed, they enable data to be recombined in any way desired, and therefore provide powerful vehicles for synthesis, analysis, and reporting.

Raw data are data in their original form, i.e., data that have not been altered, summarized, or grouped into broader categories. Raw data can exist in many forms: as hand-written information on field data forms and in notebooks, as unaltered photographs, sound and video recordings, remote sensing imagery, and Global Positioning System (GPS) files. They may be typed or uploaded directly into a computer upon acquisition, or transcribed later from data sheets.

Derived data are raw data that have been processed, or converted to another form using some automated or manual process. Raw natural resources data are often processed and packaged for summation, statistical analysis, and graphical display, or the production of maps and other information products.

Legacy data are data existing prior to the adoption of this data management plan, and at risk of becoming obsolete due to software or metadata limitations.
Programmatic data are data developed specifically for a project with well-defined objectives and specific requirements describing data use and application. As an example, satellite imagery may

be acquired for a geographic area to provide an analytical data framework for monitoring landscape change over a period of time.

Non-programmatic data are data developed for other programs, but are of value to meeting current objectives. For example, the same data acquired for monitoring landscape change might be reclassified and used to evaluate the severity of historical wildfires.

Sensitive data are data that through loss, unauthorized access, or modification, could be used in such as way as to adversely affect valuable resources, the national interest, the conduct of federal programs, or individual privacy. Examples of sensitive natural resources data might include the locations of rare flora or fauna species, caves, or cultural sites.

Metadata is information about data. A complete set of metadata describes all aspects of the data in question, including the 'who, what, where, when, how, and why the data were collected, along with any processing that has occurred. Metadata are considered an essential component of any good data set.

1.6. Making it Work

This plan offers many practical guidelines and solutions, but successful implementation will require continuous formal and informal training and education of staff and cooperators. Good data management is 'an approach and way of thinking', rather than a technological smorgasbord (Michener and Brunt 2000). We assume that Network staff and cooperators are interested in sound data management practices. However, we also recognize that managers and researchers often want minimal intrusiveness and operate under limited time schedules and budgets. Support for data management can be offered by I&M and park technical staff. However, successful implementation will depend on the active participation of everyone who collects and manages data, and so we provide guidelines for data management responsibilities associated with the roles of individuals within park natural resources programs.

To be successful, a significant paradigm shift will be required in how people think about, and work with, data and information. Project managers must be more cognizant of their data and information-related tasks and align to the changes prescribed throughout this plan. Programs will need to invest substantially more resources toward data management than they currently do; the SIEN I&M Program will be expending 33% of its overall budget to data and information management. We acknowledge that positive changes to existing programs will not happen quickly or easily. Hopefully, this plan will serve as a catalyst for incremental change in Network activities that eventually will lead to the adoption of best data management practices – the techniques, methodologies, and standard operating procedures that, through experience and research, have been proven to lead reliably to desired results). Adoption will ultimately improve the production of timely and usable scientific information about the status and trends of park resources. This information can then be made available to park managers and scientists, as well as outside researchers, who collectively seek to understand and conserve those resources.

1.7. References

Boetsch, J. R., B. Christoe, and R. E. Holmes. 2005. Data management plan for the North Coast and Cascades Network Inventory and Monitoring Program. United States National Park Service. Port Angeles, Washington. 88 pp.

Michener W. and J. Brunt. 2000. Ecological Data: Design, Management and Processing. Blackwell Science, Oxford, England. 180 pp.

Mortenson, D. 2005. Data Management Plan for the Inventory and Monitoring Program, Southwest Alaska Network. United States National Park Service. Anchorage, Alaska. 234 pp.

1.8. Appendixes

Appendix 1A) 1998 National Parks Omnibus Act

1.9. Credits

This chapter was written by Pat Lineback and Rosamonde Cook, with assistance from Andi Heard and Meryl Rose. Portions were adapted from Boetsch et al. (2005) and Mortenson (2005).

Chapter 2. Infrastructure and Systems Architecture

Modern information management infrastructure and system architecture represent the foundation of network information management systems. *Infrastructure* refers to the system of computers and servers that are functionally or directly linked through computer networking services. *System architecture* refers to the applications, database systems, repositories, and software tools that make up the framework of our information management enterprise.

Park and Network programs rely heavily on park, regional and national information technology (IT) personnel and resources to maintain its computer infrastructure and systems architecture. Local park program IT specialists are primarily responsible for development and management of local architecture with oversight and support from regional and national IT programs. Development and management of network architecture is more broadly distributed between local, regional and national IT staffs and direct participation by other local and national programs including GIS and I&M staffs.

Ongoing maintenance of computer resources includes, but is not limited to, hardware replacement, software installation and support, security updates, virus-protection, telecommunications networking, and backup/archiving processes. As technology relating to computer hardware resources advances, and the Network's computer infrastructure evolves, modifications will be made to meet information management objectives. Communication and collaboration with park IT specialists is essential to ensuring adequate resources and service continuity for SIEN computer systems architectures.

2.1. Objectives

- Ensure staff have appropriate access to electronic files that are secure and protected from accidental and malicious loss
- Simplify GIS software installation and management, and expedite access and retrieval of GIS data using automation tools.
- Identify and implement appropriate collaborative technologies and tools that enhance networking and communications at both an intra and interagency level.
- Develop a collective long-term strategy for SIEN parks towards identifying and developing a content management system and intranet portal for storing, managing, searching, and disseminating electronic files.

2.2. Laws and Policies

All Department of Interior (DOI) agencies are required to standardize on Microsoft Office products for desktop application and use of spreadsheets, documents, databases, and presentation software (DOI, Chief Information Officer, November 30, 2005).

All Department of Interior (DOI) agencies are required to standardize on ESRI™ products when using GIS software (DOI, Chief Information Officer, November 30, 2005)

In general, Network and park staffs will conform to any federal, Department, and Bureau data infrastructure and systems architecture policies and standards. Further, I&M staff will conform to national I&M Program standards and mandates to assure program integration and information sharing.

Specific infrastructure and systems related Network policies are listed below.

- The network directory structure will have a security system allowing appropriate individuals the ability to manage and maintain restricted files.
- Generally, computer users should not encrypt files and folders because of the potential for lost encryption keys and unrecoverable data. Park IT staffs are responsible for ensuring that file encryption capabilities have been disabled on individual and network computers.
- Installation of ArcGIS software, updates, and tool extensions should be automated and delivered through software installation and management portals.
- SIEN Parks should implement modern data structures and data models that optimize flexibility and scalability and integrate with other national applications.
- SIEN Parks (DEPO excluded) should employ software tool(s) that simplify and streamline staff access to spatial data and metadata.
- GIS software access will be delivered to remote users via a network-based application delivery system that utilizes master spatial data.

2.3. General Standards and Guidelines

These standards and guidelines expand and clarify local and national policies and offer accountability and metrics for measuring staff and program performance.

2.3.1. Network Computer Resources

An important element of an information management program is a reliable and secure network of computers and servers. Other than personal computers, Network programs such as I&M do not have separate computer infrastructure including file or application servers. To improve collaboration and cooperation, any new or existing Network programs will be integrated into existing park computer and network infrastructures. This includes integration and management of electronic files on existing or planned park server architectures.

There are multiple components associated with the NPS infrastructure including park level, regional, and national systems. Each hosts different parts of information systems that are maintained and secured by various park, Network, regional, and national IT and resource specialists and staff. IT duties for Network programs such as I&M are provided mostly by IT staff at SEKI. This includes hosting and managing electronic files being created, managed, and disseminated by Network staff and cooperators. It is anticipated that YOSE IT staff will provide additional future specialized IT support as Network parks begin implementing this plan. As an example, YOSE IT has Citrix administration expertise that will enable park staff to utilize ArcGIS software and spatial data across wide area networks with lower bandwidth connections. Implementation of this application will broaden GIS availability and simplify spatial data administration lowering costs and leveraging the use of GIS.

A basic network framework for the SIEN I&M program is described in Appendix 2A - Network Framework for SIEN Inventory and Monitoring.

2.3.2. Network Security

Local and wide area networks currently conform to Department of Interior security guidelines. All sensitive electronic files should be placed in protected folders with limited read and write access. Electronic file and directory permissions administration will be partially decentralized with file and folder administrative rights controlled by limited Network staff including IT and a few trained program staff. A database of file permissions will be developed and maintained. Deployment of a Microsoft utility will enable the management and restoration of New Technology File System (NTFS) file and folder permissions on network servers. YOSE staff needing read or write access to the SEKI directory structure or SEKI staff needing read or write access to the YOSE directory structure will be enabled through Microsoft Active Directory and NTFS permissions.

There are very rare circumstances where encryption of sensitive electronic files and/or folders may be necessary. As an example, files on removable media or portable devices may rarely need additional protection from unauthorized access or use. Appendix 11F and 11G, General Support Services Backup Activity Plans for SIEN Parks, provide specific guidance to Network staff who must implement electronic file encryption or other security measures (encryption policy and procedures are under development).

2.3.3. National Applications and Tools

Various national offices, including the NPS Natural Resource Program Center (NRPC) and the national I&M program actively develop and implement national-level, program-wide information management systems. Other programs such as cultural resources and maintenance have their own distributed applications that often intersect and also have broader program application and use. Principal applications and data repositories developed for the I&M and Service-wide programs, and referenced throughout this document, are summarized below (see Appendix 2B for more details).

The NPS Data Store – an Internet-based graphical search interface that links data and metadata to a searchable server on which data are organized by NPS units, offices and programs. Access permission is required.

NPS Focus Digital Library and Research Station – a decentralized digital imagery and data management system, implemented through a central public Internet portal sponsored by the NPS Office of the Chief Information Officer. Includes to the NPS Data Clearinghouse.

NPS Data Clearinghouse – the central repository for NPS GIS data available to the public. Implemented through the NPS Focus gateway.

NPS Automated National Catalog System (ANCS+) – the official curatorial cataloging system of the NPS.

NatureBib – the master database for natural resources bibliographic references for the NPS. Implemented on a dual system of public and secure servers to protect sensitive information.

NPSpecies – the master species database for the NPS. Includes park-specific lists of the plant and animal species that occur in or near the parks along with records (vouchers, observations, and literature citations) of physical or written evidence of their occurrence and status. Implemented on a dual system of public and secure servers to protect sensitive information.

Biodiversity Data Store – an Internet-based repository of documents, GIS and other data sets that contain information of plant and animal species that contribute to knowledge of the biodiversity in the National Parks

NPSTORET (also known as NPS Water Quality Database Templates) – a NPS database designed to facilitate park-level standardized reporting for STORET, an Internet-based interagency water quality database developed and supported by the Environmental Protection Agency (EPA) to house local, state, and federal water quality data collected in support of managing the nation's water resources under the Clean Water Act.

2.3.4. SIEN I&M Web Site

The SIEN I&M web site provides to the public, general information about the Network and its parks, and the I&M Program, including information on inventories, monitoring, data management, and reports & publications as directed by WASO's general webpage templates. The website will also be used to serve products such as results of I&M vital signs monitoring (e.g., executive briefs, progress reports, trend reports, etc.) and the SIEN Data Management Plan. Future plans include the inclusion of a linked blog site, for updating park staff, partners, and interested parties on recent program highlights and newsworthy happenings. An Intranet portal is also planned, where park and Network staff can share information and data that are not ready for public posting.

Structure and content of the SIEN I&M web site comply with National Park Service standards, FOIA requirements for accessibility of public information, and Section 508 of the Rehabilitation Act which ensures access to electronic information by people with disabilities (http://www.section508.gov/). Web development adhered to the national I&M program's Web Author Instructions, Practices, and Tools (http://www1.nrintra.nps.gov/im/datamgmt/webdev/index.cfm), and to policies and guidance on natural resources web publishing provided in the following documents: (1) Director's Order 70: Internet and Intranet Publications (Appendix 2F); (2) NPS Internet/Intranet Publications Manual (http://inside.nps.gov/helpdesk/manual.htm); and (3) NPS Natural Resource Web Manual (http://www1.nrintra.nps.gov/webnotes/NRWebManual/index.cfm). Sample Network page models were provided by WASO I&M support staff, along with security coding and support during development in quality assurance and quality control.

2.3.5. Collaboration Technologies

A small group of Network park staff – the Collaborative Technologies Work Group – evaluated current collaborative technologies and developed recommendations aimed at improving remote collaboration between groups and/or individuals at physically separate locations. Effective and efficient communications are essential and staffs must consider using both emerging and existing

technologies as a means to improve communications. The I&M program is an good example of a program requiring collaboration between employees at four national parks spread across a large geographic area making face-to-face meetings difficult.

This workgroup completed a needs assessment and reviewed technologies relating to meeting facilitation, audio conferencing, content management, file versioning, and personal communications. One of their most significant findings was that NPS security policy limits the availability of various collaboration technology tools. NPS security policies are becoming increasingly restrictive leaving parks with less flexibility for adopting some collaborative technology. Under current policy, NPS programs are limited to technologies within the NPS intranet. Specifically, NPS employees are prohibited from using any application hosted on external websites.

A brief summary of some of the evaluated collaborative tools and recommendations are listed below. A more thorough detailed review and assessment is included in Appendix 2C, Collaboration Technologies Needs Assessment, Sierra Nevada Network National Parks.

2.3.5.1. Web Meetings: There are many web and video conferencing applications available commercially, however, all but Lotus Sametime Meeting are hosted on external servers and, therefore, prohibited by current NPS IT policy. Network staffs have not tested the Sametime application.

2.3.5.2. Content Management: An important objective of the Sierra Nevada Network is the development and implementation of a strategy for a content management system and local intranet portal for storing, managing, and retrieving electronic files. A content management system is a computer software system for organizing and facilitating collaborative creation of documents and other content. Further, this system should enable efficient searching and retrieval of needed electronic files. This Network creates thousands of electronic files each year in an increasingly complex environment. It is essential to begin developing strategies for improving the management and dissemination of these files using consistent and structured methods. Adoption of a content management system should be preceded by a formal needs assessment that considers comprehensive staff needs. Unfortunately, current NPS policy limits the Network from selecting and implementing most content management systems regardless of how effective or efficient any particular application might be.

A limited needs assessment was completed to better understand the technology and its use. Several applications were evaluated which provide content management services. These included:

- Wiki – An open source content manager supporting the FAQ's of the Santa Monica Mountains Recreation Area website.
- Drupal – An example of a Wiki open source content manager.
- Microsoft Sharepoint Services – Commercial content manager.
- Microsoft Sharepoint Portal Server - Commercial content manager.
- WebEx WebOffice – Commercial content manager.

The collaborations technologies work group recommended adopting Microsoft Sharepoint Services, which is hosted by the PWR regional office.

2.3.5.3. Personal Communications: The use of email and telephones to conduct personal communication between individuals at separate locations is ubiquitous. Instant messaging compliments these and adds additional functionality. Email is comparatively slow and a telephone call limits access to an individual's desk or cell phone availability. Instant messaging is near real-time and offers additional flexibility for communicating with one or more employees. Lotus Notes Sametime Instant Messaging is currently available to all NPS employees with a NPS computer network login profile. The collaborative technologies work group's recommendation was to better utilize and take advantage of Lotus Notes Sametime Instant Messaging.

2.3.6. Geographic Information Systems

ESRI GIS applications and products are regularly used by Network staff. This includes use of map products or analyses generated from GIS software. Additionally, about 75 staff across SIEN and local USGS programs have technical ArcGIS skills and abilities. Skill levels range from infrequent users with beginning skills to advanced users who can complete complex analyses.

A brief overview and current GIS issues at each of the four Network parks are described below.

2.3.6.1. Devils Postpile National Monument: Currently, SEKI staff provide GIS support to DEPO and manage their spatial data. This small 800 acre park unit has limited and infrequent needs for GIS support at the local level.

2.3.6.2. Sequoia and Kings Canyon National Parks: Over a period of 20 years, the SEKI GIS program has evolved from a single Unix workstation located in a closet and running public domain GRASS software to a GIS program today that includes about 30 staff who are using GIS software on their desktop computers. Additional history on the evolution of the SEKI GIS program is described in Appendix 2D.

ArcGIS license manager software runs on two servers (both USGS and NPS have their own license allocations) at Ash Mountain Headquarters and distributes client software licenses to local area network staff. Recent monitoring has shown that it is not unusual for ten staff to be simultaneously using ArcGIS software within the Ash Mountain LAN. ArcGIS software installation, updates, and extensions are available using the SEKI intranet, but require user interactions.

To run ArcGIS *off* network, individual hardware dongle keys must be installed and configured on each computer.

A larger direct and indirect GIS user base has created challenges that are partially addressed throughout this plan. This rapid growth and change in use and application of GIS at SEKI has caused severe growing pains including: the inability to adequately develop and maintain essential spatial data, inadequate technical support and training for staff, and inadequate planning and adapting to changes in data architectures and changing technology. Flat budgets make it increasingly difficult to support increasing demands and requirements. Users need more support

with their projects due to GIS integration including data development, mapping, and consultation. Creating the data storage and archival infrastructure that allows distributed access to data has become more difficult as more individuals tap into the wealth of GIS data and have changed the underlying data architecture requirements. Federal policy requires federal agencies to make GIS data and metadata available for sharing with the public and other organizations. There are not enough resources to adequately complete these tasks. Coordination requirements across SEKI programs and at an interagency level have dramatically increased with the growing GIS user base. Development and migration of existing spatial data to new data models that are flexible and support multiple programs requires a new level of commitment requiring more staff participation and engagement by multiple divisions. This is difficult given the SEKI traditional hierarchical management structure that has promoted autonomy within each division. The long-term GIS budget outlook is grim and we expect these issues will continue to worsen.

2.3.6.3. Yosemite National Park: Appendix 2E, Early GIS History at YOSE, describes the roots and pioneering deployment and development of GIS up to 1992. YOSE staff will need to develop a narrative for this section of document.

2.3.6.4. Recommendations: At the park level, the GIS programs in this Network run independently of one another, but the nature and complexity of GIS issues are similar for both SEKI and YOSE. There are some common infrastructure solutions that are applicable to these parks:

- Automate GIS Software Installations and Updates – Because of the increasing numbers of GIS software users, it is most efficient to implement ESRI software installation protocols that streamline installation and updates including silent updates for existing ArcGIS software users. Initial software installations should be completed through a flexible intranet portal, but future updates conducted through silent and bulk updates.
- NPS Theme Manager - There are hundreds of spatial data layers available to staff. They are hosted in multiple locations and require intimate familiarity with spatial data to find, understand, and use these data. It is inefficient, particularly for infrequent users to find and use the data they need. Therefore, both SEKI and YOSE should deploy and manage NPS Theme Manager to deliver data and metadata to Park staffs. This ArcGIS extension, developed by the Alaska Regional Office, provides a front end for spatial data and makes it easier for staff to locate and access spatial data. The installation and management can be semi-automated and updates roll out to users each time they start a new ArcGIS project.
- Provide ArcGIS software access to park GIS users with low bandwidth connections – There are some GIS users in remote park locations that have lower bandwidth connections in wide area network environments. Without replicating master spatial data to their personal computers, it is difficult to provide these users with access to contemporary data and updated ArcGIS software. A recommended strategy for providing access to master spatial data and ArcGIS software is deployment of server-based ArcGIS through a Citrix application interface. Citrix is operational at both YOSE and SEKI. Preliminary testing of ArcGIS software indicates broader deployment of Citrix-served ArcGIS is an effective solution.

2.3.7. Database Applications Development

Desktop versions of SIEN databases will be in the latest Microsoft Access format unless otherwise specified in the project study plan. Desktop databases produced under contract will also use Microsoft Access unless otherwise specified in advance.

SIEN projects will invest in modular, standalone project databases that share design standards and links to centralized lookup tables, and can be developed, maintained, and archived separately. There are numerous advantages to this strategy:

- Data sets are modular, allowing greater flexibility in accommodating the needs of each project area.
- Individual project databases and protocols can be developed at different rates without a significant cost to data integration.
- Any project database can be modified without affecting the functionality of other project databases.
- We avoid a large initial investment in a centralized database and the concomitant difficulties of integrating among project areas with very different – and often unforeseen – structural requirements.
- Potentially greater efficiency for interdisciplinary use.

The SIEN will develop a set of master look-up tables or an Integrated Data Dictionary (also referred to as a Buffet of Fields and Tables [BOFAT] by WASO), that will allow databases across projects and across disciplines to remain as consistent as possible. Likely lists of standard attributes and values to be developed include park codes, place names, watershed codes, species identifiers, and vegetation or land type classifications (See Chapter 5 – Database Design for more details). To the extent possible, SIEN will work with WASO and other Networks to develop and utilize shared look-up tables for data consistency and inter-operability at a higher level across programs.

As database applications dictate, SIEN and park staffs should evaluate the need for a more robust enterprise relational database management system such as Microsoft SQL Server.

2.3.8. Word Processing

All reports and other textual documents will be finalized in editable electronic format in the NPS standard current version Word Processor or Adobe Acrobat, unless otherwise specified in the project study plan. Distribution copies will be converted to the latest version of Adobe Acrobat.

2.3.9. Digital Data Formats

2.3.9.1. Databases: Well thought-out database design standards are necessary to promote compatibility among data sets that will be aggregated and summarized, encourage sound database design and facilitate interpretability of data sets. As much as possible, SIEN standards for database design will mirror those conveyed through the Natural Resource Database Template (NRDT). The NRDT is a flexible, relational database template, designed for storing inventory, monitoring, and research data (including raw data collected during field studies). The template was designed to be used in the development of standalone databases and databases that interact

with GIS software (e.g., ArcView or ArcGIS). A description of the NRDT and a working database is presented in Chapter 5 – Database Design. Additional information can be found at (http://science.nature.nps.gov/im/ apps/template/index.cfm).

2.3.9.2. Spatial data: The coordinate system standard for SIEN parks is Universal Transverse Mercator with North American Datum 1983. Generally, existing spatial data should be migrated to this data projection and it should be used for any new spatial data development. Most significant data for all four Network parks are already in this modern coordinate system format.

Currently, Network and park staff use many different raster and vector data formats for storing and managing spatial data. Vector data formats commonly include Arc/Info coverages, shapefiles, and personal geodatabases. Raster data formats commonly include MRSID, TIFF, GEOTIFF, and Grid structures. ESRI software products have historically demonstrated excellent backward compatibility with older data structures and there is no SIEN requirement that data formats be migrated to the more modern data structures.

With the imminent release of ArcGIS 9.2, a new data structure called a *file geodatabase* will be released. Beginning in 2007, the *file geodatabase* will be the recommended data structure for significant new spatial data development in the SIEN. The *file geodatabase* is a spatial "container" used to hold a collection of datasets. Each dataset is held as a file that can scale up to 1 terabyte in size. Network and park staffs should develop strategies for migrating existing data and developing new spatial data into the *file geodatabase* format. This data format has numerous advantages including: a) providing a relatively simple and scalable solution for GIS users; b) works across operating systems; c) scales to handle very large datasets; d) has an efficient data structure optimized for performance and storage and has feature geometry storage requirements one-third of what is required by shapefiles and personal geodatabases; and e) offers more flexibility for multi-user editing and use of the same spatial data layer.

The national NPS GIS office (WASO-GIS) is facilitating the development of standardized data models for spatial data. These data models are designed to be flexible, scalable, and integrate with other national applications. GIS tools are being developed to simplify development and management of this data. Current WASO-GIS spatial standardization efforts include buildings, trails, and vegetation. As these new data models are released, Network parks should implement these new data models and convert legacy data into these new data models.

2.4. References

Boetsch, J. R., B. Christoe, and R. E. Holmes. 2005. Data management Plan for the North Coast and Cascades Network Inventory and Monitoring Program. USDI National Park Service. Port Angeles, Washington. 88 pp.

Daley, R. 2005. Data and Information Management Plan, Greater Yellowstone Network Inventory and Monitoring Network. United States National Park Service, Bozeman, Montana. 82 pp.

Department of Interior, Chief Information Officer. November 30, 2005. Office of Chief Information Officer, Directive 2006-003, Distribution of the Department of the Interior Technology Reference Model Version 3.0.

Mortenson, D. 2005. Data Management Plan for the Inventory and Monitoring Program, Southwest Alaska Network. United States National Park Service. Anchorage, Alaska. 234 pp.

National Park Service. 2001. Director's Order 70: Internet and Intranet Publishing.

Svancara, L. K. 2006. Upper Columbia Basin Network, Information Management Plan. Natural Resource Report NPS/PWR/UCBN/NRR—2006/000. National Park Service, Moscow, Idaho.

2.5. Appendixes

2A) Computer Network Framework for SIEN Inventory and Monitoring
2B) National Repositories for Digital Data and Information
2C) Collaboration Technologies – Evaluated Applications
2D) Sequoia and Kings Canyon GIS History and Current Issues
2E) Yosemite GIS Development and History Before 1992
2F) Director's Order #70: Internet and Intranet Publishing

2.6. Credits

This chapter was written by Pat Lineback with contributions from Anne Birkholz (Southern Sierra Interagency Data Analyst), Paul Gallez (Yosemite NP), Bob Basham (Information Technology, Sequoia and Kings Canyon NP), Dan Sohn (Information Technology, Yosemite NP), and Rosamonde Cook, and with information provided by Boetsch et al. (2005), Daley (2005), Mortenson (2005), Svancara, (2006).

Chapter 3. Project Development and Data Management Workflow

To better understand the information management needs of the SIEN, it is useful to understand the general work flow of project development and the information management tasks associated with each stage. There are two main types of projects handled by Network natural resources staff and the Inventory and Monitoring Program:

1. short-term, which may include individual park research projects, inventories, or pilot work done in preparation for long-term monitoring or research.

2. long-term, including Network vital-signs monitoring projects central to the I&M program and multi-year research projects and monitoring performed by other park programs, agencies and cooperators. Long-term projects will often require a higher level of documentation, peer review, and program support.

For information management, the primary difference between short- and long-term projects is an increased need to adhere to and maintain standards for long-term projects. Maintaining standardization from year-to-year will be necessary when comparing data over an extended period of time (decades for long term monitoring).

3.1. Five Stages of Project Development

To ensure the development of high quality scientific information, data management must be a component of all aspects of project development. Projects can be divided into five primary stages:

- planning and approval
- design and testing
- implementation
- product integration
- evaluation and closure

Each stage is characterized by a set of activities carried out by staff involved in the project (Figure 2). Primary responsibility for these activities rests with different individuals according to the different phases of a project. Additional discussion of the different roles and responsibilities of park and Network staff can be found in Chapter 4 of this plan. Specific project tracking methods and documentation are presented in Chapter 12.

1. *Planning and Approval* – This is the stage at which many of the preliminary decisions are made regarding project scope and objectives. Funding sources, permits and compliance are also addressed at this time. Primary responsibility rests with project leaders and program administrators. Although this phase lacks specific data management activities, it is important that data managers remain informed. This is especially true as timelines for deliverables are finalized. All contracts, agreements, and permits should include standard language that describes the formats, specifications, and timelines for project deliverables.

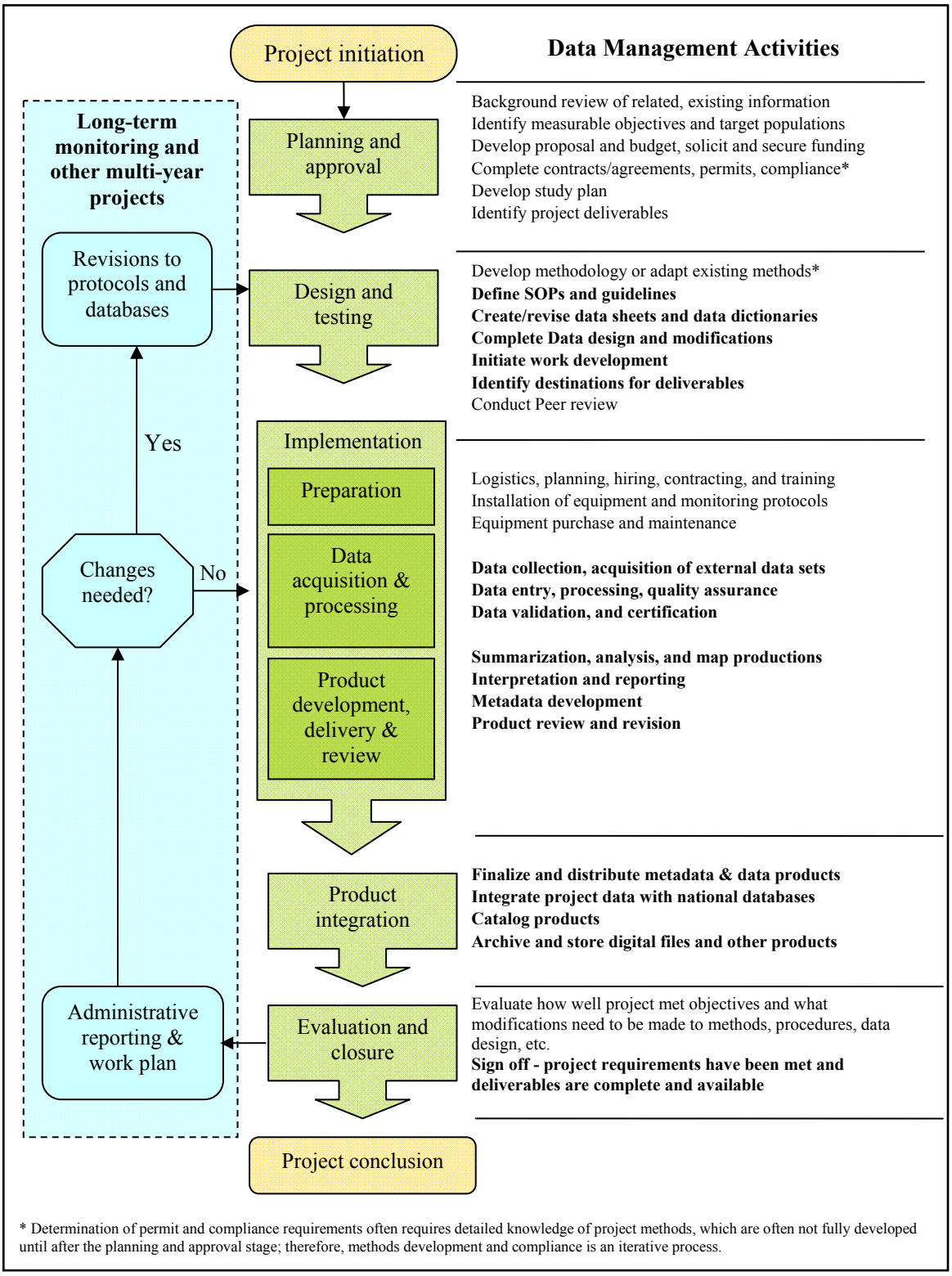

Figure 2. Data management steps during the five stages of project development with data management activities involved in each phase. Core activities are in bold.

2. *Design and Testing* – All details regarding how data will be acquired, processed, documented, analyzed, reported and made available to others are worked out. The project leader is responsible for developing and testing project methodology, or for modifying existing methods to meet project objectives. It is critical that project leaders and data managers work together during this phase. The dialog between these individuals will help to build and reinforce good data management throughout the project. By beginning collaborative development as soon as possible after project approval, data integrity and quality can most easily be assured.

   An important part of this collaboration is the development and detailed documentation (Chapter 10) of project databases, including relational diagrams, data dictionaries, business rules, and front-end programming, along with the formal metadata. Devoting adequate attention to this aspect of a project is the single most important part of assuring the quality, integrity and usability of the resulting data. Although it is possible that post-hoc changes will occasionally be required, it is important to minimize these through careful initial design work.

3. *Implementation* – Data are acquired, processed, error-checked (Chapter 6) and further documented (Chapter 8), and products such as reports, maps, GIS themes, and other products are developed. The project leader oversees all aspects of implementation – from logistics planning, contracting, training, and equipment procurement to data acquisition, documentation, report preparation and final delivery. Data management staff function primarily as facilitators – providing training and support for database applications, GIS, GPS and other data processing applications; facilitation of data summarization, validation and analysis; and assistance with the technical aspects of documentation and product development. The specific roles of data management staff will depend primarily on the technical capabilities of the project staff. As much as is possible, these roles should be worked out in advance of 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 (i.e., protocol, study plan, contracts, agreements or permits).

4. *Product Integration and Distribution* – Data are merged from the working database to a master database maintained on the Network server, certified data sets and their metadata are finalized and posted in national repositories, and products are distributed or otherwise made available to their intended audience (Chapter 10). Data are merged to the master only after the annual working data set has been certified for quality by the project leader. Product integration may include creating records for reports and other project documents in NatureBib, posting imaged documents to NPFocus, and updating NPSpecies and ANCS+ to reflect any new species occurrence information or specimen vouchers derived from the project. These updates allow the information from the project to be searchable and available to others via service-wide search engines. Certain projects may also have additional integration needs, such as when working jointly with other agencies.

   In general, all raw and derived data products, metadata, reports and other documentation should be delivered to the data steward assigned to the project. Administrative records

should be delivered to appropriate park and Network staff as specified. All project deliverables should be distributed according to specifications, which should be stipulated in all protocols, contracts, agreements, and permits. Products that do not meet program requirements should be returned for revision.

5. *Evaluation and Closure* – Records are updated to reflect the status of the project and its associated deliverables in a Network project tracking application (Chapter 12). 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 non-cyclic projects, this phase represents the completion of the project. After products are catalogued and made available, program administrators, project leaders, and data managers should work together to 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. For monitoring protocols, careful documentation of all changes is required. Changes to methods, Standard Operating Procedures (SOPs), and other procedures are maintained in a tracking table associated with each document. Major revisions may require additional peer review.

3.2. The Data Life Cycle

Data take on different forms during various phases of a project, and are maintained in different places as they are acquired, processed, documented, analyzed, reported, and distributed. What we refer to as the "data life cycle" is characterized by a series of events that we can model to facilitate communication (Figure 3) as follows:

1. *Acquire data* – Data are acquired in digital or analog form. Digital data can be recorded on handheld computers and PDAs, tablets, or laptop computers. Analog data are entered on field data sheets.

2. *Archive raw data* – Copies of all raw data files are archived intact. Digital files are copied to the digital library (the set of LAN folders created for the project); hard copy forms are either scanned and placed in the digital library or are copied and placed in the archives. Archiving or scanning of hard copy data forms may occur at the end of a season as a means of retaining all marks and edits made during the verification and validation steps.

3. *Enter/import data* – Analog data are entered manually and digital data files are uploaded to the working database.

4. *Verify, process, and validate* – Accurate transcription of the raw data is verified; data are processed to remove missing values and other flaws; and data are validated through visual inspection and queries to capture missing data, out-of-range values, and logical errors.

5. *Documentation and certification* – Develop or update project metadata and certify the data set. Certification is a confirmation by the project leader that the data have passed all quality assurance requirements and are complete and documented. It also means that data and metadata are ready to be posted and delivered.

6. *Upload data* – Certified data are uploaded from the working database to the master project database. This step might be skipped for short-term projects where there is no need to distinguish working data for the current season from the full set of certified project data.

7. *Archive versioned data set* – The SIEN Project Data Certification Form is completed (Appendix 3A). Copies of the certified data and metadata are placed in the digital library. This can be accomplished by storing a compressed copy of the working database or by exporting data to a more software-independent format (e.g., ASCII text).

8. *Disseminate data and update national databases*– Certified data and metadata, and digital image products are posted to national repositories (the NPS Data Store, Biodiversity Data Store, NPS Focus) to make them more broadly available to others. National databases, including NPSpecies, NPSTORET, and ANCS+ are updated with data obtained from certified data sets.

9. *Reporting and analysis* – Certified data are used to generate data products, analyses, and reports, including semi-automated annual summary reports for monitoring projects. Depending on project needs, data might be exported for analysis or summarized within the database.

10. *Distribute information products* – Information products such as reports, maps, and checklists are disseminated to the public through the SIEN website and NPS Focus, and catalogued in NatureBib.

11. *Share data and information* – Data, metadata, reports and other information products can be shared in a variety of ways – by FTP or mailing in response to specific requests, or by providing direct access to project records to park staff and cooperators.

12. *Track changes* – All subsequent changes to certified data are documented in an edit log, which accompanies project data and metadata upon distribution. Significant edits will trigger reposting of the data and products to national databases and repositories.

13. *Store products* – Reports and other data products are stored according to format and likely demand, either in the digital library, on off-line media, or in the document archives.

14. *Catalog project products* – Catalog products and all information associated with a project, including results of analyses and paths of dissemination. Project tracking databases can be useful tools for this purpose.

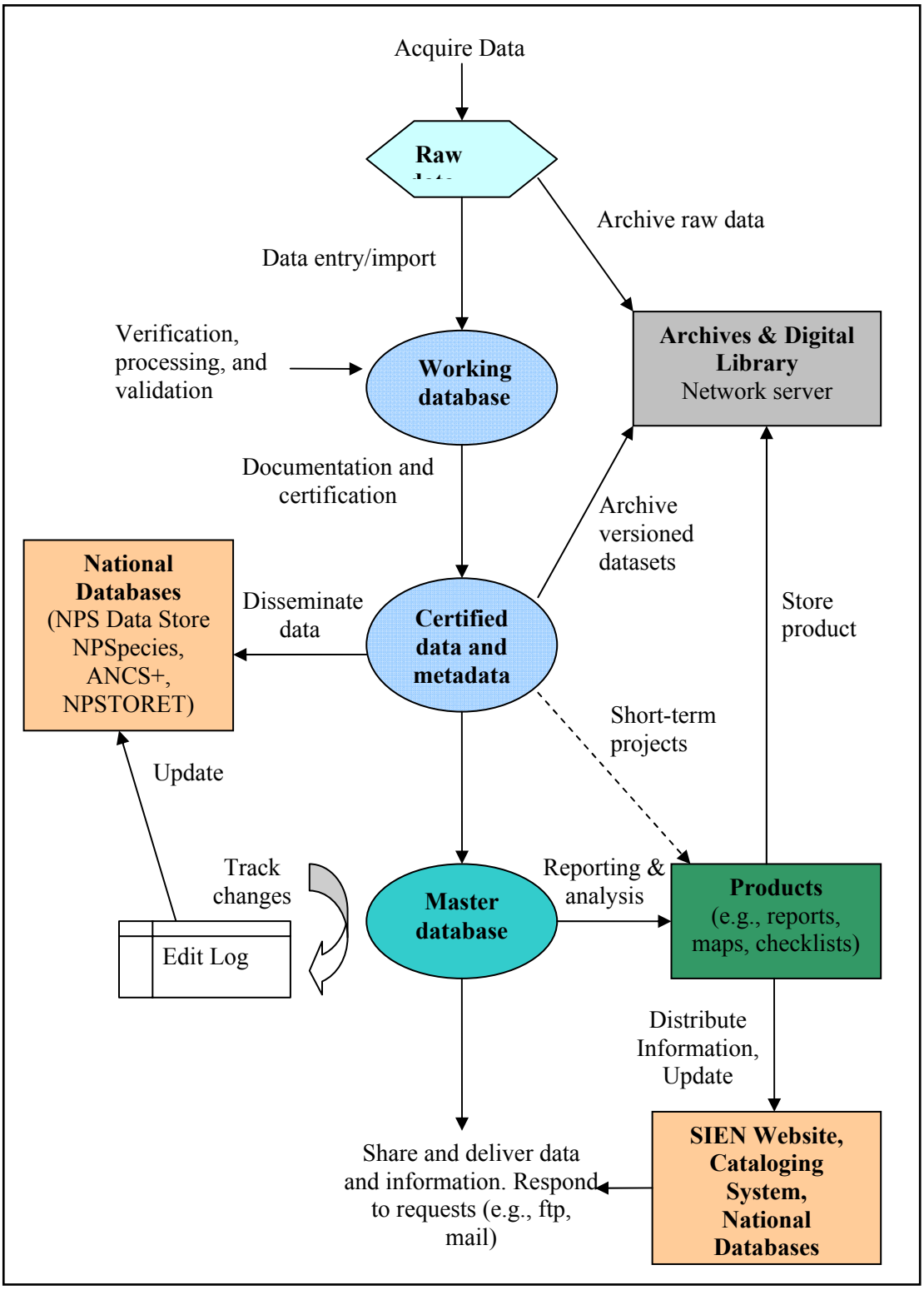

Figure 3. The data life cycle.

3.3. References

Boetsch, J. R., B. Christoe, and R. E. Holmes. 2005. Data management Plan for the North Coast and Cascades Network Inventory and Monitoring Program. USDI National Park Service. Port Angeles, Washington. 88 pp.

3.4. Appendixes

3A) Project Data Certification Form

3.5. Credits

This chapter was written by Rosamonde Cook adapted from concepts and material developed by Boetsch et al. (2005).

Chapter 4. Data Management Roles and Responsibilities

4.1. General Standards and Guidelines

An increasing demand for more detailed, high quality data and information about natural resources and ecosystem functions requires a group of people working together to steward data and information assets. Knowledgeable individuals from many disciplines must come together to ensure that data are collected using appropriate methods, and that resulting datasets, reports, maps, models, and other derived products are well managed. Datasets and the presentations of these data must be credible, representative, and available for current and future needs. Table 1 summarizes general types of data stewardship activities and the roles typically associated with them.

Table 1. Categories of data stewardship involving Sierra Nevada Network personnel

Stewardship Activity	Description of Activities	Principal Roles
Production	Creating data or information from any original or derived source. This includes recording locations, images, measurements, and observations in the field, digitizing source maps, keying in data from a hardcopy source, converting existing data sources, image processing, and preparing and delivering informative products, such as summary tables, maps, charts, and reports.	Project Crew Leader, Project Crew Member, Data/GIS Specialist or Technician
Analysis	Using data to predict, qualify, and quantify ecosystem elements, structure, and function as part of the effort to understand these components, address monitoring objectives, and inform park and ecosystem management.	Ecologists, Resource Specialists
Management	Preparing and executing policies, procedures, and activities that keep data and information resources organized, available, useful, compliant, and safe.	Project Leader, Data Manager, GIS Manager, IT Specialist, Database Manager, National-level Data and Information Managers
End Use	Obtaining and applying available information to develop knowledge that contributes to understanding and managing park resources.	Project Leader, Park Managers and Superintendents, Others

Although each position is associated with only one category in the table according to overriding responsibilities, many positions contribute to multiple categories. The degree to which specialization can occur with regard to data management responsibilities will depend on program

resources, staff and funding. For a very small program, all of these activities might be accomplished by a single individual. For a large program, such as the SIEN I&M Program, a much higher degree of diversification and specialization are required. Table 2 summarizes the roles and responsibilities of personnel that may participate in a program such as this. Roles are listed 'from the ground up' to help demonstrate the hierarchy of responsibilities. For example, a project leader is ultimately responsible for the activities listed in the field level roles of crew leader and crew member. It is possible that one person/position has more than one role. For example, the Ecologist and Project Leader roles and responsibilities may fall under one position. See Appendix 4A for a much more detailed definition of individual responsibilities, and Appendix 4B for a discussion of coordination among roles and responsibilities.

Table 2. Summary of Roles and Responsibilities for a Network Inventory and Monitoring Program.

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. Maintain connections between the LAN and the Internet. Work with the Network Data Manager and GIS liaisons to establish a directory structure that provides local access and security for natural resource data. 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. Review and approve proposed changes to project protocols prior to implementation.
Project Leader	Oversee and direct operations for one or more Network projects. Maintain communication with project staff, Network Data Manager, and resource specialists regarding data management.
Resource Specialist	Understand the objectives of the project, the resulting data, and their scientific and management relevance. Make decisions with regard to validity, utility, sensitivity, and availability of program data. Describe, publish, release, and discuss the data and associated information products.

Table 2. Summary of Roles and Responsibilities for a Network Inventory and Monitoring Program. (continued)

Role	Primary responsibilities related to data management
GIS Manager	Support park management objectives. Coordinate and integrate local GIS and resource information management with Network, regional, and national standards and guidelines.
Data Manager	Provide planning, training, and operational support for the awareness, coordination, and integration of data and information management activities, including people, information needs, data, software, and hardware. Coordinate internal and external data management activities.
Database Manager	Apply particular knowledge and abilities related to database software and associated application(s).
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.
National-level data management support staff	Provide service-wide database design, support, and services, including receiving and processing to convert, store, and archive data in service-wide databases.
End Users (e.g., park managers, researchers, other agency staff, members of the public)	Appropriate use and application of data and derived products and for providing feedback for improvements.

Data management is about people and organizations as much as it is about information technology, database theory, and applications. For large park resources programs, such as the SIEN Inventory and Monitoring program to work effectively, everyone within the program, and all those in collaboration with it, must have stewardship responsibilities for the production, analysis, management, and/or end use of data produced by the program. In order to meet new data management goals and standards developed by the National Park Service and its constituents, program staff must understand how data and information flow (discussed in Chapter 3) and what their roles and responsibilities are in this process. A role is a function or position (e.g., Data Manager). A responsibility is a duty or obligation (e.g., review data records).

4.2. Collaboration

The Natural Resource Challenge states that collaboration among the National Park Service, other public agencies, universities, and non-governmental organizations is necessary to effectively acquire, apply, and promulgate the scientific knowledge gained in National Parks. A detailed approach to collaboration in data management should be developed for all Network natural resources programs.

4.3. References

Beer, M., Nance, E., Wright, A, Powell, M. and R. DenBleyker. 2005. Northern Colorado Plateau Network, Data Management Plan. National Park Service, Inventory and Monitoring Program, Northern Colorado Plateau Network, Moab, Utah. 86 pp. plus appendices.

Daley, R. 2005. Data and Information Management Plan, Greater Yellowstone Network Inventory and Monitoring Network. United States National Park Service, Bozeman, Montana. 82 pp.

Mortenson, D. 2005. Data Management Plan for the Inventory and Monitoring Program, Southwest Alaska Network. United States National Park Service. Anchorage, Alaska. 234 pp.

Stevens, S. M. and G. Entsminger. 2004. Information Management Plan for the Inventory and Monitoring Program, Northeast Coastal and Barrier Network. United States National Park Service. Kingston, Rhode Island. 99 pp.

Wilder, D. 2005. Data Management Plan, Central Alaska Network. United States National Park Service. Inventory and Monitoring Program, Fairbanks, AK. 86 pp.

4.4. Appendixes

4A) Sierra Nevada Network Inventory and Monitoring Program, Data Management Responsibilities

4B) Sierra Nevada Network Inventory and Monitoring Program, Coordination of Data Management Roles and Responsibilities

4.5. Credits

This chapter was written by Rosamonde Cook, adapted from concepts and material developed by Stevens and Entsminger (2004), Beer et al. (2005), Daley (2005), Mortenson (2005), and Wilder (2005).

Chapter 5. Databases

5.1. Objectives

Adopting sound database design principles that will help facilitate interpretability, ensure integrity, and promote compatibility of natural resources data collected in the parks.

5.2. Laws and Policies

To help meet these objectives, the NPS I&M Program has developed a standardized database template for inventory and monitoring projects, the Natural Resource Database Template (NRDT) (http://science.nature.nps.gov/im/apps/template/index.cfm). I&M Network programs are expected to design databases used for natural resource inventories and vital signs monitoring on this template.

5.3. General Standards and Guidelines

5.3.1. The Natural Resource Database Template

The National Park Service I&M Program's Natural Resource Database Template (NRDT) is a set of Microsoft Access relational database tables that parks and networks can use to develop applications for capturing natural resource inventory and monitoring data.
The NRDT:

- Provides both a data interchange standard and a standard MS Access database core that allows flexibility in application design.
- Serves as a starting point for application development that can be extended as necessary to accommodate any inventory or monitoring field sampling protocol.
- Standardizes location and observation data to facilitate the integration of data sets.
- Acts as a design platform for developing database applications in MS Access allowing users to enter, edit, display, summarize, and generate reports for inventory or monitoring data sets.
- Integrates with other I&M data management systems and data standards including the NR-GIS Metadata Database, Geographic Information System (GIS) tools and data, the NPS GIS Committee Data Layers Standard and the NPS Metadata Profile.

5.3.2. Database Development

Communication is a vital part of developing a suitable database design for individual projects. One mechanism for this communication is collaborative development of data models by data managers and project leaders. Data models combine diagrams with associated descriptions and are completed in three stages: conceptual, logical and physical. Each is outlined below, using as an example, a database created for monitoring the status of mountain yellow-legged frog (*Rana muscosa*) populations following removal of non-native fish from high mountain lakes in Sequoia and Kings Canyon National Parks.

5.3.2.1. Conceptual Data Models : Conceptual data models are constructed to graphically portray processes specifically related to the implementation phase of a project – especially those involving acquisition, processing, and QA/QC of data (Figure 4). They are software-

independent, free of details and focus on capturing enough information needed to accurately depict project data design. Conceptual data models contain the following:

- An overview of the sampling scheme in layman's terms.
- A flow diagram of procedures, and when and what information is being collected or produced.
- A description or mock-up illustration of how data should be presented.

5.3.2.2. Logical Data Models: A logical data model is an abstract representation of database entities, their relationships and key attributes (Figure 5). Logical data models are intended to provide a conceptual framework for the database developer as well as a schematic that can be easily interpreted by the layman. They are not intended to provide full representation of the physical database, but rather to facilitate communication and understanding between developers and users. A logical data model should be produced early in project development (ideally beginning with early protocol development) and serve as a guide to the physical data model that documents the actual implementation of the database.

A logical data model should include the following five components:

1. *Database tables* – sufficient to model the sampling scheme. Enough information should be provided to allow users to understand how key information and data are distributed among tables in the database (e.g., sampling location, frequency, and measurements to be taken.
2. *Field attributes* – data types, units of measure, acceptable value ranges, and other business rules.
3. *Logical relationships* – how data tables are logically related (e.g., each location will be visited numerous times, one or more species might be observed at each location per visit)
4. *Structural hierarchies* – the structure and order of relationships between data tables, which can be determined once the logical relationships are known, e.g.,
   - Site (e.g., watershed)
     - sampling event (location id, date, time)
       - i. number of frogs per lifestage counted
       - ii. water temperature, air temperature, presence of fish
5. *Views* – how the data will be viewed or what operations will be performed (e.g., queries to search for out of bound values, summary reports of individuals counted per lake).

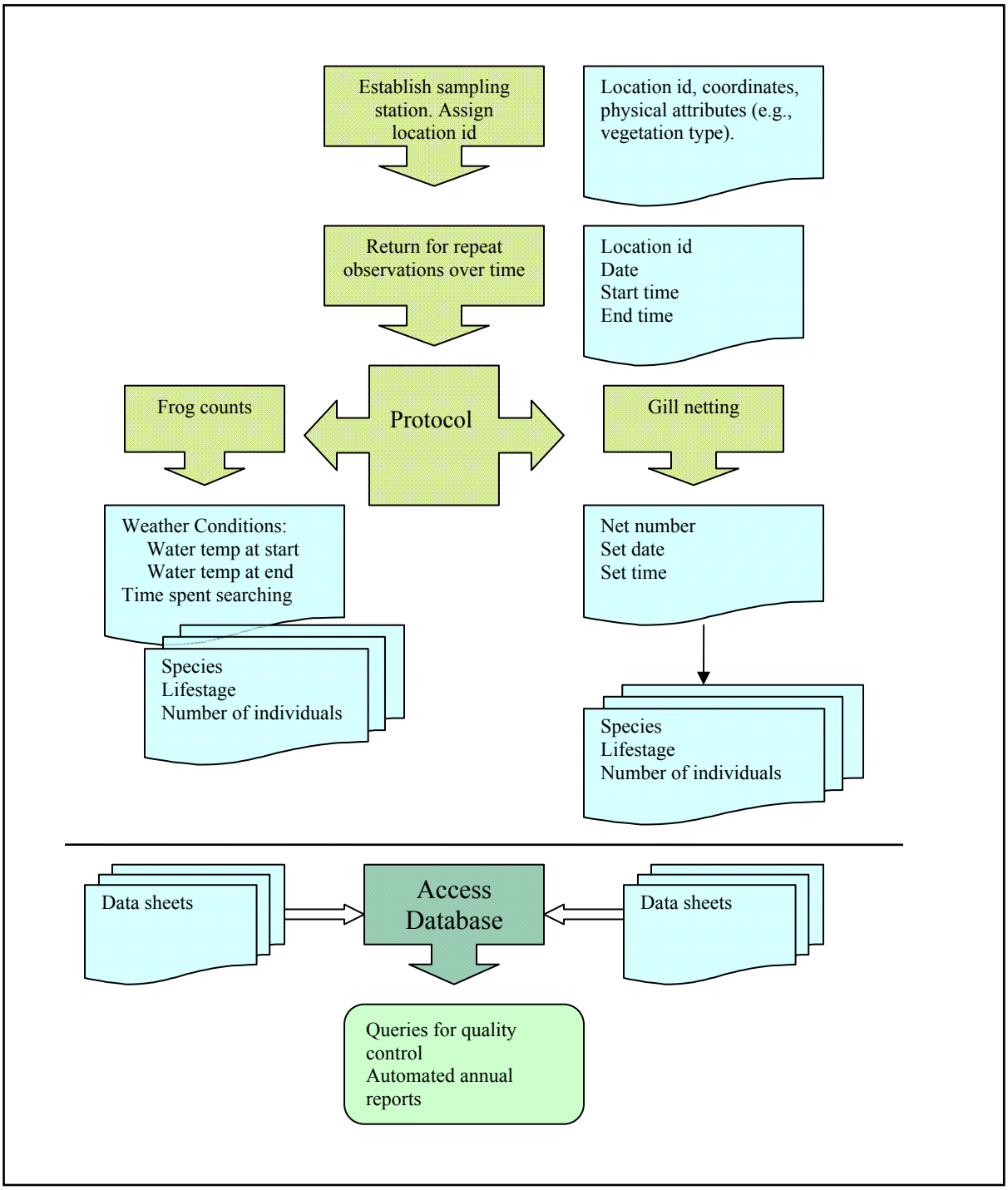

Figure 4. Schematic of a conceptual data model for monitoring the status of mountain yellow-legged frog populations following removal of non-native fish in high mountain lakes in Sequoia and Kings Canyon National Parks.

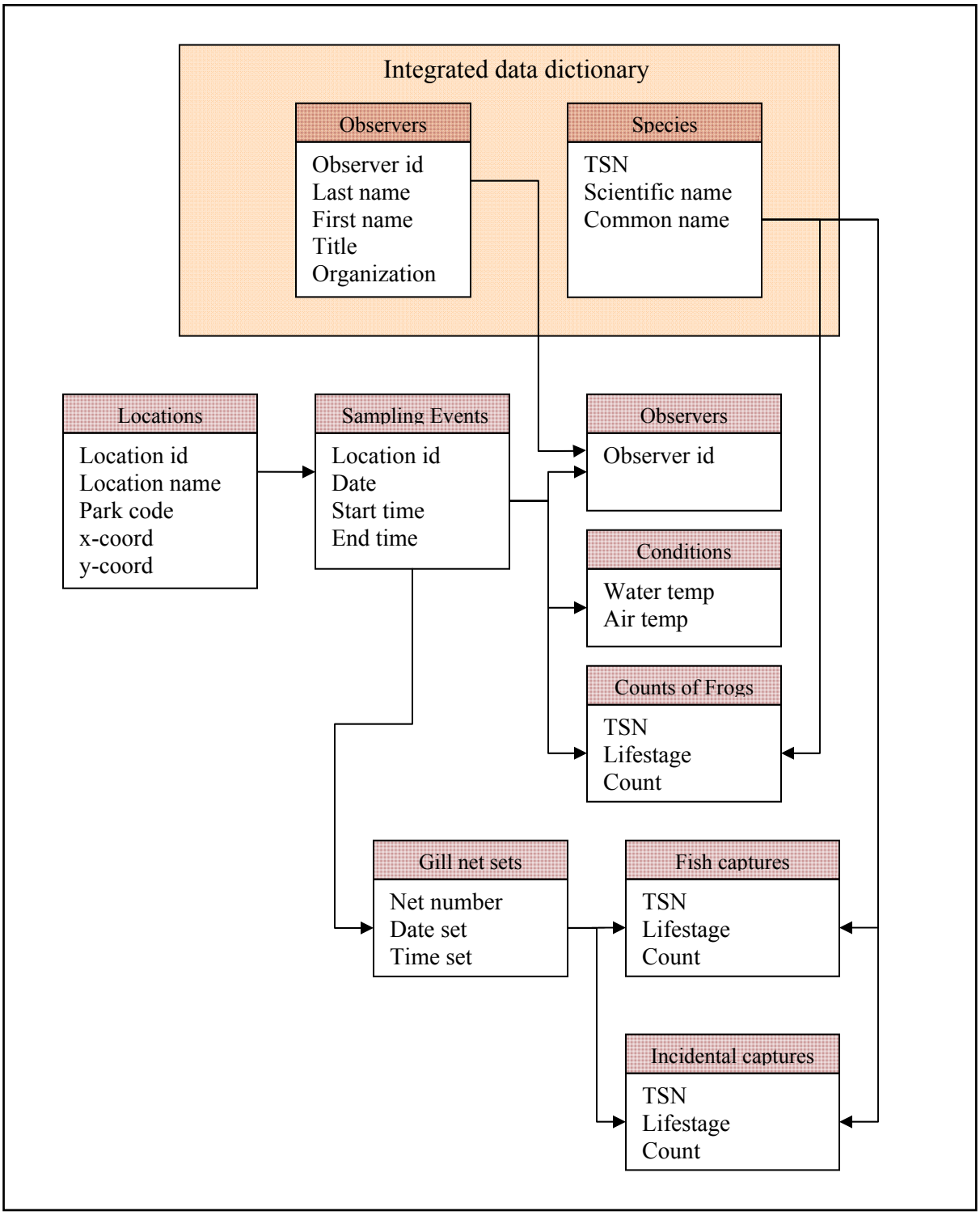

Figure 5. Schematic of a logical data model for monitoring the status of mountain yellow-legged frog populations following removal of non-native fish in high mountain lakes in Sequoia and Kings Canyon National Parks.

SIEN will develop a set of master look-up tables or an Integrated Data Dictionary (also referred to as a Buffet of Fields and Tables [BOFAT] by WASO), that will allow databases across projects and disciplines to remain as consistent as possible. Likely lists of standard attributes and values to be developed include park codes, place names, watershed codes, species identifiers, and vegetation or land type classifications.

5.3.2.3. Physical Data Models: The physical data model depicts the structure of the actual database, with all of its data tables, field definitions, and relationships (Figure 6). Although the logical and physical models might appear similar, the physical model provides enough detail to construct the physical database. The NRDT defines three levels of required or optional tables for the core database:

- Mandatory – necessary components of all databases. Includes: *tbl_Locations* (for sample site location information), *tbl_Events* (sampling events: location, date, and time), *tbl_DB_Meta* (for metadata) and *tbl_DB_Revisions* (revision history)
- Mandatory if Applicable – standardized tables that become part of a database as needed. Includes *tbl_Event_Contacts* (observers and others), *tbl_Event_Group* (for grouping sampling events), and *tbl_Event_Details* (information on who and when the event record was entered)
- Optional tables and fields – contain results of sampling or laboratory analysis and are therefore project-specific. All inventory and monitoring databases will contain one or more of these tables.

The mountain yellow-legged frog database was developed prior to the NRDT table and field naming standards. In the future, SIEN standards for names, fields, tables and other database objects will mirror those conveyed through the NRDT as much as possible.

However, each database must ultimately meet the needs of the network scientists. Considerations for these needs may include interactions with other agencies and ease of use, maintenance, integration, and customization. Where there are differences between local and national standards, documentation of the rationale for these differences will be developed. In addition, documentation and database tools (e.g., queries that rename or reformat data) will be developed to ensure that data exports for integration are in a format compatible with current national standards.

Not shown in Figure 6 are the links to the lookup tables in the Integrated Data Dictionary. This has yet to be created by the SIEN, although the Network already possesses a substantial table of place names which are already being used by several database applications. Tables will also be created with the names of park personnel involved in natural resources projects; vegetation types; wildlife habitat types; sampling methods, and others. Species will be identified by ITIS Taxonomic Serial Numbers, available in the Access file *StdClass_Master.mdb* which is part of the NPSpecies desktop database system, and updated annually from the ITIS database.

To the extent possible, SIEN will work with WASO and other I&M networks to develop and utilize shared look-up tables for data consistency and inter-operability at a higher level within the I&M Program.

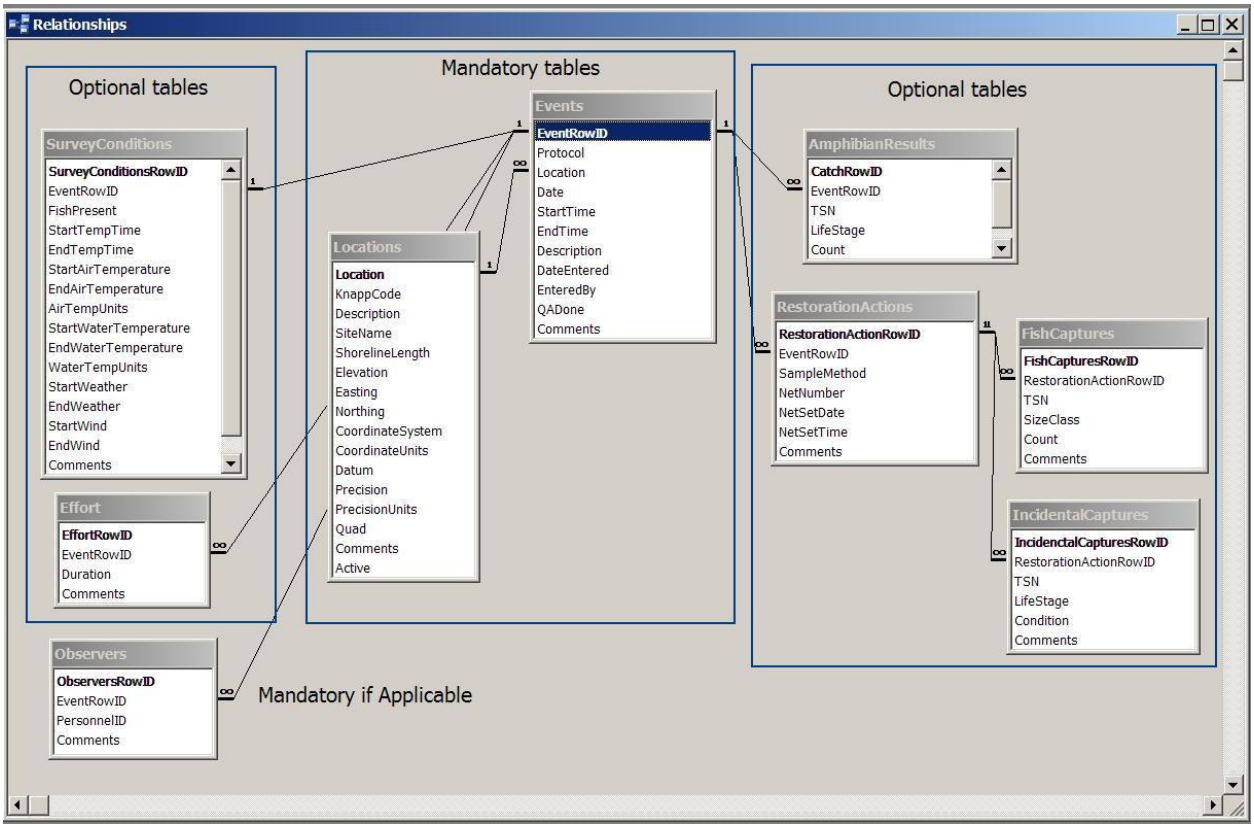

Figure 6. Schematic of the physical database used for monitoring the status of mountain yellow-legged frog populations following removal of non-native fish in high mountain lakes in Sequoia and Kings Canyon National Parks. Note: this diagram does not include data type and length information, or domain specifications, which should be declared as part of a physical database model.

It is especially important to remember that not everything can be thought of from the start and, to support change, data modeling should be iterative and interactive.

5.4. Quality Assurance/Quality Control

The success of any natural resources program is dependent on the quality of the data it collects, manages, and disseminates. The concepts of quality assurance (QA) and quality control (QC) were defined in Chapter 7 and discussed with respect to each phase of a project's lifecycle. Relational Database Management Systems can be designed to incorporate and automate many QA/QC procedures associated with data collection and processing. The following include some of the practices that will help ensure quality data:

- Set unique constraints (or unique key indexes) on every table in a database. Unique constrains are multiple key indexes set to allow no duplication, so as to prevent the entry of the same data record more than once. The index is set on the fields that make the records in a table unique. For child tables, it always includes the foreign key from the

parent table. Duplication of entries is a common problem in databases without unique constraints.

- Implement all value constraints and business rules on the backend of the database, not on the forms used for data entry.

- Create separate forms for data entry/editing and review, or use mode control on single forms with both functions.

- Populate lookup tables as much as possible before data entry begins, and restrict access to them by data entry personnel.

- Set value ranges and validation rules on fields wherever possible to control the range of values that can be entered.

- To the extent possible, set the *limit to list* property to *True* on fields that receive input from a value list or lookup table.

- Include queries or programming modules that run checks for logical errors and out of bound values on calculations.

- Create field forms that best match data entry forms.

- When using hand-held computers in the field for data entry, the same set of controls should be placed on data fields as exist in the master database.

5.5. Database Documentation

5.5.1. Basic Requirements

Relational databases will be documented according to the standards outlined above. Complete documentation will also include a narrative overview with a description of the business rules employed, entity relationship diagrams, and documented programming code. Relational databases will also utilize internal documentation such as table and field descriptions and will include a table to track modifications.

MS Access databases should be documented with the built-in Database Documenter tool, and documentation should include all objects. This will produce a pre-formatted report containing complete specifications of properties, relationship types and attributes, table indexes, and user and group permissions on every object in the database. This information, combined with the programming code would enable any developer to recreate the structure and function of the database.

The NPS Metadata Profile currently does not support all of this information, so it will need to be stored independently of the formal metadata, although, as with the formal metadata, it should be stored in a folder with the database.

Some metadata can now be extracted automatically from an MS Access database. The NPS Database Metadata Extractor MS Access add-in, Version 1.0 for MS Access 2003 2002 2000, automatically harvests entity (table) and attribute (field) metadata, including value ranges (domains) from MS Access databases. It further allows the user to edit and review the harvested metadata, make batch edits, and export the metadata to a FGDC-compliant XML file. Exported XML can be used in the Metadata Tools & Editor either by opening it to start a new metadata record or by updating a template to fill in section 5 of an existing record.

Project databases developed for the I&M Program will include a mandatory table, *tbl_Db_Meta,* which will contain a description of the purpose of the database, and a link to metadata records in the NPS Data Store.

5.5.2 Revisions

Every alteration in a project database that occurs after data collection and entry have begun should be documented. A table of revision history can be included within the database itself or maintained as a separate log. For I&M project databases, the table *tbl_DB_Revisions* is mandatory. This table includes contact person; date, reason, and description of the revision; and a link to *tbl_Db_Meta.*

5.6. References

Beer, M., Nance, E., Wright, A, Powell, M. and R. DenBleyker. 2005. Northern Colorado Plateau Network, Data Management Plan. National Park Service, Inventory and Monitoring Program, Northern Colorado Plateau Network, Moab, Utah. 86 pp. plus appendices.

Boetsch, J. R., B. Christoe, and R. E. Holmes. 2005. Data management plan for the North Coast and Cascades Network Inventory and Monitoring Program. United States National Park Service. Port Angeles, Washington. 88 pp.

5.7. Appendixes

5A) Sierra Nevada Network Database Documentation Template

5.8. Credits

This chapter was written by Rosamonde Cook, with some material adapted from Beer et al. (2005) and Boetsch et al. (2005).

Chapter 6. Acquisition, Processing and Reporting

Large, multi-scale natural resources programs increasingly rely on data and information gathered from multiple sources. This chapter describes the general steps involved with acquiring, processing, and reporting data to meet standards established by the NPS I&M Program, although these standards could apply broadly to any of the programs covered by this Data Management Plan. Also included are guidelines for the acquisition and processing of physical objects (photographs, voucher specimens) which are often collected as part of resource management, inventory and monitoring, and other research projects. Instructions specific to particular projects should be developed and included with the protocols for those projects.

6.1. Objectives

- Acquire, through a variety of sources, data and information needed by SIEN Park managers to properly manage and maintain the natural resources of their parks.

- Acquire data and information needed to understand broad-scale changes in the environment that impact ecosystems on a regional or national level in cooperation with NPS and collaborative monitoring efforts.

6.2. Laws and Policies

Current laws and policies encourage programs to make use of existing (or non-programmatic) spatial data before investing in the acquisition of new data sets. Specifically, Executive Order 12906, Section 3(d) states that each agency must adopt internal procedures to ensure the agency accesses the National Geospatial Data Clearinghouse before it expends federal funds to collect or produce new spatial data, to determine whether the information has already been collected by others or whether cooperative efforts to obtain the data are possible.

As part of any new project, the SIEN will use the National Geospatial Data Clearinghouse and NPS Data Store for searching existing natural resources data set. The SIEN will also use the Internet, agency contacts, and other means to find data that may benefit its programs.

6.3. General Standards and Guidelines

General standards and guidelines for the acquisition and processing of programmatic and non-programmatic data (as defined in Chapter 1) are provided below. Quality assurance and control procedures, which are essential for ensuring the production of quality data and information products, and which comprise critical components of all stages of the data life cycle (Chapter 3), are discussed in greater detail in Chapter 7.

6.3.1. Programmatic Data

Project leaders and data managers are responsible for ensuring data collection, data entry, verification, validation, storage, and archiving are consistent with Network standards. In addition to general Standard Operating Procedures (SOPs) that define Network-wide requirements, protocol-specific SOPs may be developed that detail procedures and/or methodologies. Leads

and cooperators on I&M projects should always receive a copy of Sierra Nevada Network I&M Program Checklist of Project Specifications (Appendix 6A)

6.3.1.1. Data Collection: Listed below are some of the tools available for field data collection. All methods involve some trade-offs in terms of expense, efficiency, and susceptibility to data entry/transcription errors. Project protocols should provide detailed specifications on how the following tools are used with individual projects and some justification for the choice of methods.

Field Forms – the most common method of recording field data. Field forms are inexpensive but require neat, legible handwriting. There generally exists greater opportunities for error during the collection/data entry process compared with other methods because data entered on forms must later be key-entered into project databases. Thus, this method tends to require more data entry time and more rigorous QA/QC.

Field Computers – increase data collection and data entry efficiency. Data can be directly downloaded to office desktops, eliminating the data entry step. QA/QC checks can be built directly into the database, further reducing data entry error and processing time. The drawback to field computers is usually greater expense and sometimes, training time for field crews. There are two types of computers used in the field.

- **Personal Data Assistants (PDAs) –** the small size and relative low cost of these devices make them attractive options for collecting field data. PDAs can be weatherized fairly easily and inexpensively. However, they have limited internal memory and so data must be continually backed-up to larger storage devices. Most run either Windows CE or Palm operating systems which may require additional processing/programming to transfer/create the structure of the master database in the field units. Working with relational database structures can be difficult, or even a practical impossibility if the database is a complex one. In these cases, uploading field data can require great care and special skill.
- **Tablet PCs –** possess the same properties as most laptops and provide the user with the convenience of a touch screen interface. They are bulkier, heavier, more expensive and harder to weatherize than the PDAs, but more powerful as well. They are the best choice for field projects that are very data-intensive or rely on complex data structures. Because they run Windows XP (Tablet Edition) and Microsoft Office software, MS Access databases can be directly transferred from the desktop to field units without additional programming. Data from the field units can in turn, be uploaded to the desktop with minimal effort.

In some cases, the best choice is a combination of field forms and computers, for example, where large amounts of notes or comments need to be recorded in the field.

Automated Data Loggers – are mainly used to collect ambient information such as weather data or water quality information. These units must be properly calibrated and maintained, which requires proper training of field crews and SOPs that outline these procedures. Two types of automated data loggers are available.

- **Permanently deployed devices –** provide continuous or more frequent data (compared to field staff), provide data during conditions unsafe for field staff (i.e., severe storms, flood events), and with proper equipment can provide near real-time data. Data still must be retrieved, although this can be accomplished remotely – via satellite, landline, or cell phone. Providing sufficient and reliable power may be an issue. Automated data-loggers require proximity to AC power, installation of solar power, or regular site visits to change batteries. Wilderness regulations may restrict the type of and location where a device is deployed.
- **Portable hand-held devices –** are deployed for sampling only during site visits.

GPS Units – Two types of GPS units are used during field work in SIEN parks to collect location (geographic coordinates, altitude) information.

- **Recreation Grade Units –** include both self-contained units (e.g., Garmin Map76CSx) and companion units for PDAs (e.g., Holux). These work well for collecting general position information but are not recommended for high accuracy (sub-meter) location information.
- **Survey Grade Units –** such as some Trimble GPS receivers. These are good for collecting accurate sub-meter location information but are expensive.

See Field Data Collection with Global Positioning Systems (Appendix 6B) and Spatial Data Standards (Appendix 6C) for details on methods of collection and processing of spatial data.

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

Photographs – provide an excellent visual record of field visits and are useful for capturing point records of long-term study sites. They also serve well for automated data collection by remotely recording information using web cams or trip cameras. Digital photo management will follow the procedures outlined in Sierra Nevada Network Digital Photo Management Strategy (Appendix 6D). Slides and photographs should be stored and archived according to guidelines outlined in Chapter 11.

Remotely Sensed Data – includes satellite imagery and aerial photography. Remote sensing can be a powerful tool for characterizing and analyzing landscape data, as well as readily capturing data within areas of low accessibility. Considerations for selecting remote sensing imagery are as follows:

- Accuracy and resolution needed
- Frequency of measurement
- Costs
- Licensing for public use
- Ortho-rectification standards

Each remote sensing product is unique. It is imperative that users fully understand the product being used and that products be accompanied by well-documented metadata. Any projects involving remote sensing should include consultation with a professional remote-sensing specialist, and project managers should consider the trade-offs between accuracy and cost among different imagery sources (Table 3).

The SIEN GIS library currently includes satellite imagery coverage for most of the parks, along with base cartographic layers, digital orthophoto quadrangles, and natural resource theme layers.

Table 3. Satellite resolution, swath width, area, cost, and history of various remote-sensing imagery sources (sorted by cost)

Sensor	Multi-spectral Resolution (m)	Panchromatic Resolution (m)	Swath (km)	Area (sq km)	Cost/sq km ($)*	History
Quickbird [T]	2.4	0.7	16	272	29.73	2001-present
Ikonos [T]	4	1.0	11	121	27.03	1999-present
Spot5 [T]	10	2.5 and 5	60	3599	0.77	2002-present
Spot	66	10	60	3599	0.05**	1986-present
ETM+	30	15	185	34221	0.03	1999-2003
TM	30	NA	185	34221	0.03	1982-1999
ASTER	15	NA	60	3599	0.02	1999-present
MSS (ESTS)	79	NA	185	34221	NA	1972-1997
AVHRR	1100	NA	2700	7290138	NA	1978-present
MODIS	250	NA	2330	5429053	NA	1999-present

* cost uses the multi-spectral unit specifications
** cost for post-1998 imagery (pre-1998 imagery is half-price)
[T] requires tasking

Any additional remote sensing data acquired will likely require some spatial or spectral processing depending on how they are received. All such data sets should be received in a geo-referenced format although some may need to be transformed to the SIEN standard (UTM NAD 83). The extent of any additional processing will depend on how the data are to be used and what

type of information is to be extracted from them. Specific requirements and procedures should be outlined in project protocols or SOPs.

6.3.1.2. Data Processing: Ideally, each project will have a database developed prior to the collection of any data in the field. SIEN I&M program databases will be developed in conjunction with project protocols and will be based on the I&M Natural Resources Database Template (NRDT) (Chapter 5) and will include built-in procedures for QA/QC. Ideally, data processing should proceed as follows, and as soon as possible after data are collected, if using field forms.

- Field crews enter all data into an approved project database, under the supervision of the project manager.
- Field crews periodically forward project data files to the project leader and/or data manager (refer to individual protocols for specific requirements).
- All data undergo QA/QC procedures (see Chapter 7 for more specifics on data verification and validation).
- The data manager maintains the master copy of the database and updates it with certified data files received from the project leader.
- National databases are updated as per the procedures outlined in Chapter 10.

6.3.1.2.1. Spatial data: See Field Data Collection with Global Positioning Systems (Appendix 6B) and Spatial Data Standards (Appendix 6C) for complete details on methods for data collection and processing of spatial data.

6.3.1.2.2. Photographs: Photos taken as part of a project's data collection protocol constitute data and need to be organized, documented and preserved in conjunction with all other project data. The processing, documentation, and storage of photographs, including digital photos is detailed in Appendix 6D. In general, the level of processing required will depend on the purpose of the photograph. Photos are regularly used by many park programs such as resource management, maintenance, fire, and research.

Editing of digital photos may be done to improve orientation or correct for lighting conditions (e.g., rotation to best orientation and removal of 'red eye'), but should never include alterations that change the original content of the photo. Photos may be cropped to remove edge areas that grossly distract from the subject. Poor quality photos can be deleted or destroyed, except where the subject is unique. Photos of medium quality should be assessed against existing photos of the same subject; if they duplicate a subject with no enhancement of quality or perspective, they may be deleted or destroyed. Working photos should be stored either in a workspace within a specific project or within a user's photo library. Digital aerial photos and imagery should be georeferenced. All digital data photos should have geographic coordinates associated with them.

6.3.1.2.3. Remote Sensing Data: The processing of remotely-sensed data is often project specific. Therefore, project plans, protocols, and SOPs should document these procedures. Some steps will be common to all data sets, including:

- Images are geometrically registered using nearest-neighbor resampling methods and co-registered to UTM NAD83. Registration accuracy is assessed.
- Images are radiometrically corrected and converted to exo-atmospheric reflectance.
- Atmospheric corrections are made, if applicable.
- All iterations (raw, intermediate corrections, and final) are maintained until project completion when raw and final products are archived.

6.3.1.2.4. Vouchers: The acquisition and processing of biological specimen vouchers will be guided by the policies outlined in Sierra Nevada Network Inventory and Monitoring Program Voucher Collection and Documentation Standards (Appendix 6E).

Project leaders will provide parks with material and data, in MS Excel format or comma-delimited ASCII files (.csv) for automated uploading into ANCS+ (Appendix 6F) and input to NPSpecies (Appendix 6G). Data provided to non-NPS curators will be in MS Excel format as well. Specimens are owned by the parks in which they were collected and are curated by park staff unless agreements have been made with outside institutions. See Chapter 11 for more details.

6.3.2. Non-programmatic Data

A large amount of the data used to manage the natural resources of SIEN parks is collected by entities outside of the Network (universities and other parks, agencies, and NPS programs). These kinds of data collection efforts are referred to as *data discovery* or *data mini*ng. Data collected and products produced by such efforts provide a great deal of information about park natural resources and are therefore relevant to many of its programs. These data can be classified as follows:

> *Current or ongoing* – These are data that are currently being acquired or were recently acquired, but are known to be of value and are targeted for incorporation into existing or planned Network projects. Their acquisition and use should follow very specific guidelines identified by project protocols. Sources can be other Network park programs or external NPS sources. These data might be used to complement program data collection, to fill in gaps of missing data, or as a basis for comparison.
>
> *Legacy* - Those data found and compiled through the data mining process. These may include vertebrate and vascular plant species data, other important natural resource inventory data, specimen or voucher data, bibliographic data, and existing monitoring datasets.

Data discovery is the process of locating and identifying these useful data sets. Non-programmatic data can be obtained from academia, private organizations and non-profit groups, as well as local, state, and federal government agencies. Within the NPS, park, regional and national programs are all potential sources of information.

Parks within the SIEN may use base funds or receive funding through Natural Resources Protection and Preservation (NRPP) programs to support park-level projects. Typically, these include:

- **Park-based biological inventories** - Network parks often conduct their own park-based inventory projects.
- **Park-based monitoring projects** - Parks also engage in monitoring (such as vegetation or water quality) which can produce information valuable to other programs, such as I&M.
- **Park and multi-park based projects** - include other studies or projects conducted at the park or regional level that do not fall into one of the previous two categories (e.g., restoration projects).

NPS regional and national programs support all parks within the Pacific West Region and are good resources for natural resources information. These include:

- **Air –** National-level programs collect data, maintain databases, assure data quality, and perform the trend analyses relevant to SIEN air quality issues.
- **EPMT -** Exotic Plant Management Teams (EPMT) collect and maintain data regarding the presence of exotic species in many parks, and develop and document the methods used to treat these species. Data are stored in the Alien Plant Control and Monitoring Database (APCAM) and maintained by the EPMT data manager.
- **Fire Program –** Data on fire occurrence within the SIEN are maintained both at individual park and regional levels. National databases such as Fire-Pro, SACS and Fire Program Analysis (FPA) have been, and will be, used to maintain information regarding fire incidence and the resources dedicated to fire management.
- **GIS –** The SIEN is supported by regional GIS specialists to help ensure regional GIS data are available and accurate. Much of these data are also available through the NPS GIS Clearinghouse.
- **Wildlife Management -** The regional Threatened and Endangered (T&E) Species Specialist provides management support and coordinates reporting of T&E species populations.

Additional data sources are listed in Table 4 and Appendix G of Appendix 10B.

Future data mining efforts within the SIEN should adhere to the guidelines and protocols outlined in Sierra Nevada Network Guidance for Data Mining and Discovery (Appendix 6H). Information collected during the data discovery process should be maintained either electronically or in hard copy format depending on how it was collected, and should be documented as fully as possible (Chapter 8). Data should be disseminated to the appropriate repository as summarized in Table 4 and discussed in Chapter 10.

Data discovery is an integral part of project development, but efforts should not be limited solely to project development needs. This should be an ongoing process requiring regular data searches and visits to Network parks to ensure that Network parks maintain as much material relevant to

Table 4. Summary of possible data sources for different types of information and repositories where they are maintained.

Type of Data	Possible Source	Repository
Bibliographic / Literature	• Online literature databases (e.g., First Search or Biosis) • Library catalogs (e.g., academic or research institutions) • Park archives through ANCS+	• NatureBib • Reference cabinets for hard copy materials • Digital archive for electronic materials
Geographic Data	• Regional centralized GIS data • Federal and state geographic data clearinghouses • Local, state, and federal government offices • Regional and park GIS specialists	• NPS Data Store • Digital archive
Biologic / Natural Resource Data	• Voucher collections (museums, parks, universities) • Network parks	• NPSpecies • Digital archive

managing their natural resources as possible. Encouraging data sharing among parks will assist in this process and may alleviate the need for regular searches of park records.

6.3.2.1. Data Collection: The collection of data from non-programmatic sources should follow program specifications as outlined by project protocols or SOPs. Procedures should be standardized as much as possible and include:

- Contacting data stewards and informing them of program needs.
- Establishing Memoranda of Understanding (MOUs) if needed.
- Developing a contingency plan in case the data source is no longer available.
- Determining whether data can be consistently exported/imported and establishing a schedule.
- Determining how the data will be stored and integrated into the program.
- Determining how errors will be addressed.
- Determining if documentation is adequate and if not, completing where necessary.
- Identifying any interest in the exchange of program data and information with the outside program.

Agency or organizational stewards of these data often have the expertise to conduct proper quality control procedures and the capability to function as a repository and clearinghouse for validated data. In some cases, portions of external databases may be incorporated into SIEN databases and thereby made more accessible to staff.

6.3.2.2. Data Processing: Much of the data identified during the discovery process are likely to be the legacy type. As time and resources permit, legacy data should be converted to file formats compatible with current software standards. Hardcopy references and other materials containing legacy data can be scanned and saved as .pdf files and stored in a program's digital library (Chapter 11). All legacy data sets should be reviewed and cataloged as follows:

- Enter all biodiversity data into NPSpecies (this is especially important for park-based biological inventories).
- Enter all natural resource reports and publications related to SIEN parks into NatureBib. Hard copies should be stored in the appropriate park collections and electronic copies archived in the proper directory on the Network file servers (Chapter 11).

All GIS data should be stored in the proper projection (UTM NAD83) and accompanied by FGDC-compliant metadata (Chapter 8).

6.3.3. Analysis and Reporting

Obtaining meaningful results from data summary and analysis is essential to providing useful information for natural resource managers and scientists. Thus it is incumbent on data managers and stewards to provide valid data in formats that support scheduled and ad hoc display, query, analysis, summary, and reporting. Routine and scheduled data summary, analysis and reporting requirements and procedures should be identified in project protocols.

The development of data products should be guided by project objectives, protocols and data management SOPs. Products defined in advance, such as routine summaries and output formats can be automated through the use of queries and reports stored in a project's database. Project leaders should work with data managers to create, and test these queries. Queries can be used to view raw or summarized data, or to output either type as custom-formatted reports or as files that can be imported by other analytical tools (e.g., statistics software). Other queries can be built to facilitate data exploration and unscheduled analysis. Specific needs should be determined by individual projects.

Most likely, data will be exported from project databases for most statistical analyses beyond means, standard deviations, and other descriptive statistics. SIEN will use third party statistical software (i.e., R, SAS, and SPSS) for generating frequency distribution plots, tests for normality analysis of variance, time series analysis, and others.

6.3.4. Changes to Procedures

Changes to established data collection procedures are discouraged unless there are acceptable, valid reasons for altering the methodologies. Ideally, all problems should be identified during the design and testing stages of the project and changes implemented prior to the collection of any field data. Protocols should attempt to identify any foreseeable issues that might occur as well as contingencies to address them. Inevitably, unforeseen problems may occur which require procedure/protocol revision after data collection has begun. Significant changes to protocols should be approved by the principal investigator, project leader and data manager. The key official should evaluate the proposed changes and determine if additional peer review is required

before accepting them. All changes should be carefully documented within project SOPs and any associated databases.

6.3.5. Maintaining an Edit Log

Any changes that occur as a result of a change in a project's protocol should be documented in the formal documentation developed for the database (Chapter 5).

6.4. References

Boetsch, J. R., B. Christoe, and R. E. Holmes. 2005. Data management plan for the North Coast and Cascades Network Inventory and Monitoring Program. United States National Park Service. Port Angeles, Washington. 88 pp.

Daley, R. 2005. Data and Information Management Plan, Greater Yellowstone Network Inventory and Monitoring Network. United States National Park Service, Bozeman, Montana. 82 pp.

Mortenson, D. 2005. Data Management Plan for the Inventory and Monitoring Program, Southwest Alaska Network. United States National Park Service. Anchorage, Alaska. 234 pp.

Stevens, S. M. and G. Entsminger. 2004. Information Management Plan for the Inventory and Monitoring Program, Northeast Coastal and Barrier Network. United States National Park Service. Kingston, Rhode Island. 99 pp.

6.5. Appendixes

6A) Sierra Nevada Network I&M Program Checklist of Project Specifications
6B) Field Data Collection with Global Positioning Systems
6C) Spatial Data Standards
6D) Sierra Nevada Digital Photo Management Strategy
6E) Sierra Nevada Network Inventory and Monitoring Program Voucher Collection and Documentation Standards
6F) Data Dictionary and Formatting Specifications for ANCS+ Voucher Records
6G) Field Definitions for NPSpecies Voucher and Observation Records
6H) Sierra Nevada Network Guidance for Data Mining and Discovery
10B) NatureBib Standards Manual

6.6. Credits

This chapter was written by Rosamonde Cook and Leona Svancara, adapted from concepts and material developed by Stevens and Entsminger (2004), Boetsch et al. (2005), Daley (2005), and Mortenson (2005).

Chapter 7. Quality Assurance and Quality Control (QA/QC)

The success of any natural resources program is dependent on the quality of the data it collects, manages, and disseminates. Analyses performed to detect ecological trends or patterns require data that are recorded properly and have acceptable precision and minimal bias. Poor quality data can limit detection of subtle changes in patterns and processes, can lead to incorrect interpretations and conclusions, and could greatly compromise the credibility of the program managing it.

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

7.1. Objectives

- Ensure natural resources projects produce high quality and credible data that can be confidently used by managers, researchers and the public
- Implement standard quality assurance and quality control procedures to meet the first objective

7.2. Laws and Policies

Director's Order #11B, "Ensuring Quality of Information Disseminated by the National Park Service," was issued in order to promote data quality (Appendix 7A). It defines 'quality' as incorporating three key components: objectivity, utility, and integrity.

> *Objectivity* consists of: 1) presentation: whether disseminated information is being presented in an accurate, clear, complete, and unbiased manner within a proper context; and 2) substance: ensuring accurate, usable, and reliable information.
>
> *Utility* refers to the usefulness of the information to its intended users.
>
> *Integrity* refers to the security of information, meaning protection from unauthorized access or revision is ensured that information is not compromised through corruption or falsification.

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

The SIEN holds as policy that techniques and procedures chosen for a project will maximize data quality.

7.3. General Standards and Guidelines

To ensure that the SIEN produces and maintains data of the highest possible quality, procedures have been established to identify and minimize errors at each project stage associated with the data life cycle (Figure 7). QA/QC procedures specific to any project should be specified in a project's protocols and SOPs. However, some general concepts apply to all Network projects. The general guidelines presented below were primarily adapted from the I&M Program's *Draft Data Management Protocol* (Tessler & Gregson 1997) and ideas contained in Michener and Brunt (2000).

Although a data set containing no errors would be ideal, the cost of attaining 95%-100% accuracy may outweigh the benefit. Therefore, at least two factors are considered when setting data quality expectations:

- frequency of incorrect data fields or records, and
- significance of error within a data field.

Errors are more likely to be detected when data sets are clearly documented and what constitutes a 'significant' error within *that* data set is understood. The significance of an error can vary both among data sets and within a single data set. For example, a two-digit number with a misplaced decimal point (e.g., 99 vs. 9.9) may be a significant error while a six-digit number with an incorrect decimal value (e.g., 9999.99 vs. 9999.98), may not. However, one incorrect digit in a six-digit species' Taxonomic Serial Number could indicate a different species. QA/QC mechanisms are designed to prevent data contamination, which occurs when a process or event introduces either of two fundamental types of errors into a data set:

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

Selected QA/QC procedures relative to the amount of planning and quality control necessary to have confidence in the data are illustrated in Figure 8. The most effective mechanism for ensuring that a project produces high-quality data is to determine procedures that direct project staff through accurate data collection, entry, and validation, and adhere to them.

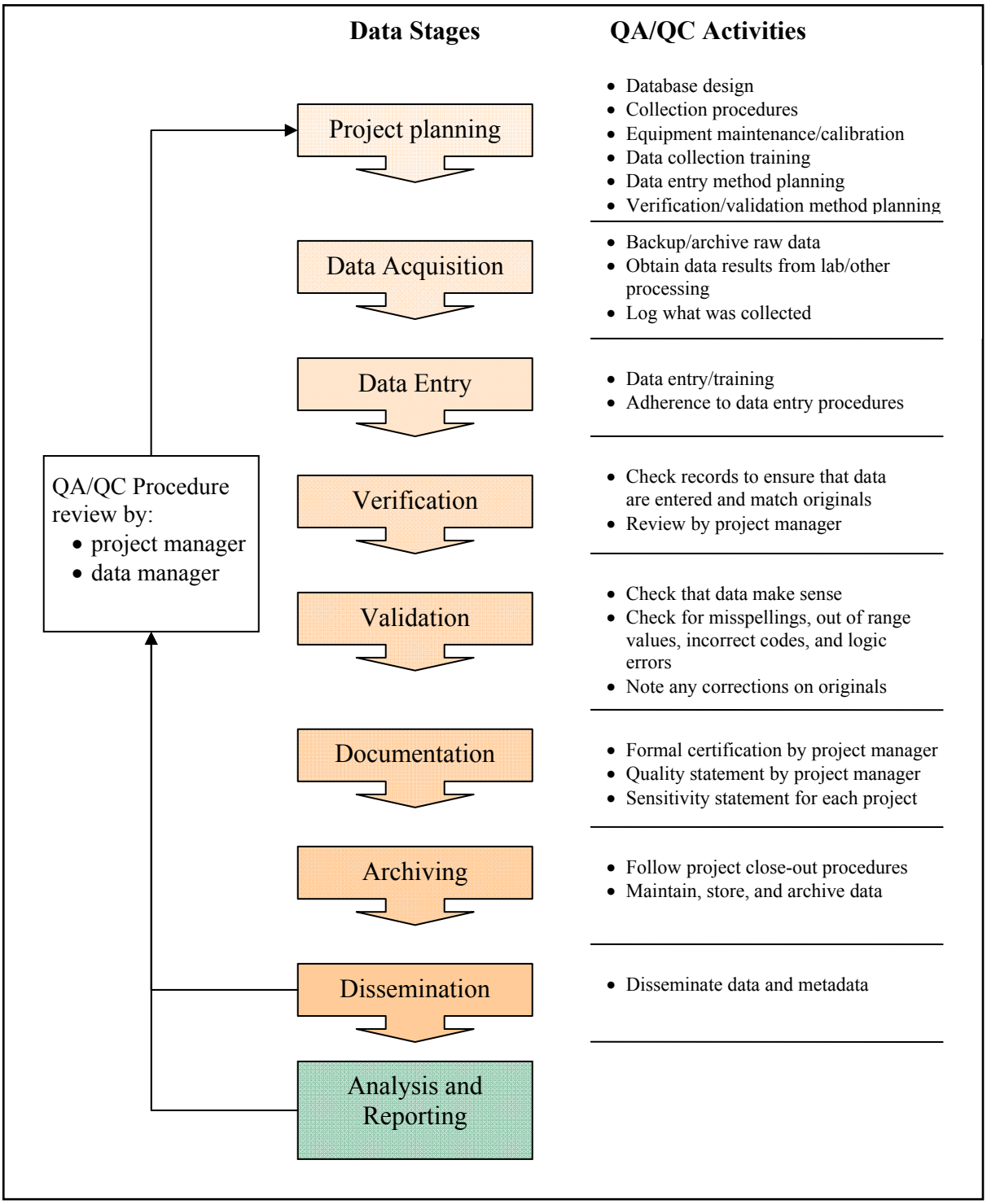

Figure 7. General course of data and associated Quality Assurance/Quality Control procedures. Quality Control with regards to data analysis is specific to each project and addressed in appropriate standard operating procedures.

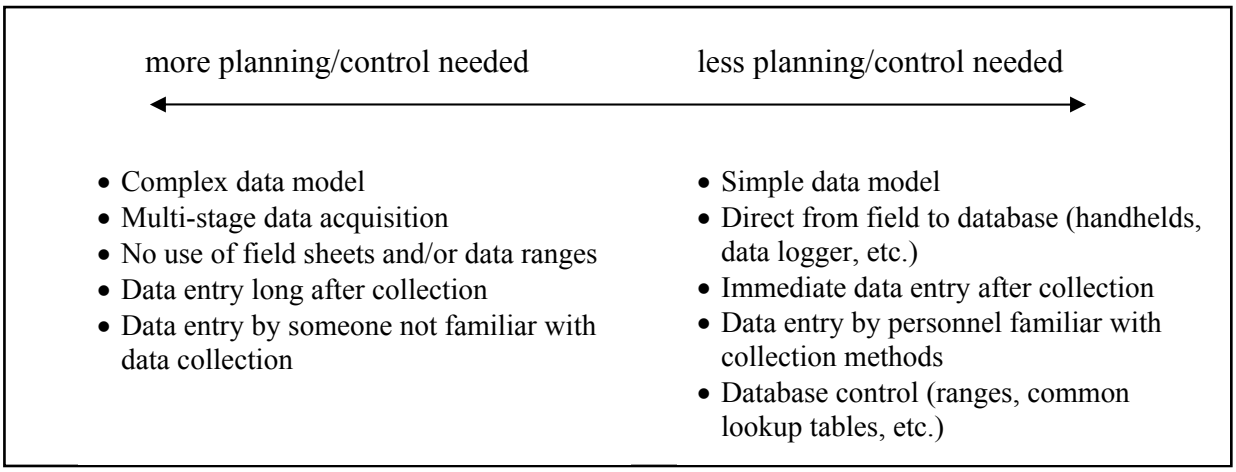

Figure 8. Some common information management elements influencing the amount of QA/QC needed.

7.3.1. Project Planning and Data Design

Quality assurance begins with choice of project methods and information to be gathered. Quality assurance will be achieved by stipulating that:

- Common lookup tables are created for parameter values recorded by multiple projects (e.g., weather variables, standard equipment, field personnel, etc).
- Each vital sign protocol has SOPs that address core information management practices (e.g., field crew training, use of handheld computers, equipment maintenance, and data backup, entry, verification, and validation, etc.).
- Each project uses handheld computers for data collection if possible. If not, at a minimum, standardized field sheets are used.
- Databases adhere to the standards outlined in this plan (Chapter 5)
- Database data entry forms, if necessary, resemble the field sheets.
- Automated error checking features will be included in database applications.
- Database application design will maximize the use of auto-fill, auto-correct, range limits, pick lists, and other constraints specific to projects.
- Database applications will include a means to track the date a record is created or modified, the name of the person creating/modifying the record, and errors reported on the data after dissemination.
- Database maintenance logs will be maintained for each SIEN database and housed in association with database files.

7.3.2. Data Collection

Chapter 6 addressed general data acquisition and initial handling as well as changes to data collection protocols. Attention to detail during the data collection phase, however, is crucial to overall data quality. Unlike a typographical error that occurs during data entry, an incorrect observation in the field is not easily corrected.

The SIEN adopts the following guidelines regarding data collection that affect data quality:

- Field crews will receive proper training.
- Any project using field equipment will include a calibration and maintenance SOP that will also specify establishment of an equipment maintenance log.
- Wherever possible and appropriate, data loggers or field-based computers will be used to collect data. When this is not possible, data will be recorded on formatted, project-specific data sheets that reflect the overall design of the project and are designed to minimize the amount of writing necessary to effectively record observations.
- When field sheets are necessary, or more appropriate, the format will reflect the computer data entry interface to help ensure all relevant information is recorded and subsequent data entry errors are minimized.
- Completed field forms will be proofed for errors each day in the field. Edits will be made by drawing a single line through the information to be changed, adding any replacement information in clear writing next to the original entry, and initializing the changes.
- Separate SOPs will describe the use of data loggers or field-based computers and will include direction for daily review and back-up.

7.3.3. Data Entry

Data entry is the process whereby raw data are transferred from paper field forms into an electronic data format. When data are gathered or stored digitally in the field (e.g., on a data logger), data entry consists of the transfer of data (downloading) to a file 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, although errors are unavoidable. 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 SIEN adopts the following guidelines regarding data entry:

- To the extent possible, data entry will be automated. This may simply entail downloading data from field-based computers but may include the application of new technology to allow for machine-driven data entry (e.g., voice-recognition software).
- Data will be entered as soon as reasonably possible after collection.
- Data entry will be completed by someone familiar with data collection. The project leader (with assistance from the data manager if needed) must ensure that data entry staff are familiar with the database software, database structure, and any standard codes used by the Network. At a minimum, data entry technicians should know how to open a data entry form, create a new record, and exit the database properly. They must also learn how to correct mistakes made while typing.
- If feasible, data will be entered by two qualified persons; one person will read the observations and the other will enter the data.
- Data will be entered into pre-designed database forms that resemble field sheets and maximize error control.

7.3.4. Data Verification

Data quality is appraised 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 (i.e., planning). *Data verification* checks that the digitized data match the source data, while *data validation* checks that the data make sense. Although data entry and verification can be handled by personnel who are less familiar with the data, validation requires in-depth knowledge about the data.

The SIEN adopts the following guidelines regarding data verification:

- Project leaders are responsible for specifying in the project protocol one or more of the data verification methods available and ensuring proper execution. At the discretion of the project leader, additional verification methods may be applied.
- Data verification should be carried out by staff familiar with data collection and entry.
- All records (100%) will be verified against original source data.
- A subset of randomly selected records (10%) will be reviewed after initial verification by the project leader. If errors are found, the entire data set will be verified again.
- A record of the verification process for each dataset, including number of iterations and results, will be prepared by the project leader as part of formal metadata generation (see Chapter 8 for more details).
- Spatial data collected as part of the project will be viewed in a GIS and visually inspected for accuracy (e.g., points located outside park boundaries, upland locations occurring in water).

7.3.4.1. 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 and immediately corrects any errors. 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. Upon completion of data entry, all records are printed and compared with the original values from the hard copy. Errors are clearly marked and corrected in the database as soon after data entry as possible. Reliability increases if someone other than the person keying the data performs the review. Alternatively, two technicians (one reading from the original data and one checking the entered data) can perform this review.

3) Duplicate data entry. The data entry person completes all data entry as normal. Random records are selected (e.g., every *nth* 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 on any mismatches. Disparities are manually reviewed and corrected 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.

Data verification can be improved by calculating summary statistics and identifying duplicate or omitted records. For example, the number of known constant elements, such as the number of sampling sites, plots per site, or dates per sample can be evaluated. Databases can also be built with controls to prevent duplicate records (Chapter 5), although tests must be performed to identify missing records. The more checks that are devised to test for completeness, the greater will be one's confidence in the quality of the data set data.

7.3.5. 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. It is essential that we validate all data as truthful and do not misrepresent the circumstances and limitations of collection.

General step-by-step instructions are not possible for data validation because each data set has unique measurement ranges, sampling precision, and accuracy. Specific guidelines should be written into all project protocols and SOPS. Invalid data commonly consist of misspelled species names or site codes, wrong dates, or out-of-range errors in parameters with well defined limits (e.g., pH). More interesting and often puzzling errors are detected as unreasonable metrics (e.g., stream temperature of 70°C) or impossible associations (e.g., a tree 2 feet in diameter and only 3 feet high). These types of erroneous data are called *logic errors* because they produce 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 SIEN adopts the following guidelines regarding data validation:

- Project protocols will address a process for data validation that includes at least one of the available methods.
- 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, with the editor's initials.
- Modifications of the field data will be clear and concise while preserving the original data entries or notes (i.e., no erasing).
- Validation efforts will also include a check for the completeness of a data set since field sheets or other sources of data could easily be overlooked.
- Use of automated routines and/or data summary and visualization (e.g., histograms, line plots, and basic statistics) will be maximized to identify possible logic and range errors.
- Use of database programming will be maximized to control data entry. This will be achieved via the use of lookup tables and/or field-type design in a database (e.g., yes/no field-types).

7.3.6. Methods for Data Validation

The following general methods can be used as guidelines:

1) *Data entry application programming.* Certain components of data validation are built into data entry forms. This method is essentially part of the database design and is discussed earlier in this chapter. Not all fields, however, have appropriate ranges known in advance. 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 (see Chapter 5 for more detail).

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' extreme.

Data quality assurance procedures should not try to *eliminate* outliers. Extreme values naturally occur in many ecological phenomena; eliminating these values simply because they are extreme is equivalent to pretending the phenomenon is 'well-behaved' when it is not. Eliminating data contamination is perhaps a better way to explain this quality assurance goal. When an outlier is detected (via GIS, database, graphic, and statistical tools for ad-hoc queries and displays), the possibility of contamination will be evaluated and noted.

3) *Other exploratory data analyses.* Palmer and Landis (2002) suggest calculations for assessments of precision, bias, representativeness, completeness, and comparability may be applicable and, for certain types of measurements, evaluation of detection limits may also be warranted. Normal probability plots, Grubb's test, and simple and multiple linear regression techniques may also be used (Edwards 2000).

7.3.7. Review, Conformance and Communication

The National Park Service requires QA/QC review and approval prior to communicating or disseminating data and information. 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 the related data and information. Mechanisms must also be in place for receiving and addressing comments or complaints pertaining to data quality (see also Chapter 10).

As part of the close-out and evaluation stage of each SIEN project, QA/QC procedures will be reviewed by the project leader and recommendations for change will be included in the annual report. Similarly, SIEN data management staff will review and revise the QA/QC procedures included in this information management plan and/or SOPs as needed.

To ensure the highest quality, data custodians should conduct periodic audits to ensure compliance with the information management plan and protocol QA/QC procedures. Such

quality checks promote a cyclic process of continuous feedback and improvement of both the data and quality planning process. Audits 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

The final step in the QA/QC process is preparation of summary documentation that assesses overall data quality. The statement of data quality is composed by the project leader and incorporated into formal metadata for the data set. Metadata should also provide information on the specific QA/QC procedures applied and the results of review. Typically, data quality information will be conveyed as part of FGDC-compliant metadata (Chapter 8) and will be available via the NPS Data Store (Chapter 10).

7.3.8. Roles and Responsibilities

Producing and maintaining high quality data is the responsibility of everyone involved with the handling of project data. It is essential that each member of the team have a stake in data quality, and is responsible for the quality of the results generated from his or her tasks. While Chapter 4 discusses data management roles and responsibilities, selected QA/QC duties are emphasized here.

Project leaders need to:

- be aware of QA/QC procedures in 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.

Project technicians must:

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

The data manager is responsible for:

- developing Network-wide SOPs to ensure data quality
- making project leaders, technicians, and others involved 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 quality control operations.

7.4. References

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

Beer, M., Nance, E., Wright, A, Powell, M. and R. DenBleyker. 2005. Northern Colorado Plateau Network, Data Management Plan. National Park Service, Inventory and Monitoring Program, Northern Colorado Plateau Network, Moab, Utah. 86 pp. plus appendices.

Edwards, D. 2000. Data quality assurance. Pages 70-91 in Michener, W. K., and J. W. Brunt (eds), Ecological Data: Design, Management and Processing, Methods in Ecology Series. Blackwell Science Ltd., Malden, Massachusetts.

Hart, M. and U. Gafvert. 2005. Data Management Plan: Great Lakes Inventory and Monitoring Network. National Park Service Great Lakes Inventory and Monitoring Network Report GLKN/2005/20.

Michener, W. K., J. W. Brunt, J. J. Helly, T. B. Kirchner, and S. G. Stafford. 1997. Nongeospatial metadata for the ecological sciences. Ecological Applications 7:1, 330- 342.

Palmer, C. J. 2003. Approaches to quality assurance and information management for regional ecological monitoring programs. Pp. 211-225 in Busch, D. E. and J. C. Trexler (eds) Monitoring Ecosystems: Interdisciplinary Approaches for Evaluating Ecoregional Initiatives. Island Press, Washington, DC.

Palmer, C. J., and E. B. Landis. 2002. Draft. Lake Mead National Recreational Area Resource Management Division: Quality System Management Plan for Environmental Data Collection Projects.

Tessler, S., and J. Gregson. 1997. Draft Data Management Protocol. NPS I&M Program. http://www1.nrintra.nps.gov/im/dmproto/joe40001.htm.

7.5. Appendixes

7A) Director's Order #11B: Ensuring Quality of Information Disseminated by the National Park Service

7.6. Credits

This chapter was written by Rosamonde Cook and Leona Svancara, with concepts and material developed by Angell (2005), Beer et al. (2005), and Hart and Gafvert (2005).

Chapter 8. Dataset Documentation

Documenting data is a time-consuming task for NPS staff and cooperators. Thorough documentation however, is essential for preserving the integrity and longevity of data and the products of its analysis, and is therefore, an essential component of sound data management.

8.1. Objectives

- Document all significant geographic (GIS) and tabular datasets to SIEN standards described in this chapter.
- Maintain and leverage the investment made by NPS staff and cooperators in producing quality data.
- Aid the discovery of relevant data by NPS staff, collaborators, and the public.
- Preserve integrity and longevity of data and its documentation indefinitely.

8.2. Laws and Policies

Executive Order 12906 (Appendix 8A) requires all federal agencies to fully document all spatial data collected or produced, either directly or indirectly, since 1995 with metadata that meet specific standards developed by the Federal Geographic Data Committee (FGDC) as outlined by its *Content Standard for Digital Geospatial Metadata* (CSDGM) (Federal Geographic Data Committee 1994). Executive Order 12906 also directs agencies to plan for legacy data documentation and provide metadata and data to the public.

The NPS GIS Committee requires all new spatial data sets be described according to FGDC standards and the NPS Metadata Profile. The *NPS Metadata Profile* (http://science.nature.nps.gov/nrdata/docs/npsprofile.cfm) extends the original seven sections of the CSDGM to include the *FGDC Biological Data Profile, ESRI Metadata Profile (ESRI 2003), and the NPS Profile.* Profiles are simply formatted sections of information that are included in a formal metadata document. The NPS Metadata Profile is utilized by the Natural Resource and NPS Data Store for managing GIS and other data for both internal use and publication to the NPS GIS Clearinghouse within the NPS Focus Digital Library and Research Station.

The *FGDC Biological Data Profile* (FGDC Biological Data Working Group, and USGS Biological Resources Division 1999) contains elements (i.e., specific information sections) for describing biological data sets (both spatial and non-spatial). Metadata that contain these elements can be disseminated to the National Biological Information Infrastructure (NBII, www.nbii.gov/datainfo/metadata) Clearinghouse and the NPS Data Store.

The *ESRI Metadata Profile* (ESRI 2003) contains elements to better document aspects of ESRI data formats, particularly objects stored in geodatabases, and allow ArcCatalog© to maintain metadata automatically.

The *NPS Profile* contains information about the park units to which the data pertain.

8.3. General Standards and Guidelines

From project development through final delivery of information products, the Network Data Manger and project leaders will place a high priority on documenting the aims, quality, and meaning of data, and allow for the time this will take when scheduling. Detailed protocols provide an important source of information about the data produced by a project and therefore project leaders will track and document protocol versions (Chapter 12). Metadata will adhere to strict standards, and their development will be guided by the practices described below. Crews will be trained to record decisions made in the field that affect data quality or meaning. Project leaders will start metadata at the onset of the project and set the tone for record-keeping throughout the course of the project.

8.3.1. Content Standards

Network metadata content standards will adhere to those set by FGDC and NPS policy as described above. The CSDGM consists of seven sections (Sections 1-7), all of which contain elements required for spatial data. The NPS Metadata Profile adds another section (Section 0). The Biological and ESRI Profiles are added as elements to Sections 1-7. A summary of the CSDGM and NPS Metadata Profile sections is provided below.

Section 0, NPS Information –purpose of the metadata, relevant park unit(s), and data steward.

Section 1, Identification Information – who produced the data set, when and why it was produced, and where it is from. Constraints on access (e.g., for sensitive data) and use are also recorded in this section. Also includes the geographic extent, bounding altitudes, taxonomy, and analytical tools used in processing (for biological data).

Section 2, Data Quality – the accuracy of attributes and geographic positions and the procedures used to ascertain accuracy. This section also documents the completeness and lineage of the data set. Lineage includes source(s) of the data and processing steps, and methodologies used (for biological data).

Section 3, Spatial Data Organization – methods of spatial reference. Mandatory for spatial data.

Section 4, Spatial Reference Information – coordinate system definitions. Mandatory for spatial data.

Section 5, Entity and Attribute Information – attribute names, definitions, codes and their meanings and other information essential to a basic understanding of the data.

Section 6, Distribution Information – methods and contacts used for obtaining data. Also documents information critical for using biological data formatted in ASCII.

Section 7, Metadata Reference Information – includes who created the metadata, when it was created, the profile used, and the frequency of update.

Metadata that are *fully compliant* with FGDC and NPS standards have entries in Section 0 and all element fields in Sections 1-7 where the *Optionality* field contains the term 'mandatory' or 'mandatory if applicable'. A mandatory element is one which must be populated for every data

set. A mandatory if applicable element is one which must be populated if the data set exhibits the characteristic being documented by the metadata element. For example, the element defining the vertical coordinate system would be mandatory if a data set contains elevation data.

Metadata that are *minimally compliant* with FGDC and NPS standards have entries in Section 0 and all 'mandatory' and 'mandatory if applicable' element fields in Sections 1, 6, and 7, and Section 2 for biological data. These include the fields used by the NPS Data Store. Data sets documented to this extent can be distributed via the Data Store's online upload utility. See Appendix 8B for guidance on creating metadata for the NPS Data Store.

Different types of data and information require different kinds and levels of documentation. Standards for documentation of SIEN data sets are as follows:

Spatial Data will contain, at a minimum, all of Section 0 and the required elements of Sections 1-7.

Non-Spatial Data will include, at a minimum, all elements of Section 0, the required elements of Sections 1, 6, and 7, and Section 2 for biological data. The minimum requirements for non-spatial data therefore meet the requirements for minimum compliance with the FGDC and NPS standards.

Relational Databases will be documented according to the standards outlined above. Complete documentation will also include entity relationship diagrams, business rules, and programming code. The NPS Metadata Profile currently does not support this type of documentation, so it will be stored separately from the formal metadata, in a folder with the database. Relational databases will also utilize internal documentation such as table and field descriptions and will include a table to track modifications. See Chapter 5 for more information on documenting relational databases.

***Legacy Dat*a** will be documented to the extent possible according to the standards outlined above. Metadata that accompany legacy spatial and non-spatial data sets are suitable for upload to the NPS Data Store if they include entries sufficient for minimum compliance with FGDC and NPS standards. Priority for documentation will be:

- Data sets needed for current project development
- Data sets used frequently by park staff or cooperators
- Historic data sets archived for possible future use

Any contracts entered into by the SIEN with data miners will stipulate the submission of FGDC and NPS-compliant metadata. The Network Data Manager or project leader will assist with metadata acquisition by providing tools, format protocols and file transfer services.

Legacy metadata in SIMMS databases should be converted to Extensible Markup Language (XML) as outlined in SIEN Standard Operating Procedures: Migrating metadata from SIMMS to XML (Appendix 8C).

Data from Outside the Network are data that are generated and/or managed outside of SIEN programs but used in analysis with SIEN data or distributed in any manner by the NPS. These data require the same level of documentation produced for SIEN-generated data, including but not limited to, data produced under contract with the NPS. Metadata will be requested from the originating entity by the Network Data Manager or project lead.

Generally, external data will not be posted on Network or park local or wide area networks (LAN/WAN) without accompanying metadata. This will include any metadata downloaded with the data, plus additional information regarding date of download and any alterations made to the data by NPS staff. Staff posting data to the SIEN LAN or WAN will make a reasonable effort to make up for any deficiency in the original metadata, but should not create new metadata for data from well-known sources such as USGS digital line graph (DLG) data. NPS staff will occasionally post 'value added' external data (e.g., a digital elevation model clipped to park boundaries and converted to the standard projection) to the NPS Data Store, and associated metadata will reflect the source data as well as processing prior to upload.

Sensitive Data. Metadata documentation (Section 1, Constraints on Access) provides one means of labeling sensitive data in order to ensure their protection and integrity over time. Sensitive data therefore will be documented according to the standards outlined above, although these data and their metadata will not be uploaded to the NPS Data Store. Instead, they will be archived in secure locations on Park and/or Network servers.

Digital Photos. A national metadata standard for digital photos is currently under development by WASO. In the meantime, digital photos will be documented according to the specifications outlined in Sierra Nevada Network Digital Photo Management Strategy: Goals, Policies, and Guidelines (Appendix 6C) and will include at a minimum:

- Author – who took the photo, with agency affiliation
- Title – who or what is in the image
- Location – where the image was taken
- Date – when the image was taken
- Park code
- Access constraints – who may view the image
- Copyright information – restrictions on using the image
- Contact information – who to contact for further information

8.3.2. Format and Storage Standards

To enable uploading to the NPS Data Store and further public distribution by NPS servers, metadata must be formatted in a manner that will allow parsing with the USGS Metadata Parser (see below). Proper formatting can be most easily accomplished using the metadata tools described below.

Metadata and other documentation will be stored as close to their associated data files as possible, with formal metadata files saved in XML. Metadata may be exported in text, Hypertext Markup Language (HTML) or other formats as needed but the original XML file will be maintained with the data as the master copy.

Hypertext links within the metadata document will point to accompanying datasets stored on Network servers. In general, a single metadata document will apply to both raw and certified versions of the data.

8.3.3. Metadata Tools

ESRI's ArcCatalog© is a multifunction application for managing spatial data and for editing FGDC compliant metadata. ArcCatalog© uses stylesheets to display XML metadata in a format for easy viewing and editing. Because ArcCatalog© metadata are linked to the datasets they describe, certain inherent properties of datasets, such as bounds, coordinate system, feature count, and attribute names, can be automatically populated and maintained in the metadata.

NPS Metadata Tools and Editor (Appendix 8D) is a custom software application for authoring and editing NPS metadata. It extends the basic functionality of ArcCatalog© for managing spatial metadata and provides a stand-alone tool for creating and manipulating non-spatial metadata outside of ArcCatalog©. The Metadata Editor formats metadata according to the NPS Metadata Profile. Like ArcCatalog©, editing is done with stylesheets. A variety of stylesheets are offered, each tailored to a specific type of metadata and displaying only the elements or sections pertinent to that type. NPS-specific stylesheets are based on the NPS Metadata Profile extension and contain all elements in the FGDC, ESRI, and Biological metadata standards and profiles plus NPS-specific elements. Metadata can be parsed with mp, the USGS MetaParser which comes bundled with the Metadata Tools. Output format is XML, suitable for upload to the NPS Data Store. Metadata Tools includes utilities for searching, cataloging, parsing, and spell-checking metadata records.

NPS Database Metadata Extractor is an add-in for MS Access 2000-2003. The extractor automatically harvests entity (table) and attribute (field) metadata from MS Access databases, including domains. It further allows the user to edit and review the harvested metadata and make batch edits and to export metadata to a FGDC-compliant XML file. Exported XML can be used in the Metadata Tools & Editor either by opening it to start a new metadata record or by updating with template to fill section 5 of an existing record. This tool will eventually become part of the NPS Metadata Tools & Editor.

Metadata Parser (mp) is a program developed by the USGS that comes bundled with both ArcCatalog© and the NPS Metadata Tools and Editor. It is used to validate metadata records by checking the syntax against the CSDGM and to generate compliant output files for posting to clearinghouses. The parser generates a textual report indicating errors in the metadata, primarily in the structure, but also in the values of some of the scalar elements where values are restricted by the standard.

The "Metadata in Plain Language" questionnaire (Schweitzer 2006, Appendix 8E) is a user-friendly information sheet (MS Word© document) that can be used to elicit metadata needed to complete all FGDC required elements for data sets not formatted for geographic information systems (e.g., shapefiles, coverages, and geotiffs). It currently does not include questions specific to the Biological Data Profile nor the NPS Profile and therefore, does not alone, meet the content standards specified by the NPS Metadata Profile. Additional information must be obtained to comply with these specifications.

8.3.4. Work Flow Process

The Network will establish a standard operating procedure for metadata generation and maintenance (Appendix 8F). In general, metadata development begins with project design and planning (Chapter 3). Metadata documents should be created with the tools recommended above to ensure standardization and the ability to upload to the NPS Data Store. Other tools may be used as long as the metadata produced meet the appropriate requirements for content, format and organization. Initially, some sections may be populated by SIEN staff with information from formal written project protocols. As the project progresses, the document will be completed with information such as field season dates, protocol changes, and taxonomy information. Completion will occur after the dataset is certified by the Network Data Manager or project leader. The Network Data Manager or project leader ensures that updated metadata are uploaded to the NPS Data Store. The relationship between local applications and tools, and national-level storage and delivery system is illustrated in Figure 9. See details for data dissemination in (Chapter 10).

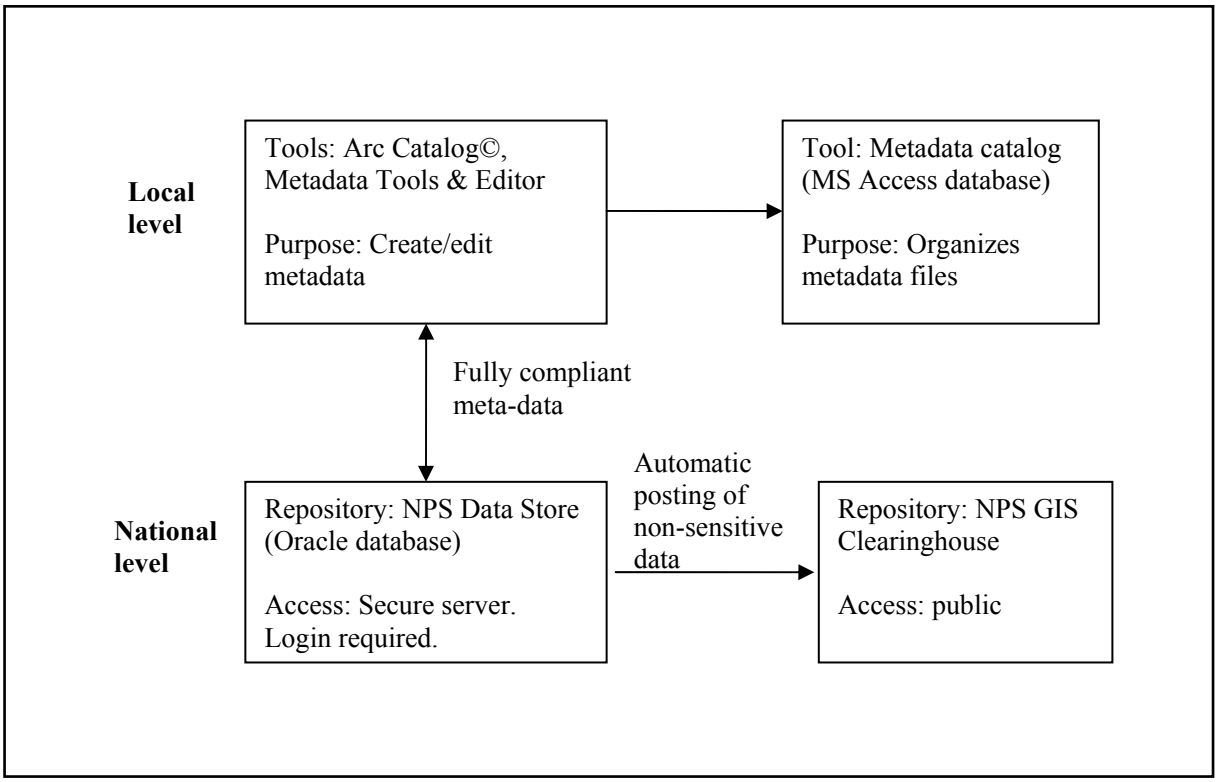

Figure 9. The NPS integrated metadata management system.

8.3.5. Maintenance and Access

Once metadata are complete, the Network Data Manager or project leader saves one copy with the primary dataset at the local level, and uploads an XML-formatted copy to the NPS Data Store.

Metadata can easily be searched with Metadata Catalog, one of the tools in the NPS Metadata Tools and Editor. Metadata Catalog creates a Microsoft Access database that stores the location, filename and other information about all metadata files in a selected directory.

8.4. References

ESRI. 2003. ESRI Profile of the Content Standard for Digital Geospatial Metadata. Environmental Systems Research Institute, Redlands, California.

Federal Geographic Data Committee. 1994. Content standards for digital spatial metadata (June 8 draft). Federal Geographic Data Committee. Washington, D.C.

FGDC Biological Data Working Group, and USGS Biological Resources Division. 1999. Content Standard for Digital Geospatial Metadata - Biological Data Profile, FGDC-STD-001.1-1999 Federal Geographic Data Committee. Washington, D.C.

Mortenson, D. 2005. Data Management Plan for the Inventory and Monitoring Program, Southwest Alaska Network. United States National Park Service. Anchorage, Alaska. 234 pp.

Schweitzer, P. 2006. Metadata in Plain Language. United States Geological Survey, Reston, Virginia. http://geology.usgs.gov/tools/metadata/tools/doc/ctc/.

8.5. Appendixes

6C) Sierra Nevada Network Digital Photo Management Strategy
8A) Executive Order 12906: Coordinating Geographic Data Acquisition and Access, the National Spatial Data Infrastructure
8B) NR GIS Data Store: Metadata Authoring Guidance
8C) Migrating Metadata from SMMS to XML
8D) NPS Metadata Tools and Editor: A Comprehensive Editing Tool for National Park Service Metadata
8E) Tools for Creation of Formal Metadata: Metadata in Plain Language
8F) Sierra Nevada Network Guidelines for developing FGDC compliant metadata for both spatial and biological data

8.6. Credits

This chapter was written by Rosamonde Cook, with concepts and material developed by Mortenson (2005) and information provided by Ginger Bradshaw and Chris Dietrich.

Chapter 9. Data Ownership and Sharing

For the purposes of releasing data, the term *sharing* means releasing to an individual or entity upon request or as part of an agreement to do work that benefits the parks. The term *dissemination* includes releases to the public via publications or other standard routes of delivery such as the NPS Data Store. Chapter 10 provides detailed standards and guidelines for data dissemination. Guidelines for data sharing, including sensitive data procedures, are described in this chapter.

9.1. Objectives

- Establish clear guidelines for the ownership and sharing of natural resources data and information.
- Protect sensitive data from unauthorized access and inappropriate use.

9.2. Laws and Policies

The National Park Service defines conditions for the ownership and sharing of data, natural resource collections, and results based on research funded by the United States government in two principal documents - the Freedom of Information Act, 5 U.S.C. § 552, referred to as FOIA (Appendix 9A), and Office of Management and Budget (OMB) Circular A-110 (Office of Management and the Budget, 1999).

FOIA stipulates that the United States government, including the National Park Service, must provide access to data and information of interest to the public, regardless of whether or not they were created by the federal government. FOIA is intended to establish a right for any person to access federal agency records that are not protected from disclosure by exemption or by special law enforcement record exclusions. Under the terms of FOIA, agencies must make non-protected data and information products available for inspection and copying in public reading rooms and/or from the Internet.

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

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

The 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. OMB Circular A-110 establishes property standards within cooperative agreements with higher institutions and non-profit organizations. Section 36, “Intangible Property” describes the administrative

requirements pertinent to data and ownership, stipulating that all data and materials collected or generated using National Park Service personnel and funds become the property of the National Park from which they were collected.

For additional information, see:

- The Freedom of Information Act 5 U.S.C. § 552, As Amended By Public Law No. 104-231, 110 Stat. 3048 (Appendix 9A).
- Sierra Nevada Network FOIA and Sensitive Data Guidelines (Appendix 9B).
- A Synopsis of Laws and Policies Governing Public Dissemination of Natural Resources Data (Appendix 9C). Also contains the text to Section 36 of OMB Circular A-110.
- U.S. Department of the Interior FOIA regulations, instructions, and general information at http://www.doi.gov/foia/.
- Draft Director's Order #66: Freedom of Information Act and the Protection of Exempted Information (National Park Service, 2004a).
- Reference Manual RM-66A: FOIA processing (National Park Service, 2004b).
- Draft Reference Manual RM-66B: Handling protected information (National Park Service, 2004c).

9.3. General Standards and Guidelines

Data and information will be shared internally through Network channels, externally through online databases and repositories, and by special request as appropriate, and only after quality control procedures have been fully implemented.

9.3.1. Collaborative Agreements

To ensure that proper ownership, format, and development of Network products are 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 Network must cite OMB Circular A-110 under the *Reports and Deliverables* Section. 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."*

Every cooperative or interagency agreement or contract must include a list of deliverables and products clearly defined within each agreement or contract. Details on formatting and media types that will be required for final submission must be included. Agreements and contracts must list all products expected to result from the project. These include, but are not limited to, field notebooks, photographs (hardcopy and digital), specimens, raw data, and reports.

The following statement must be included in the Reports and Deliverables section of all agreements and contracts made by the SIEN Inventory and Monitoring Program:

> *"All reports and deliverables must follow Sierra Nevada Network Product Specifications".*

See Sierra Nevada Network I&M Program Checklist of Project Specifications (Appendix 6A) and the documents referenced therein for further details on data acquisition, documentation, and reporting standards and requirements.

Project leaders should provide a schedule of deliverables that includes sufficient time for the NPS to review drafts before scheduled final submissions.

9.3.2. Due Notification

SIEN staff will notify investigators prior to making datasets available to the public. This will allow each investigator the opportunity to make a request in writing to further restrict access to the dataset by the public. Network staff will review the investigator's request and determine whether the request will be granted and for how long the dataset will remain restricted. Details on how data and information products will be made available to the public are provided in Chapter 10.

9.3.3. Access Restrictions on Sensitive Data

With regard to natural and cultural resources, sensitive data and information are those that could be used to harm, remove, or destroy sensitive resources protected by units of the National Park System. Network staff members are responsible for managing access to sensitive data handled by their programs. These responsibilities include:

- Identify and classify all potentially sensitive park resources and information.
- Provide to superintendents, a complete list of potentially sensitive park resources for which data may be collected as part of any project. The superintendent then determines what information should be protected.
- Ensure that all protected information, digital and hardcopy, is properly identified and marked.
- Ensure that all references to protected information are removed or obscured in any reports, publications, maps, data, or other form made available to the public.
- Inform investigators working on Network projects 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 will identify all potentially sensitive park resources to the principal investigator for each project. Reciprocally, the principal investigators for each project must identify any known references to potentially sensitive park resources.

Agreements with partners will include a standard confidentiality agreement which states that, among other things, the researcher will not share protected data or information with any other party, public, private, or academic, without channeling requests through the NPS project lead. Example language is provided in Appendix 9B. Researchers may fulfill such requests, but there will be a confidentiality agreement between NPS and the new holder of the data.

When sharing data and information on sensitive resources with other federal agencies, the other agency will be told of the need to channel any requests for the data or information through the NPS, as per NPS Director's Order #66 (National Park Service 2004a) and accompanying Reference Manuals 66A (National Park Service 2004b) and 66B (National Park Service 2004c).

9.3.4. Answering Requests under the Freedom of Information Act

Data stewards will forward records requested pursuant to the FOIA to the requester via the Park FOIA Officer in a timely manner. Data stewards will maintain records subject to the FOIA in a manner so as to make them rapidly available and to leave no doubt that the most recent version of accurate, non-sensitive information is provided to requesters. Data stewards are responsible for reading, understanding, and following applicable laws, policy, and procedural guides.

If information requested under the FOIA has not already undergone sensitivity classification, it will be thoroughly but rapidly assessed for presence of sensitive information and, if such information is found, the data steward will work with the appropriate Park and regional FOIA Officers to formulate a response to the requester, which may include complete denial, complete fulfillment, or partial fulfillment once sensitive information have been cleared from the requested records.

9.4. References

Hart, M. and U. Gafvert. 2005. Data Management Plan: Great Lakes Inventory and Monitoring Network. National Park Service Great Lakes Inventory and Monitoring Network Report GLKN/2005/20.

National Park Service. 2004a. Draft Director's Order #66: Freedom of Information Act and the Protection of Exempted Information. http://www1.nrintra.nps.gov/DO66/ DO_66_Final_ Draft.doc.

National Park Service. 2004b. Draft Reference Manual RM-66A: FOIA processing. http://www1. nrintra.nps.gov/DO66/RM-66A-Draft.doc.

National Park Service. 2004c. Draft Reference Manual RM-66B: Handling protected information. http://www1.nrintra.nps.gov/DO66/RM-66B-Draft.doc.

Office of Management and the Budget. 1999. Circular A-110. http://www.whitehouse.gov/omb/ circulars/a110/a110.html#1

Stevens, S. M. and G. Entsminger. 2004. Information Management Plan for the Inventory and Monitoring Program, Northeast Coastal and Barrier Network. United States National Park Service. Kingston, Rhode Island. 99 pp.

9.5. Appendixes

6A) Sierra Nevada Network I&M Program Checklist of Project Specifications

9A) The Freedom of Information Act 5 U.S.C. § 552, As Amended By Public Law No. 104-231, 110 Stat. 3048

9B) Sierra Nevada Network FOIA and Sensitive Data Guidelines
9C) Synopsis of Laws and Policies Governing Public Dissemination of Natural Resources Data

9.6. Credits

This chapter was written by Rosamonde Cook, adapted from material developed by Stevens and Entsminger (2004) and Hart and Gafvert (2005).

Chapter 10. Data Dissemination

Under the terms of Freedom of Information Act (5 U.S.C. § 552), public agencies must make non-protected data (described in Chapter 9) and information available for inspection and copying in public reading rooms, the Internet, or via requests through a specified process. Providing well-documented data in a timely manner is one of the most important goals of the I&M Program, and critical to the success of the program. This chapter describes the methods by which quality natural resource data and information collected by the SIEN are made available to park managers, researchers, educators, and the general public.

10.1. Objectives

- Ensure non-sensitive data are easily discoverable and obtainable.
- Ensure data that have not yet been subjected to substantial quality control are not released to the public, unless necessary in response to a FOIA request.
- Distribute data with complete and accurate metadata that clearly identify who collected the data, what data were collected, and where, when and how the data were collected.
- Identify and protect sensitive data from unauthorized access.
- Maintain a complete record of data distribution/dissemination.

10.2. Laws and Policies

Three major pieces of legislation set the tone for information dissemination and protection by federal agencies. They are the Freedom of Information Act, 5 USC Section 552 (FOIA), the National Parks Omnibus Management Act of 1998, 16 USC Section 5937 (NPOMA), and OMB Circular A-110 (Office of Management and the Budget, 1999). Each provides for the wide and proactive sharing of information and each recognizes instances where such sharing may be harmful.

The FOIA requires that federal agencies provide access to most data and information holdings, whether developed by employees or by cooperators. Agencies are directed to post commonly requested items in public reading rooms and/or on public websites. Nine exemptions provide for withholding information, such as that relating to law enforcement actions and confidential personnel records that are protected. Of these, the third exemption allows withholding information that is protected under other statutes. These are summarized in Appendix 9C and include:

- The Archaeological Resources Protection Act, which requires information on the location and nature of archaeological resources to be protected.
- The National Historic Preservation Act, which directs that information regarding historic resources should be withheld if disclosure would cause significant invasion of privacy or risk harm to historic resources.
- The Federal Cave Resources Protection Act, which prohibits release of information regarding the nature and location of caves.

- Executive Order 13007, Protection and Accommodation of Access to Indian Sacred Sites, which allows for maintaining the confidentiality of Indian Sacred Sites.
- The National Parks Omnibus Management Act, which provides protection of sensitive natural and cultural resource information.

The NPOMA, heralded as the NPS "research mandate" (Harmon 1999), established the Cooperative Ecosystem Studies Unit and I&M programs. While NPOMA clearly requires a higher level of information sharing with the research community and the interested public, as well as use in management decisions, the exemptions in Section 207 allow for protection of information about resources not previously specified in other legislation.

Another key policy statement, OMB Circular A-110 establishes federal government rights to use and share information gathered with federal support. The Circular also specifies that such information is subject to the FOIA and its exemptions. See Chapter 9 for details.

10.3. General Standards and Guidelines

The SIEN will employ a number of distribution methods that ensure information collected and developed as part of Network programs are made widely available to park employees and the public.

SIEN parks develop and implement a map center concept that focuses on delivering both electronic and hardcopy maps. Map delivery should be efficient and meet the primary and contemporary mapping needs of all park program staffs.

Data and information will be disseminated principally through the Internet and only after quality control procedures have been fully implemented, unless a request is made pursuant to the FOIA with consideration to sensitivity and ownership classification (Chapter 9).

SIEN staff will notify investigators prior to making datasets available to the public. This will allow each investigator the opportunity to request in writing to further restrict access to the dataset by the public. Network staff will review the Investigator's request and determine whether the request will be granted and for how long the dataset will remain restricted.

10.3.1. Data and Information Distribution Mechanisms

According to FOIA (specifically the 1996 amendments), all information routinely requested must be made available to the public via reading rooms and/or the Internet. The Network's principal means of distributing inventory and monitoring data will be the Internet (Table 5). A number of Internet-based databases and repositories have been developed as part of the NPS I&M Program to store and disseminate a variety of natural resource information. They include:

- The Sierra Nevada Network public website
- National public applications, NPSpecies and NatureBib
- The NPS Data Store
- The Biodiversity Data Store

Table 5. Repositories of SIEN public and sensitive data and information.

Item	Repository
Reports (public) - digital	SIEN LAN servers, SIEN public website, NPS Data Store, NPS Focus
- hard copy	SIEN I&M library, YOSE Library, USGS Libraries
- bibliographic	NatureBib
Network-generated digital datasets and data products (public, non-sensitive) • Certified data and data products (including photographs) • Metadata	SIEN LAN servers, NPS Data Store, Biodiversity Data Store, NPSpecies, EPA STORET, the California Data Exchange Center (weather data), and the California Data Exchange Network (water resources data)
Network-generated digital datasets and data products (NPS staff, sensitive) • Raw, validated data • Analytical products • Metadata • Reports • Digital photos • Digital presentations, etc.	SIEN LAN servers
Project products • Specimen vouchers • Photographic film	SEKI and YOSE Museums or other curation facilities, according to project protocol

10.3.1.1. The SIEN Website: The Network will regularly provide updated information about inventories and monitoring projects, including annual reports and detailed project reports through its public web site (http://www1.nature.nps.gov/im/units/sien/index.htm). Links, and a brief description, will be provided to each of the national databases and repositories described below.

10.3.1.2. National Databases: NPSpecies will be updated with observations and vouchers collected during the course of SIEN projects, if these data are deemed important to the further documentation of species already present on certified park species lists, or constitute additions to these lists (Appendix 10A). Data will either be manually entered into the online databases by SIEN or Park staff, or if present in sufficient quantity (e.g., more than 100 records), sent to the WASO office for upload after first ensuring proper formatting. NatureBib will be updated as new documents become available (Appendix 10B). These databases are available via both a secure server and a public server. The public can not access records flagged 'sensitive' as NPS or Park-Only.

Water quality data collected to meet federal regulatory requirements are managed according to guidelines from the NPS Water Resources Division (WRD), which also oversees the integrated water quality monitoring portion of the I&M Program. NPS requires all I&M data to be uploaded to EPA's STORET National Data Warehouse. WRD developed NPSTORET – a desktop database application for I&M networks to help manage data entry, documentation, and transfer. Data from NPSTORET are transferred periodically STORET (Figure 10). Individual networks are free to use NPSTORET for data entry and maintenance, or to develop a customized database compatible for data exchange and delivery. The SIEN may choose the latter and build a desktop application that would also interface with the California Environmental Data Exchange Network (CEDEN) client-server database, a central repository for water resources data collected within

California and operated through the State's Department of Water Resources. In that case, formal procedures, in accordance with WRD's Electronic Data Deliverable (NPSEDD) file specifications, will be developed for transferring data into NPSTORET for delivery to WRD (Appendix 10C).

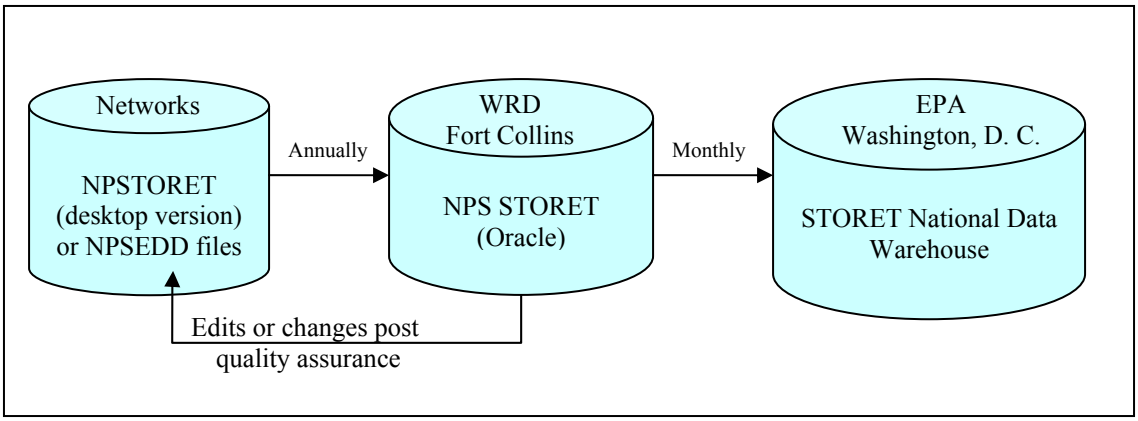

Figure 10. Flow diagram for water quality data from I&M Networks to the National Data Warehouse.

Data from short-term studies (such as biological inventories) will be provided to the public through the NPS Data Store and/or the Biodiversity Data Store within two years after data collection was completed, or following publication of the investigator's results (whichever comes first).

Data from long-term air quality monitoring conducted by Network parks or by cooperators working with the parks, including visibility, gaseous pollutants (mainly ozone), atmospheric deposition (wet and dry), and meteorology are available through the NPS Air Resources Division website at: http://www2.nature.nps.gov/air/monitoring.

Long-term monitoring data collected as part of the I&M Vital Signs Monitoring Program will be made available on the NPS Data Store at four or five year intervals, or when trend analyses have been completed and reported on by the Network. The NPS Data Store is available on a secure server and open to individuals approved for access. All data distributed to the Data Store and not defined as sensitive will be available to the public through the NPS GIS Clearinghouse. Before data are posted, the investigator or project leader will be asked to verify the final dataset and metadata, and to identify any sensitive data or information. Sensitivity will be documented in the metadata which will accompany all data sets (Chapter 8).

Non-sensitive digital images, including photos, drawings, maps, text, and GIS DOQ/DRG images will be made available to the public through the NPS Focus digital library.

Data from the Basic Natural Resources Inventories conducted through the NPS Natural Resource Program Center and the I&M Program are stored in a variety of locations (Appendix 2B).

10.3.2. Sensitive Data

Data released outside of the federal government without a confidentiality agreement is considered a release of information under the Freedom of Information Act and triggers the application of the "release to one, release to all" principle. There are many venues for release of sensitive data that constitute release to the public. Therefore, information should be reviewed for sensitivity by qualified NPS staff prior to inclusion in these media. They include:

- Internet posting
- Interpretive programs, media, and activities
- News releases
- Resource management storing and sharing activities
- NEPA compliance documents

Inadvertent access may be deemed a release depending on the circumstances. Network staff will institute quality control and quality assurance measures to ensure that any person uploading records to public databases is familiar with the procedures for identifying and entering protected sensitive information (Appendix 10A, Appendix 10B). See Appendix 9B on classifying and handling sensitive natural resources data.

10.3.4. Feedback Mechanisms

The SIEN website will provide an opportunity for NPS staff, cooperators and the public to provide feedback on data and information gathered as part of Network programs. A "comments and questions" link will be provided on the main page of the site for general questions and comments. A more specific "data error feedback" link will direct comments to SIEN staff pertaining to errors found in website-accessible data.

The following procedure describes the process that the SIEN will use to receive, verify, and correct data errors identified by web site users.

- Users send a notification through the SIEN website.
- An acknowledgment of receipt is sent to the user.
- Information is input into an error log table incorporated either in the project database or a specific error tracking database developed for the Network.
- Network staff determine whether the data in question are correct or incorrect. If correct, the user is informed that no corrections are to be made and the information stands. If incorrect, Network staff make the appropriate corrections and notify the user and original data collectors.
- The Network website is refreshed with the corrected information.

10.4. References

Harmon, D. 1999. The New Research Mandate for America's National Park System: Where It Came From and What It Could Mean. George Wright Forum 16(1):8-23.

Mortenson, D. 2005. Data Management Plan for the Inventory and Monitoring Program, Southwest Alaska Network. United States National Park Service. Anchorage, Alaska. 234 pp.

Office of Management and the Budget. 1999. Circular A-110. http://www.whitehouse.gov/omb/circulars/a110/a110.html#1.

Stevens, S. M., and G. Entsminger. 2004. Information Management Plan for the Inventory and Monitoring Program, Northeast Coastal and Barrier Network. United States National Park Service. Kingston, Rhode Island. 99 pp.

10.5. Appendixes

2B) National Repositories for Data and Information
9B) Sierra Nevada Network FOIA and Sensitive Data Guidelines
9C) A Synopsis of Laws and Policies Governing Public Dissemination of Natural Resources Data
10A) NPSpecies Data Standards Manual
10B) NatureBib Data Standards Manual
10C) Transferring Vital Signs Water Quality Data to NPSTORET

10.6. Credits

This chapter was written by Rosamonde Cook, with information provided by Ginger Bradshaw and material developed Stevens and Entsminger (2004) and Mortenson (2005).

Chapter 11. Records Management and Object Curation

Documenting park resources and their management is an essential part of National Park Service administration. The potential for loss of data, documents, or objects can come from a variety of sources, including catastrophic events (e.g., fire, flood, and earthquake), user error, hardware failure, software failure or data corruption, theft, and intentional acts of vandalism. The responsibility for the preservation of park records begins with each employee and thorough implementation of effective best practices is crucial to ensuring long-term preservation and accessibility of park records.

The Federal Records Act (44 U.S.C. 3301) defines records as:

> *"... all books, papers, maps, photographs, machine readable materials, or other documentary materials, made or received by an agency of the United States Government under Federal law or in connection with the transaction of public business and preserved as evidence of the organization, functions, policies, decisions, procedures, operations, or other activities of the Government or because of the informational value of data in them".*

The law includes records created under contracts or other agreements with other agencies, organizations, or individuals.

Documents most necessary for fulfillment of the NPS mission are considered *mission critical* records. The Records Management Handbook (National Park Service 2005) defines such records as resource management records – "*the permanently valuable, substantive program records documenting the actual work of the Park Service in managing critical cultural, natural, and informational resources.*" Records documenting natural and cultural resources and their management contain information that may affect the future management of these resources and are irreplaceable. They should receive the highest priority in records management activities and resources and should receive archival care as soon as practical after their creation.

Resource management records may include inventories, drawings, field notes, maps, photographic negatives, prints and slides, audio and video recordings, and electronic files and media. All are considered records and their management and preservation are addressed in this chapter.

This chapter also provides guidelines and standards for the curation of natural history specimens and other non-record objects. Such collections, and their associated records, are by definition museum objects, but have much in common with resource management records. Most significantly, both result from park studies as defined in Management Directive 4 of the Records Management Handbook:

> *"The term studies means short- or long-term scientific or scholarly investigations or educational activities that may involve natural resource surveys, inventories, monitoring, and research, including data and specimen collection. Studies include projects conducted by researchers and scholars in universities, foundations and other*

institutions; tribal colleges and organizations; other federal, tribal, and state agencies; and NPS staff. The data and information acquired through studies conducted in parks will be made publicly available".

Digital data pose particular challenges to long-term preservation. Technological obsolescence is a significant cause of information loss, and data can quickly become inaccessible to users if stored in out-of-date software programs or on outmoded media. Effective maintenance of digital files depends on proper management of a continuously changing infrastructure of hardware, software, file formats, and storage media. Major changes in hardware can be expected every 1-2 years, and in software every 1-5 years (Vogt-O'Connor 2000). As software and hardware evolve, data sets must be consistently migrated to new platforms, or they must be saved in formats that are independent of specific platforms or software (e.g., ASCII delimited files). This chapter describes procedures for both *short* and *long-term* preservation and management of electronic files storage.

An essential part of any archival holding (digital or non-digital) is its accompanying explanatory materials (Olson and McCord 1998). Thus, effective long-term maintenance of these products depends on thoughtful and appropriate data documentation. Chapter 6 provides detailed policies, guidelines, and practices for documenting data. Chapter 12 provides detailed policies on project tracking and organization.

11.1. Objectives

- Adopt and implement robust standard file directory structure specifications.
- Where practical, maintain *mission critical* and *permanent* paper records in both hardcopy and electronic file format, and protect and preserve them <u>indefinitely</u>.
- Maintain datasets no more than two versions behind current software versions, or store in American Standard Code for Information Interchange (ASCII) format, complete with data and file documentation.
- Ensure information can be easily obtained, shared, and properly interpreted by a broad range of users.
- Ensure backup, storage, and recovery practices for electronic files equal or exceed the minimum standards established by the NPS Office of Chief Information Officer.
- Store all electronic files on servers in a networked environment using approved file-naming standards and file directory structures.
- Maintain all data, programmatic, and administrative electronic files <u>indefinitely</u>.
- Ensure all short and long term projects are well-documented, organized, and protected according to local and national standards and guidelines.
- Ensure all collected natural history specimens and objects are well-documented and preserved in perpetuity.

11.2. Laws and Policies

The most relevant guidance for managing NPS records and object curation are found in the following references:

- The NPS Records Management Handbook (National Park Service 2005a)
- Management Directive 4 of the NPS Records Management Handbook: Studies and Collections. Provides direction for approving and managing scientific research
- NPS Director's Order 19: Records Management (National Park Service 2001a). Provides direction for managing and preserving records
- NPS Director's Order 19, Appendix B: Records Disposition Schedule (National Park Service 2001b). Determines the long term disposition of specific NPS records
- NPS Director's Order 24: Museum Collections Management (National Park Service 2000). Provides direction for managing and preserving museum collections
- 36 CFR 2.5: Research Specimens (Code of Federal Regulations, 2001). Provides direction for managing natural history collections
- NPS Management Policies (National Park Service 2006). Describes requirements for the long term housing of records and objects
- NPS Museum Handbook (National Park Service 2005b). Provides the overarching guidance for archival procedures

11.2.1. Records

NPS Records Disposition Schedule (National Park Service 2001b) defines the appropriate disposition of specific record categories. The term "disposition" applies to records no longer needed for active, on-going business; it can mean permanent archiving, transfer to a Federal Records Center, or destruction. Section N of this Appendix (Natural and Social Sciences) states that all natural resource records are considered *permanent* and need to be retained in an appropriate park museum facility or at the National Archives. It also states that even non-archival copies of natural resource-related materials are "potentially important for the ongoing management of NPS resources" and should not, in any instance, be destroyed.

Directors Orders 19, Section 6.2 provides specific guidelines and standards for <u>electronic records</u>. Under these guidelines, an electronic or digital recordkeeping system may be used as the <u>primary</u> recordkeeping system <u>only</u> under the following circumstances:

- When the records in the system are, or can be made, software and hardware independent, and meet current NARA format standards for electronic records.
- When the records to be stored are temporary, with a retention period of 5 years or less. This includes most records typically created during routine administrative functions in offices on electronic mail and word processing programs.
- Inherently electronic records where there is little practical choice, (generally geo-spatial data and web pages).
- If the records to be stored are temporary, with a retention period of more than 5 years.

In all other circumstances, traditional media records of paper, photography, film, microfiche and magnetic recordings must be used for archival storage.

11.2.2. Collections

Regarding the collection of natural history specimens, NPS Management Policies states the following:

> *"Field data, objects, specimens, and features obtained for preservation during inventory, monitoring, research, and study projects, together with associated records and reports, will be managed over the long term within the museum collection. Specimens that are not authorized for consumptive analysis remain federal property and will be labeled and cataloged into the NPS cataloging system (ANCS+, or its successor) in accordance with applicable regulations".*

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

11.2.3. Local Network Policies

These policies supplement or reinforce certain national policies deemed very important to the Network.

- To the extent practicable and consistent with the other goals and policies identified in this plan, backup and archival storage of electronic files will comply with relevant existing federal, department, and bureau policy, rules and regulations.
- Park IT staffs are responsible and accountable for maintaining, backing-up, and restoring electronic files on all Network servers.
- Regular and complete backups of electronic files on personal computers are the responsibility of each staff member.
- The rotating configuration for media backups will be five years duration with at least quarterly offsite media backups supporting long-term storage.
- To the extent practicable, any electronic file that meets the definition of both *mission critical* or a *permanent* record will be printed out and managed as a hardcopy record.
- Local area network temporary, permanent, and disposition-suspended federal records and objects will meet the minimum physical and security conditions required by federal, department, and bureau laws, mandates, and policies.

11.3. General Standards and Guidelines

11.3.1. Records Management

11.3.1.1. Electronic File Naming Guidelines and Standards: All electronic files will be named according to guidelines and standards outlined in Appendix 11A – Electronic File Naming Guidelines and Standards. These will apply to all SIEN electronic files created or maintained by staff or cooperators. Network programs that are adopting them will be clearly identified in Part 2 of the appendix.

Highlights of SIEN file naming standards:

- File names should be as short as possible, but include all mandatory information.
- The electronic file name should be unique, such that the chances for another electronic file having the same name on a local area network are unlikely.
- Version numbers (e.g., version1, version2, final, etc.) are not to be used on documents. The date of file creation (yyyymmdd) will represent the version.
- Generally, use lower case characters. However, it is an acceptable practice to mix upper and lower case letters in a filename to separate a concatenation of words, such as words in a title.
- Do not use spaces in file names.
- Do not use special characters, other than the underscore (_), and periods.
- Use the underscore to separate categories in an electronic file name.
- Use generally recognizable abbreviations or spell words out.
- Avoid uncommon acronyms, abbreviations and codes. Generally, avoid codes that require users to refer to another source for the code description. NPS codes such as park code (e.g., YOSE, SEKI) are acceptable.
- Use leading zeroes for numbers 1 through 9.
- File extensions: Use the default extension recommended by the software application associated with the document. (Examples: doc, xls, mdb, rtf)

11.3.1.2. Directory Structure: Directory file structures already exist on Network servers but are generally inadequate. Electronic files will be managed within a hierarchical set of file directory structures. Appendix 11B provides an example of a very detailed layout of the directory structure for the SIEN I&M program which includes folders for administrative files, data, publications, and reports. All Network programs will shift toward a file directory structure like this one. The SEKI local area network will be the primary repository for all Network I&M electronic files, but will provide access to YOSE and DEPO employees and I&M staff. Local area network file servers will remain the repositories for individual park programs.

Key aspects of the SIEN file management strategy include the following:

- To the extent possible, all applications work (e.g., spreadsheets, GIS, documents, databases) will be conducted on a SIEN server.
- The local area network copy of all electronic files will be the primary version. Any local electronic file versions on personal computers should be considered backup files.
- Generally, only one copy of an electronic file will be maintained on a Network server. One exception to this situation would be a *master* database that is read-only, but may also have a working database copy that is being actively updated.
- Generally, all electronic files will be maintained in a networked environment that enables immediate retrieval. Although there are some exceptions, generally offline storage and retrieval of electronic files from media will be discouraged.
- Working files are kept separate from finished products.
- Finished products are typically read-only, except for "inbox" folders where users can drop things off to be cataloged and filed.

- Standards, such as naming conventions and hierarchical filing, are enforced within shared and archived sections (e.g., libraries, GIS, databases). Although less stringent in other sections, these conventions are encouraged as good practice.
- Version control is implemented in both active and archived directories. Generally, previous versions of databases will be saved in their native format and archived using the same best practices as used for current versions. Documentation of version updates and associated details will be part of the archive metadata document, and revision information and history will also be included in tables within the database files themselves.

A *Project Template Directory Structure* for short-term projects has been created (Appendix 11C). The data manager or designated staff will create a working directory based on this template for each project, which project leaders can then modify as needed (e.g., deleting directories not being used). Upon completion, projects and products are certified by project leaders and moved if needed.

11.3.2. Archival, Backup, and Storage

11.3.2.1. Project Process and Workflow: Both electronic and hardcopy files should be consolidated and packaged for archival when a project is complete or when milestones are reached. Project leaders are responsible for packaging electronic files and data and for preparing materials for the curator. Project protocols should designate who will be responsible for product archiving, integration, backup, and distribution. The workflow diagram below illustrates the paths that product archival and storage follow (Figure 11). Museum staff will follow a much more detailed procedure for actual curation.

For **digital** data, the project leader should prepare files as follows:

- Comply with SIEN electronic file naming standards
- Verify that all electronic files are organized on an NPS server according to SIEN file directory hierarchy standards.
- Comply with SIEN data documentation standards (Chapter 8).
- Include all related documents listed in the SIEN project deliverables guidance document (Appendix 6A).
- Assure that hardcopy records are printed for all mission critical or permanent records.
- Make a clear distinction between public and sensitive information.
- For I&M Projects, ensure that the I&M Project Tracking Database (Chapter 12) has been updated and is complete.

For **digital** data, the data manager or other designated staff is responsible for the following:

- Archive project information on an SIEN server according to SIEN file directory hierarchy standards.
- Assure that proper securities are assigned to electronic files.
- Complete a project backup on CD or DVD that is placed in an approved location.

- Integrate deliverables, such as final reports, into appropriate catalogs, electronic libraries and NPS clearinghouses, as described in this plan or specific project protocols.

For **non-digital** records and objects, the project leader should package information as follows:

- Complete a coversheet listing contact information, project abstract and purpose, sensitivity, use of materials, and all materials included in the package (Appendix 11D).
- Clearly label materials with park acronym, date(s), accession number, project number.
- Comply with paper and folder type requirements (e.g., acid free paper).
- Print and document selected photos.
- Label voucher specimens according to the Museum Handbook and complete loan paper work if necessary. Specimens not residing in NPS repositories should be stored in a selected facility that meets NPS museum collection standards.
- Address specimen curation in project protocols and specify name and address of facility, Memorandum of Agreement (if applicable), list of items to be curated or loaned.
- Create a CD or DVD of all related, electronic materials and label with park acronym, date created, date(s) of information, accession number, project number, list of contents

For **non-digital** records and objects, museum staff are responsible for the following:

- Catalog information into ANCS+ and properly store materials according to the Museum Handbook.

11.3.2.2. Electronic Files: Performing regular backups of data and arranging for offsite storage of backup data sets are important safeguards against data loss. Recently, the NPS developed general standards and guidelines for backup, storage, and retrieval of electronic files (Appendix 11E). This document describes high level procedures at the bureau level that provide *minimum* standards for data backup, storage, and recovery. It was produced by the NPS Office of the Chief Information Officer to provide high-level procedures for the NPS General Support System. These efforts center on providing the NPS with the ability to quickly reinstate information systems and the data required to run those systems.

A primary requirement of the NPS General Standards and Guidelines for Backup, Storage, and Retrieval of Electronic Files is that each NPS organizational unit creates a formal set of backup procedures for critical system-level and user-level information. The procedures should address documenting backups and enforcing backup policies. These requirements have been met through the development of Network General Support Services Backup Activity Plans for SEKI & DEPO (Appendix 11F) and YOSE (Appendix 11G). These plans outline specific and consistent backup routines for Network parks, including: frequency and types of backups; media rotation schedules; hardware, software and media specifications; offsite storage locations; testing procedures; staff responsibility; and data classification on individual servers.

11.3.2.2.1. Individual Employee Responsibility: Backups of electronic files on personal computers are the responsibility of each staff member. SIEN policy discourages electronic file storage on personal computers and, to the extent possible, electronic files should be managed on a Network server. Most SIEN employees are used to managing and maintaining their electronic

files on a personal computer, rather than storing and managing on a Network server. The primary advantage for shifting from personal computer to local area network storage is that backup management and security are shifted to the IT staff who use well-defined protocols and optimize protection and backup of electronic files. The IT staff will provide an individual work directory on a Network server for those staff who request personal workspace. Other than redundant backup to a personal computer's hard drive, the use of individual work directories eliminates electronic file storage on local hard drives.

In some situations, employees work in geographic locations that are not networked and are unable to develop and manage electronic files on a Network server. In situations where an employee works with a computer that is off-network, each employee will be responsible for archiving and storing electronic files before they are transferred to a SIEN server. Off-network electronic files should be regularly transferred to a networked server for permanent archiving, storage, and final disposition. Appendix 11F and Appendix 11G provide specific guidance for SIEN staff who use personal computers off network including best practices for backing up and copying electronic files to SIEN servers.

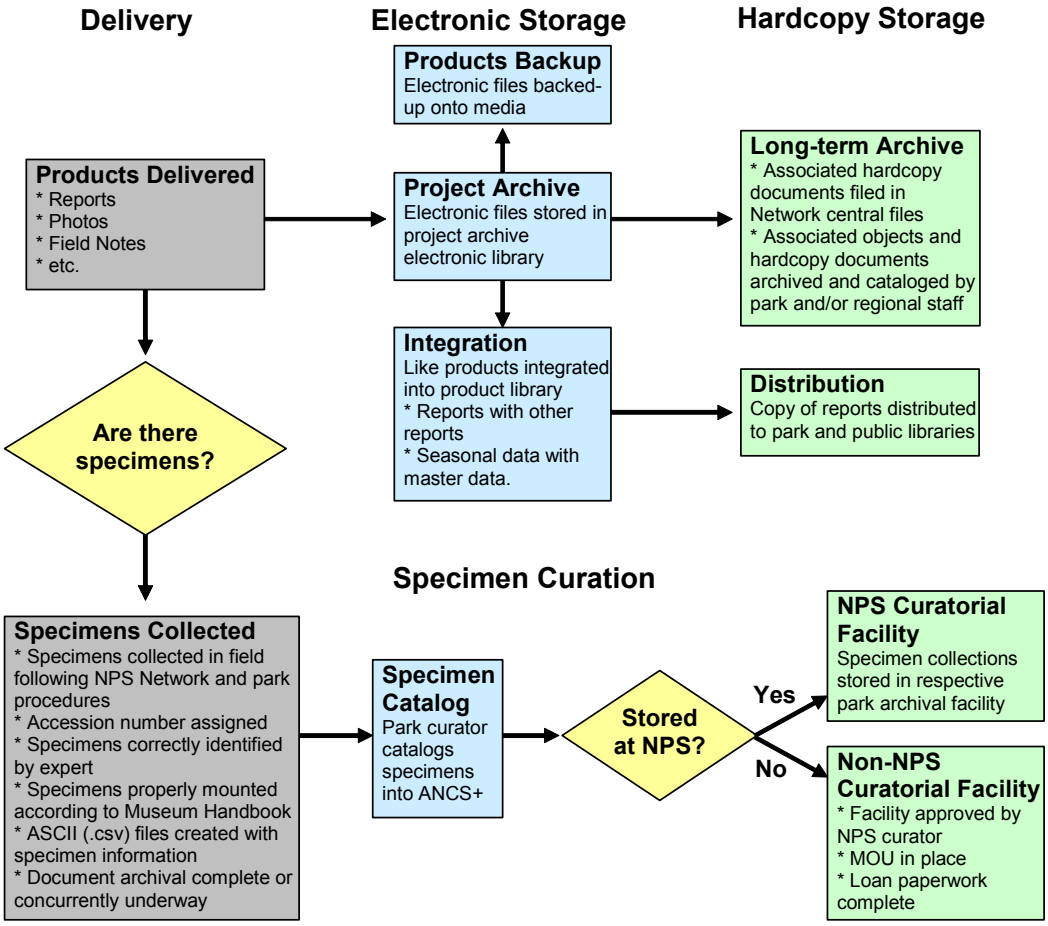

Figure 11. General workflow for storage and archival of electronic (in blue) and hardcopy (in green) scientific research products.

11.3.2.2.2:. IT Staff Responsibility: IT staff will adhere to their Network General Support Services Backup Activity Plans (described above).

11.3.2.2.3.: Short-term Projects: Short-term projects usually last from one to three years and include individual research projects, inventories, or pilot work done in preparation for long-term research or monitoring.

Upon project finalization, a set of ASCII comma-delimited text files will be created for each data table comprising the dataset. These files will be accompanied by a *readme.txt* file that explains the contents of each file, file relationships, and field definitions. The ASCII files are in addition to the native version of the dataset (typically in database format) and will help ensure the data are usable in a wide range of applications or platforms.

The SIEN will employ a utility created by the Gulf Coast Network called *Access_to_ ASCII.mdb,* which automates the creation of these ASCII files from Access databases. All ASCII files created from databases will undergo quality control to ensure the number of records and fields correspond to the source data set and that conversion has not created errors or data loss. A second reviewer (preferably the project leader) will evaluate the ASCII files and documentation to verify that tables, fields, and relations are fully explained and presented in a way that is useful to secondary users. All finalized files will be stored in the archive section of the project directory folders.

In addition to creating ASCII files, SIEN will also update completed and archived data sets that may be in older versions of MS Access, with the goal of having no dataset more than two versions behind the current version used by the Network. Converted databases will require thorough quality control if they are actively being used for data entry or analysis and if no documentation of prior quality assurance/quality control measures is available. Forms, queries, reports, and data entry will be thoroughly tested, and corrected where any problems occur. All previous versions of the data set will be saved.

11.3.2.2.4.: Long-term Projects: Long-term data are projects include vital sign monitoring and other multi-year research and monitoring projects performed by the parks, other agencies and cooperators and for which data acquisition and entry will continue indefinitely.

Long-term monitoring databases will require occasional conversion to current database formats. All active or long-term databases will conform to the current NPS software version standards.

Long-term projects also will have variable long-term data archiving requirements. Raw data should always be stored in perpetuity. Processed data sets should be archived complete before any modifications are made to project protocols. Depending on the project, it might also be necessary to preserve interim data sets (data “milestones”) over the long term. Archived data sets or subsets destined for long-term archival will be saved, whenever possible, in their native formats in addition to ASCII text files. Specific data archiving requirements for ongoing projects should be spelled out in the Standard Operating Procedures for each project.

11.3.2.2.5.: Spatial Data: Spatial data sets essential to the SIEN will be maintained in a format that remains fully-accessible by the current ArcGIS version. ArcGIS has maintained compatibility with previous data formats and, while shapefiles have retained functionality in ArcGIS, coverages may require conversion if they are no longer supported. At this time there is no practical way to save GIS data in a software or platform-independent format.

Both uncorrected and corrected GPS data (e.g., .ssf and .cor files) will be archived in their native format in addition to the corresponding GIS files that are created.

11.3.2.3.: Physical Objects and Records: The guidelines in this section apply to physical objects and records (e.g., draft and final reports, field notes, data collection forms, maps, photographs, audio and video recordings, natural and cultural history specimens) collected or produced by staff and cooperators.

The long term management of physical objects and records will require the close cooperation of the individual collectors (whether NPS personnel, contractors, or permitted researchers) and local park museum personnel. Much of the responsibility will rest with the collector who remains the sole source for all curated documentation and the principal authority for essential collection information.

Appendix 11H – Sierra Nevada Network Archive and Storage Locations identifies repositories for paper documents, and other archival objects. Each item will be stored in a suitable physical location appropriate to the object or document, which will depending on where it is in its lifecycle (Chapter 3). Each location cited in Appendix 11H is classified according to whether it can house temporary, permanent, or disposition suspended records or objects. A temporary record has value only for a specified period of time. A permanent record or object has enduring value and must be preserved forever. A disposition-suspended object or record should not be destroyed or transferred to the NARA. Depending on the classification, locations must meet certain physical and security conditions as described by the Records Management Handbook and Museum Handbook.

11.3.2.3.1.: Documents: NPS Directors Order 19, Section 4.1.1 states that anything classified as a federal record must be printed and managed in paper form. Such records include certain emails, reports, publications, maps, or other documentary materials. It is important for Network staff to understand that the NPS does not currently have an approved system for archiving electronic records. At a practical level, this means that with only few exceptions (see above), permanent records must be printed and organized in paper files.

A formal process is needed for ensuring that federal hardcopy records needing permanent preservation are handled properly. This is part of the disposition process. These records cannot simply stay in the office of creation or use for two reasons: 1) there are mandatory preservation requirements that should be handled by professionals; and 2) federal records of enduring value must, by law, be available to the public for research use. Given the existence of active local museum programs at SEKI and YOSE, the majority of collections derived from SIEN projects can be managed by their respective parks. DEPO collections will continue to be curated by SEKI.

The Technical Information Center (TIC) in Denver, Colorado, is the principal repository for planning documents for the NPS. The TIC provides temperature and humidity-controlled facilities, a professional archival staff, and meets all museum standards set by NPS. This repository may be used for technical reports for which off-site storage is considered worthwhile.

NPS Director's Order 19, Appendix B – Records Disposition Schedule has 40 pages devoted to classifying and filing NPS federal records and disposition specifications for each NPS federal record category. Generally, permanent federal records should be transferred to the NARA when they are 30 years old. To the extent possible, the SIEN will comply with the most current NPS disposition schedule.

11.3.2.3.2.: Natural History Specimens: The most explicit direction for the management of natural history collections is found in NPS Management Policies, Directive 4.2. It reads: "Specimens that are not authorized for consumptive analysis remain federal property and will be labeled and cataloged into the NPS cataloging system (ANCS+, or its successor) in accordance with applicable regulations (36 CFR 2.5)".

36 CFR 2.5 further states that collected specimens "*placed in displays or collections will bear official National Park Service museum labels and their catalog numbers will be registered in the National Park Service National Catalog"* and that the specimens and any derived data be made available to the public.

Natural history specimens collected under the auspices of the SIEN will be transferred to the Network park in which they were collected for either inclusion in the parks' museum collections or for transfer to a repository approved by the park. In this latter case, the park will enter into a formal, documented relationship with the repository and the specimens will be considered to be "on loan" with clearly stated terms and responsibilities. All loans are subject to the approval of the park superintendent.

Ideally, park curators will be involved at an early stage in the permitting or agreements process; early cooperation between curators and collectors will facilitate compliance with the current NPS Management Policies. Park curators can provide the collector with a pre-formatted Access database template, the use of which will greatly simplify eventual importation of catalog data into the Automated National Catalog System (ANCS+). At a minimum, parks will need such data in comma-delimited format (.cvs) for automated uploading into ANCS+.

11.3.2.3.3.: Photographs: While the SIEN is prepared to accept both digital and analog photographs, the particular photographic methods to be used by any given project will be developed jointly by the project leader and Network staff or cooperators. Early involvement of curatorial staff will insure that the specific preservation considerations of particular photographic methods are understood and considered.

For projects using analog methods, SIEN strongly encourages the use of Kodakchrome 35mm slide film which has the longest dark storage dye stability of any color film (Wilhelm and Brower 1993).

Slides should be labeled using indelible pigment, pH-neutral ink (e.g., Pigma Micron), or using laser-printed archival-quality slide labels. Labels must include the following information:

- Title – who or what is in the image
- Location – where the image was taken
- Date – when the image was taken
- Park Code
- Access Constraints – who may view the image
- Copyright information – restrictions on using the image
- Contact Information – who to contact for further information

Optional, but important information includes the geographic location with UTM NAD83 coordinates (SIEN mapping standard) and datum. All slides should be placed in archival, polypropylene slide sleeves and stored in the dark, preferably in an archival box.

Additionally, slides may be scanned and saved digitally in Tagged Image File Format (TIFF). If scans are used as the primary means of distributing or reproducing the images, the lifespan of the original slides will be greatly extended. Further recommendations regarding management of digital imagery can be found in Chapter 6 and Appendix 6D.
If photographic *prints* are desired, staff and cooperators are strongly encouraged to use true black-and-white films (e.g., Ilford Delta, Kodak T-Max or Tri-X). Prints and negatives will be stored in individual polypropylene sleeves and within archival boxes. Each print will be labeled on the back, using archival-quality labels that are either laser-printed or hand-labeled in pencil with the same information elements required for slides. If a contractor is submitting photographs, corresponding TIFF files may also be submitted.

Every image, regardless of format, will be entered into the SIEN photo database (guidance documents in development) where attributes such as electronic file name, keywords, project, photo description, photographer, date, and location are cataloged. All photo files and the associated photo database are housed on the archive portion of the SIEN server.

11.3.2.3.4.: Role of Curators: Park curators are a source of expertise, advice, and guidance on archiving and curatorial issues and have a role in almost all projects undertaken by the Network. Project leaders will involve park curators when projects are in the planning stage, to ensure all aspects of specimen preservation or document archiving are considered and any associated expenses are included in project budgets. Section 4.3.16 of Director's Order 24 requires "…project budgets to include funding for the basic management of collections that are project-generated". The costs for curation include cataloging, labeling, specimen preparation, initial storage of objects and specimens, and the organization and storage of project documentation.

11.4. References

Code of Federal Regulations. 2001. Title 36, Volume 1, Chapter 1, Part 2, Section 2.5 (36 CFR 2.5), U.S. Government Printing Office, Washington, D.C.

Mortenson, D. 2005. Data management plan for the Inventory and Monitoring Program, Southwest Alaska Network. U.S. National Park Service. Anchorage, Alaska. 234 pp.

National Park Service. 2000. Director's Order 24: NPS Museum Collections Management, U.S. Department of the Interior, National Park Service, Washington, D.C.

National Park Service. 2001a. Director's Order 19: Records Management. U.S. Department of the Interior, National Park Service, Washington, D.C.

National Park Service. 2001b. Director's Order 19, Appendix B: Records Disposition Schedule. Revised May, 2003. U.S. Department of the Interior, National Park Service, Washington, D.C.

National Park Service. 2005a (revised). Museum Handbook, Part II, Museum Records. U.S. Department of the Interior, National Park Service, Washington, D.C.

National Park Service. 2005b. Records Management Handbook. U.S. Department of Interior, National Park Service, Washington, D.C.

National Park Service. 2006. Management Policies 2006. U.S. Department of the Interior, National Park Service, Washington, D.C.

Olson, R. J., and R. A. McCord. 1998. Data Archival. Pages 53-57 in Michener, W. K., Porter, J. H., and S. G. Stafford. Data and information management in the ecological sciences: a resource guide. LTER Network Office, University of New Mexico, Albuquerque, NM. (available http://intranet.lternet.edu/archives/documents/data-informationmanagement /DIMES/html/ frame.htm.

Vogt-O'Connor, D. 2000. Planning digital projects for preservation and access. National Park Service Conserve O Gram 14:4. NPS Museum Management Program, Washington, DC. (available http://www.cr.nps.gov/museum/publications/conserveogram/19-21.pdf)

Wilhelm, H., and C. Brower. 1993. The permanence and care of color photographs: traditional and digital color prints, color negatives, slides and motion pictures. Preservation Publishing Company, Grinnell, Iowa.

11.5. Appendixes

6A) Sierra Nevada Network I&M Program Checklist of Project Specifications
6D) Sierra Nevada Digital Photo Management Strategy
11A) Electronic File Naming Guidelines and Standards
11B) Sierra Nevada Network Inventory and Monitoring Program Guidance and Rules for File Directory Structure
11C) Sierra Nevada Network Project Template Directory Structure
11D) Coversheet for Non-Digital Data listing Key Information
11E) National Park Service Standards and Guidelines for Backup, Storage, and Recovery
11F) Sequoia and Kings Canyon and Devils Postpile Network General Support Services Backup Activity Plan

11G) Yosemite Network General Support Services Backup Activity Plan
11H) Sierra Nevada Network Archive and Storage Locations

11.6. Credits

This chapter was written by Pat Lineback and Ward Eldredge (SEKI Archivist), with contributions from Dan Sohn (YOSE IT), Paul Gallez (YOSE Database Specialist), Bob Basham (SEKI IT), Mortenson (2005), Michael Grimes (NPS Records Officer), and Rosamonde Cook.

Chapter 12. Project Tracking and Documentation

12.1. Objectives

Develop and implement a comprehensive and cohesive procedure for tracking projects, including project status, data, and the products of analysis to support program coordination and annual reporting, and to improve accountability for Sierra Nevada Network natural resource inventory and monitoring efforts and products.

12.2. General Standards and Guidelines

12.2.1. Project Organization

All projects electronic files should be well organized in a project directory (Appendix 2H). The SIEN digital directory structure is organized at the project level, such that most or all digital files associated with a project are filed under a common root directory. Project file names will adhere to the naming conventions established for the Network (Chapter 11). Physical objects acquired as part of a project will be stored according to specification in Chapter 11 and its appendices.

12.2.2. Project Tracking

The SIEN I&M Project Tracking Database will be developed and will serve as the primary organizational tool for cataloging and searching information on I&M projects. Specifically, the purpose of this database is to:

- Maintain a list of projects – project managers and other users will be able to locate rapidly most project-related information (e.g., objectives, contact information, status, funding sources, and more) and to easily summarize this information for administrative reports.
- Track product deliverables – a comprehensive list of project deliverables will be maintained with information on permitting contracting, content and formatting specifications, delivery schedules, disposition schedules, expected due dates, and storage location. This information is generally specified at project initiation and updated throughout the course of a project.
- Manage project codes – These are intelligent alphanumeric codes used to tie together digital information in various, minimally connected NPS project tracking systems (e.g., RPRS, PMIS), along with analog materials that cannot otherwise be linked to an integrated information system. These codes are also used to link to databases and GIS themes, especially where information from multiple sources is stored together. See Appendix 12A for a list of Service-wide Project Tracking Systems.

The project tracking database will be hosted on a SIEN data server. Although primarily maintained by the SIEN Data Manager, the database will be available to project leaders and the Network Coordinator who will be able to update information as needed. The Data Manager will create custom query views and reports to help administrators manage projects, and to facilitate reporting on project status, accomplishments and delivered products.

12.2.3. Project Documentation

Project-specific protocol narratives and SOPs are the principle means by which I&M projects will be documented. Standards for these documents have been developed by the national I&M Program and are discussed for the SIEN in Mutch et al. (2006). These documents must always accompany the distribution of monitoring data.

12.2.3.1. Protocol and SOP Version Changes: Over time there will be instances when protocol narratives and SOPs need to be updated, and these updates may occur independently of each other. That is, a change in one SOP will not necessarily invoke changes in other SOPs; a narrative update may not require SOP modifications. The SIEN Project Tracking Database will track the project narrative and SOPs by version number. This information will be updated whenever any narrative or SOP document is modified. The protocol narrative and SOPs will not be distributed without a log of changes from the Project Tracking Database.

12.2.3.2. Additional Project Metadata: Long-term monitoring projects may require documentation beyond the scope of established FGDC standards and the Sierra Nevada Network Database Documentation Template (Appendix 5A). Documentation should be provided for algorithms, output files, and analytical products which may reside in different systems and formats, and could potentially be overlooked when distributing or applying project data. Data use and data request histories, and information on secondary research or publications resulting from long-term monitoring projects, should also be maintained. The Project Tracking database potentially could be used to organize this information as well. Specific methods for documenting this information will be tested for ease of use and are likely to evolve over time as needs become apparent and new solutions are developed.

12.3. References

Mortenson, D. 2005. Data management plan for the Inventory and Monitoring Program, Southwest Alaska Network. U.S. National Park Service. Anchorage, Alaska. 234 pp.

Mutch, L., Rose, M., Heard, A., Cook, R. R., and G. Entsminger. 2006. Sierra Nevada Network Vital Signs Monitoring Plan, Draft Phase III Report, National Park Service Inventory & Monitoring Program. U.S. National Park Service, Three Rivers, California.

12.4. Appendixes

2H) SIEN Directory Structure
5A) Sierra Nevada Network Database Documentation Template
12A) NPS Service-wide Project Tracking Systems

12.5. Credits

This chapter was written by Rosamonde Cook, with information provided by Mortenson (2005).

Chapter 13. Implementation

The data management plans for each of the 32 I&M Networks are the first comprehensive documents of their kind in the NPS and contain practices that may be new to staff and cooperators. However, almost every requirement stems from federal law, Executive Orders, Director's Orders, or national I&M Program guidance. The DMP helps put these requirements into context, and provides operational guidance for achieving them.

A DMP *ad hoc* steering team of 15 members was established during the summer of 2006 to help guide plan development. It comprises staff from Sequoia and Kings Canyon National Parks, Yosemite National Park, and U.S. Geological Survey field stations in Yosemite and Sequoia. This team will assist in the development of goals, roles and responsibilities, sideboards, and best data management practices, and aid development of specific DMP policies. Team members are expected to be the primary advocates for implementation of this plan and, in some cases, will help transform existing business practices for certain programs.

13.1. Education and Training

Implementation will require education and training in order to familiarize park staff and cooperators with the tools, procedures, and guidelines outlined in the plan. Formal (training sessions) and informal (one on one communication and assistance) methods will be used. These efforts will begin in 2007 and be led, at least initially, by I&M data management staff and the SEKI GIS coordinator, with participation by interested parties at all parks actively encouraged.

13.2. Milestone Goals

Goals for the first 3 years will include:

- Acceptance and understanding by all staff of targeted programs and their cooperators of the fundamentals of data and information management, including
    - File management
    - Documentation
    - Quality assurance and quality control
    - Electronic storage
    - Archive storage
- Improvement of data management practices by implementing
    - accepted database design standards
    - thorough testing of databases, data collection methods, and their integration prior to field work
    - quality assurance and control procedures at every stage of project development
- Common SOPs and guidance documents for multiple protocols
- Inclusion of detailed specifications for data management consistent with the DMP in all protocols developed for the Vital Signs Monitoring Program
- Development of procedures and outlets for communication within and among Network parks and with the public as described in this plan

Beyond the first three years, goals should include the development and assessment of:

- Methods for improving file management (e.g., a content management system), database administration and security (e.g., migration to SQL-Server), integration into the network of off-site users, and other needs identified in the DMP
- Procedures to facilitate the summarization and reporting of monitoring data to staff, cooperators, and the public in formats appropriate to each audience
- Framework and gateway for integration of monitoring data with other initiatives both within the NPS and with outside agencies
- Methods for improving information management (e.g., basic content management services including library services, search tools, and decentralized web administration), database administration and security (e.g., migration to SQL-Server), integration into the network of off-site users, and other needs identified in the DMP

Full implementation of this plan for SIEN budgeted projects will be achieved by 2010. Implementation at the broader park level is expected to be an ongoing process and will most likely occur in stages. Realistically, it will not be possible for all programs to fully adopt the practices and procedures recommended without an increase in current levels of funding and staff support. Currently, none of the Network park resources divisions have a data manager on staff, although Yosemite may add a position in the future. Programs wishing to participate should identify those areas that can be implemented given existing resources. For example, it should be relatively easy for most programs to begin implementing standards and guidelines for organizing and managing electronic files, including digital photos. Pragmatism should dictate prioritization and implementation. Critically important to successful implementation at the broader park level will be the active support of park resource management teams.

13.3. Revisions

The next plan revision should be completed within three years of plan approval, or by October 1, 2010, and then every five years afterward. Plan appendices, including SOPs, detailed guidelines, reference manuals, policy statements, etc., will likely require more frequent updates to account e.g., for changes in technology or availability of better information. These updates should be completed as needed.

13.4. Credits

This chapter was written by Rosamonde Cook and Pat Lineback.

Glossary

ArcIMS, ArcSDE, (Arc Internet Map Server and Arc Spatial Database Engine) are ESRI GIS software applications that provide tools for managing large, complex data sets within a relational database management system – ArcSDE, and serving the data over the Internet - ArcIMS.

Best practices - A technique, methodology, or standard operating procedure that, through experience and research, has been proven to reliably lead to a desired result.

Biodiversity Data Store is a digital online repository of documents, GIS and other data sets that contribute to the knowledge of biodiversity in National Park units, including presence/absence, distribution and abundance of animal and plant species. Along with NatureBib and NPSpecies, comprises the NPS Biodiversity Information System.

Certified data and metadata – Completed data and documentation for short-term projects, or one season of completed data for long-term monitoring projects. Certification is a confirmation by the project leader that data have passed all quality assurance requirements and are complete and ready for distribution. Metadata records include the detailed information about project data needed for proper use and interpretation.

Conceptual data model is a detailed model that shows the overall structure of organizational data, independent of any particular database management system or other implementation considerations. Conceptual data models are typically produced as the first step in system design, often presented as entity relationship diagrams, and frequently a precursor to the logical data model.

Data refers to observational information gathered directly (in the field) from a natural resource via specific protocols and organized for analysis, summary or reporting. Data may either in analog or digital form (generally stored either on paper or a variety of computer-compatible media), though the latter is encouraged where ever feasible. Data may exist in several states (conditions) including 'raw', 'validated' and 'analyzed'. 'Analyzed' data includes 'reported' or 'summarized' data and may represent 'information' as a final form of the data from which decisions or conclusions may be made. Ultimately, data are intended to contribute to the knowledge and decisions regarding the conditions, processes, and changes within the ecosystem.

Data set can best be considered a convenient grouping of data, or individual observations, such that the summary of the information will be meaningful to prospective users.

Edit log – A means of tracking changes to certified data.

Document – Recorded information regardless of physical form or characteristics. Often used interchangeably with record (From DO19 glossary).

Ecological indicators are a subset of the physical, chemical, and biological elements and processes of natural systems selected to represent the overall health or condition of the system.

Glossary (continued)

Entity relationship diagrams (ERDs) are high-level data models that are useful in developing conceptual designs for databases. Creation of an ER diagram, which is one of the first steps in designing a database, helps the designer(s) to understand and to specify the desired components of the database and the relationships among those components. An ER model is a diagram containing entities or "items", relationships among them, and attributes of the entities and the relationships.

Goal - The general ends toward which this data management plan are directed.

Inventories are "an extensive point-in-time effort to determine location or condition of a resource, including the presence, class, distribution, and status of plants, animals, and abiotic components such as water, soils, landforms, and climate. Inventories contribute to a statement of park resources, which is best described in relation to a standard condition such as the natural or unimpaired state. Inventories may involve both the compilation of existing information and the acquisition of new information. They may be relative either to a particular point in space (synoptic) or time (temporal)." (Source: http://www .nature.nps.gov/im/monitor/index.htm)

Legacy data existing data that has been acquired by a park unit and still has importance, but which is at risk of becoming obsolete because of software limitations or available metadata.

Local archives and digital library – Local storage of copies of data, metadata and other products generated by projects. Archives are for hard-copy items and off-line storage media, whereas the digital library is maintained live on a server.

Logical data model (LDM) is an abstract representation of a set of data entities and their relationships, usually including their key attributes. It may omit non-key attributes or use abstractions of actual types, and may include physical tables that are represented by relationships or as aggregates within a larger entity, depending on the notation used and the level of abstraction. The logical data model is intended to facilitate analysis of the function of the data design, and is not intended to be a full representation of the physical database. It is typically produced early in system design, and it is frequently a precursor to the physical data model that documents the actual implementation of the database.

Master database – Central repository for project data, used for viewing, summarizing, and analysis. Contains only data that have passed all quality assurance/quality control.

Metadata is information about data. A metadata document contains specific and detailed information about a data set, including who, what where, when, why and how the data were collected, analyzed, or manipulated.

NatureBib is the National Park Service bibliographic database. NatureBib is designed to work with NPS Focus to help staff identify, locate, and obtain NPS natural resources documents.

Glossary (continued)

National databases and repositories – Applications and repositories maintained at the national level, primarily for the purpose of integration among NPS units and for sharing information with cooperators and the public.

Natural Resource Database Template (NRDT) is a core set of database tables that serves as a foundation for building relational databases for the NPS Inventory and Monitoring Program.

Nonspatial data are data that do not include reference to geographic location and shape of, and relationships among, geographic features.

NPS Data Store an online graphical search interface that links dataset metadata to a searchable data server on which datasets are organized by NPS units, offices and programs.

NPS Focus is the National Park Service online digital library designed to provide a venue for researching and downloading natural resources documents and other digital imagery.

NPS Data Clearinghouse – the central repository for NPS GIS data available to the public. Implemented through the NPS Focus gateway.

NPS Focus (Digital Library and Research Station) – a decentralized digital imagery and data management system, implemented through a central Internet portal sponsored by the NPS Office of the Chief Information Officer. Includes to the NPS Data Clearinghouse.

NPSpecies is the master species biodiversity database for the NPS. The database lists the species that occur in or near each park, and the physical or written evidence for the occurrence of the species (e.g., references, vouchers, and observations). NPSpecies is implemented online through secure and public servers.

NPSTORET is the NPS version of the U.S. EPA STORET database, used for transferring water quality data collected by NPS programs to STORET.

Object – Physical items such as natural history specimens, cultural artifacts, photographs, audio, and video tapes.

Objective - Specific statements describing the results to be achieved.

Ortho-rectification is the process of correcting an image for distortions due to camera lens, sensor non-verticality, and terrain, a process collectively included in photogrammetry.

Orthophoto is a digital image in which the pixels have been corrected to an orthogonal projection, produced by removing errors due to tilt and relief displacement.

Glossary (continued)

Physical data model is a representation of a data design which takes into account the facilities and constraints of a given database management system. In the lifecycle of a project it is typically derived from a logical data model, though it may be reverse-engineered from a given database implementation.

Protocol (or Monitoring Protocol) refers to the formal documents and sampling processes that describe how a vital sign will be monitored; composed of a narrative section, standard operating procedures and supplementary information (databases, reports, tools, hardcopy materials).

Raw Data are environmental measurements that have been recorded but not subjected to any quality assurance or control beyond those applied during field work. Raw data are typically recorded on field sheets or in handheld computers, but may also include the products of remote sensing (aerial photographs and spectrophotometry).

Relational Database Management System (RDMS/RDBMS) is a type of database management system (DBMS) that stores data in the form of related tables.

Sensitive data - Any information, which through loss, unauthorized access, or modification could adversely affect the national interest, the conduct of federal programs, or the privacy of individuals.

Spatial data – Information about the location and shape of, and relationships among, geographic features, usually stored as coordinates and topology. Can be stored in tabular or GIS file format.

Standard Operating Procedures (SOPs) are detailed step-by-step instructions that outline a formal set of procedures for performing specific tasks.

STORET is a database application maintained by the U.S. Environmental Protection Agency that contains raw biological, chemical and physical data on surface and ground water quality collected by federal, state and local agencies, Indian Tribes, volunteer groups, academics, and others.

Structured Query Language (SQL) is a language used to interrogate and process data in a relational database. Originally developed for mainframes, all database systems designed for client/server environments support SQL. SQL commands can be used to interactively work with a database or can be embedded within a programming language to interface to a database. Programming extensions to SQL have turned it into a full-blown database programming language, and all major database management systems (RDBMS s) support the language. There is an ANSI standardized SQL, but most database management systems (RDBMS s) have some proprietary enhancement, which if used, makes SQL non-standard.

Tabular Data – Data organized into logical tables. Refers to the format of data sets rather than content. Tables usually contain information arranged in rows and columns.

Glossary (continued)

Validated data are data that have been verified according to the standard operating procedure under which the data were gathered (typically the protocol for a given project) and are deemed ready for reporting and/or analysis.

Vital Signs are ecological elements or processes chosen to represent the overall health or condition of park ecosystems, known or hypothesized effects of environmental stressors, or elements of value to humans, and are the subject of long-term monitoring by the NPS I&M Program.

Working database – A project-specific database for entering and processing data for the current season (or other logical period of time). May also constitute the master database for very short-term projects

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-45, November 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