

National Park Service
U.S. Department of the Interior



Natural Resource Program Center

Data Management Plan – Southeast Alaska Network

Natural Resource Report NPS/SEAN/NRR—2008/058



ON THE COVER

Sunset at Bartlett Cove, Glacier Bay National Park and Preserve.
Photograph by: Craig Murdoch, NPS.

Data Management Plan – Southeast Alaska Network

Natural Resource Report NPS/SEAN/NRR—2008/058

William F. Johnson and Brendan J. Moynahan
National Park Service
Southeast Alaska Inventory and Monitoring Network
3100 National Park Road
Juneau, Alaska 99801

September 2008

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

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

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

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

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

Please cite this publication as:

Johnson, W. F., and B. J. Moynahan. 2008. Data management plan: Southeast Alaska network. Natural Resource Report NPS/SEAN/NRR—2008/058. National Park Service, Fort Collins, Colorado.

Acronyms

DMP	Data Management Plan
DOI	Department of the Interior
GIS	Geographic Information System
GLBA	Glacier Bay National Park and Preserve
I&M	Inventory and Monitoring Program
IRMA	Integration of Resource Management Applications
KLGO	Klondike Gold Rush National Historical Park
NPS	National Park Service
NRDT	Natural Resource Database Template
SEAN	Southeast Alaska Network
SITK	Sitka National Historical Park
SOP	Standard Operating Procedure
VSMP	Vital Signs Monitoring Plan
WASO	Washington Support Office, Fort Collins, CO

Document Change History

Version Date	Author	Nature of Change
1/8/2008	Bill Johnson	Initial outline
2/16/2008	Bill Johnson	Outline overlaid with narrative
2/25/2008	Bill Johnson	Draft for VSMP
3/10/2008	Bill Johnson	Appendix A added
4/15/2008	Bill Johnson	Appendix B added
9/30/2008	Bill Johnson	Revised following peer review
10/15/2008	Bill Johnson	Formatted for publication

Contents

Page

EXECUTIVE SUMMARY.....	xiii
CHAPTER 1. INTRODUCTION	1
1.1 PLAN CONTEXT	1
1.2 PLAN PURPOSE	1
1.3 PLAN AUDIENCE	1
1.4 ROLE OF INVENTORY AND MONITORING	1
1.5 MISSION OF THE SEAN	2
1.6 THE SEAN DOMAIN	2
1.7 THE SEAN CUSTOMERS	3
1.8 DATA MANAGEMENT NETWORK DESIGN PHILOSOPHY	5
1.9 PLAN REVISIONS.....	9
CHAPTER 2. INFRASTRUCTURE AND SYSTEMS ARCHITECTURE	11
2.1 THE SEAN OVERALL ARCHITECTURE.....	11
2.2 STANDARD COMMERCIAL APPLICATIONS.....	12
2.3 ENTERPRISE APPLICATIONS	13
2.4 IP NETWORK AND WEB SERVER INFRASTRUCTURE.....	13
2.5 DATABASE SERVERS.....	16
2.6 MID-TIER	16
2.7 FILE STORAGE	16
2.8 INTER-NETWORK SHARED RESOURCES.....	16
CHAPTER 3. PROJECT MANAGEMENT AND DATA PROCESSING WORKFLOW	19
3.1 PROJECT MANAGEMENT	19
3.2 DATA PROCESSING WORKFLOW CONCEPTS	20
3.3 WORKFLOWS DIFFER BY PHASE	21
3.4 THE SEAN DELIVERY WORKFLOW	22
3.5 THE SEAN ACQUISITION WORKFLOW	23
CHAPTER 4. DATA MANAGEMENT ROLES AND RESPONSIBILITIES.....	25
4.1 GROUP ROLES.....	25
4.2 DEVELOPMENT ROLES DIFFER MARKEDLY FROM OPERATIONAL ROLES	26
4.3 SPECIFIC STAFF ROLES	26
4.4 SPECIFIC RESPONSIBILITIES OF PARTICIPANTS.....	27
CHAPTER 5. DATABASES	31
5.1 DEFINING “DATABASE” FROM A HOLISTIC PERSPECTIVE	31
5.2 DATABASE CONTENT IS STORED IN SPECIFIC REPOSITORIES	33
5.3 IMPLEMENTING THE AUXILIARY REPOSITORY	33
5.4 REPOSITORY ISSUES FACED IN EACH CORE FUNCTION.....	35
5.4.1 <i>Core Function 1: Information Dissemination.....</i>	<i>35</i>
5.4.2 <i>Core Function 2: Information Repository.....</i>	<i>36</i>
5.4.3 <i>Core Function 3: Information Certification</i>	<i>37</i>
5.4.4 <i>Core Function 4: Information Acquisition.....</i>	<i>38</i>
5.5 DATABASE CREATION WORKFLOW	39
5.6 INTER-ALASKA DATABASE COOPERATION OFFERS BENEFITS.....	39
CHAPTER 6. DATA ACQUISITION, PROCESSING, AND REPORTING	41
6.1 THE SEAN’S DATA FLOW	41

6.2	A DISTINCT DEVELOPMENT PHASE PRECEDES THE OPERATIONS PHASE	42
6.3	THE DEVELOPMENT PHASE	42
6.4	THE OPERATIONS PHASE	43
6.5	LEGACY DATA ACQUISITION	46
6.6	DATA ANALYSIS AND REPORTING	47
CHAPTER 7.	QUALITY ASSURANCE AND QUALITY CONTROL.....	49
7.1	PRACTICAL QUALITY ASSURANCE	49
7.1.1	<i>Quality Assurance for Paper-Based Tabular Data Acquisition.....</i>	<i>49</i>
7.1.2	<i>Quality Assurance for Computer-based Data Acquisition.....</i>	<i>49</i>
7.1.3	<i>Quality Assurance Considerations for Using Data from Multiple Sources in One Project Database</i>	<i>50</i>
7.2	QUALITY CONTROL	50
7.2.1	<i>Effective Quality Control is Based in Validation Criteria</i>	<i>50</i>
7.2.2	<i>Quality Control for Tabular Time-Series Data Products</i>	<i>51</i>
7.2.3	<i>Quality Control for Non-tabular Products</i>	<i>52</i>
CHAPTER 8.	DATASET DOCUMENTATION	53
8.1	TABULAR DATA WILL COMPLY WITH THE 8-LEVEL METADATA STANDARD	53
8.2	EXTERNAL DATA CONSIDERATIONS	53
8.3	DATA DEFINITIONS MAY BE PROSCRIBED BY A PARTNER.....	53
8.4	METADATA FOR NON-TABULAR PRODUCTS	53
CHAPTER 9.	DATA OWNERSHIP AND SHARING	55
CHAPTER 10.	DATA DISSEMINATION.....	57
10.1	DISSEMINATION IS A HIGH PRIORITY CORE FUNCTION.....	57
10.2	THE DEVELOPMENT PHASE	63
10.2.1	<i>Create Web Site Objects for Basic Information Delivery</i>	<i>63</i>
10.2.2	<i>Deploy Objects to Production Web Site.....</i>	<i>64</i>
10.3	THE OPERATIONS PHASE	64
CHAPTER 11.	RECORDS MANAGEMENT, DATA MAINTENANCE, AND ARCHIVING.....	67
11.1	DELIVERABLE DATA PRODUCTS MUST BE IN DIGITAL FORM	67
11.2	DIRECTIVES DESIGNATE SOME PROJECT RECORDS AS MISSION-CRITICAL	67
11.3	PRESERVING THE SEAN REPOSITORIES	67
11.4	LOCAL PHYSICAL ARCHIVING	68
11.5	“VERSION OBSOLESCENCE” MUST BE CONTINUALLY GUARDED AGAINST	68
11.6	ARCHIVING TECHNOLOGIES ARE UNDERGOING RAPID CHANGES	68
11.7	THE VALUE OF MONOLITHIC DATABASES FOR HOUSING REPOSITORIES	69
CHAPTER 12.	PROJECT TRACKING AND DOCUMENTATION	71
12.1	SCOPE OF PROJECT TRACKING RESPONSIBILITY	71
12.2	TRACKING IS BY SPREADSHEET	71
12.3	CURRENT TRACKING SHEET FORMAT.....	71
CHAPTER 13.	IMPLEMENTATION.....	73
13.1	IMPLEMENTATION REQUIRES MULTIPLE PASSES TO COMPLETE	73
13.2	STEP ONE – ESTABLISH THE CORE STRUCTURE USING SIX VITAL SIGNS.....	73
13.3	STEP TWO –ESTABLISH MATURE DATA MANAGEMENT STRUCTURE	74
13.4	STEP THREE –ENHANCE REPOSITORY AND DISSEMINATION CORE FUNCTIONS	75
13.5	STEP FOUR - ONGOING OPERATIONS	75
13.6	ESTIMATED WORKFORCE COMPOSITION FOR IMPLEMENTATION	75
LITERATURE CITED.....		77

APPENDIX A – DATA FLOW DIAGRAMS	A-1
APPENDIX B – STANDARD OPERATING PROCEDURES	B-1

Tables

Page

Table 1.1. The SEAN customers and their interests in using the network.....	3
Table 2.1. Facilities required to support the system architecture, by location.....	12
Table 2.2. NPS applications available to store and disseminate vital sign-related information.	13
Table 4.1. Staff requirements by Core Function. Besides the four Core Functions identified in the SEAN's underlying model, management and development functions are also included.	27
Table 4.2. Task responsibilities for each required staff role.....	28
Table 5.1. Target repositories for specific deliverables.....	33
Table 5.2. Possible auxiliary repository architectures considered, in order of desirability.	34
Table 5.3. Specific file types acceptable to the SEAN auxiliary repository.....	37
Table 6.1. Development tasks necessary to fulfill information repository, certification, and acquisition for each project.....	42
Table 6.2. Operations tasks needed to fulfill information Acquisition, Certification, and Repository Core Functions.....	44

Figures

Page

Figure 1.1. Satellite image illustrating SEAN’s geographic domain.	2
Figure 1.2. Map showing SEAN in relation to other networks in Alaska.	3
FIGURE 1.3. CONTEXT LEVEL DATA FLOW DIAGRAM. ARROWS POINTING INTO A BOX INDICATE THAT THE GROUP ACCEPTS INFORMATION FROM THE NETWORK. ARROWS POINTING OUT OF A BOX INDICATE THE GROUP PROVIDES INFORMATION TO THE NETWORK FOR SUBSEQUENT STORAGE AND/OR ACTIONS.	5
Figure 1.4. The SEAN Core Functional Model overview. Dissemination services (red) provide all data products to all customers using Internet web servers. Repositories (yellow) store the certified inventory and monitoring products called for in the protocols. Certification processes (green) assure repositories have the highest quality data and that sensitive items are restricted to authorized users. Data acquisition processes (blue) led by park staff and cooperators include a wide array of tasks, ranging from designing field forms to producing talks on long-term trends.	7
Figure 2.1. The SEAN overall architecture. In operating order, data acquisition (blue) provides deliverables to validation (green). Certified products are installed in the appropriate repository (yellow). Customers find the data using web browsers and servers (red).....	11
Figure 2.2. Details of the specific equipment and objects the SEAN directly employs to manage its data. Associated functional areas are denoted using the four colors consistently employed throughout the plan. Items in gray, however, have been added to represent development tools. Development tools are essential but are not associated with one single functional area.	15
Figure 3.1. The four stages of SEAN project management within the context of two phases: development and operations.	19
Figure 3.2. The SEAN delivery architecture detail; a partial decomposition of Figure 2.1.....	23
Figure 3.3. The SEAN acquisition architecture detail; a partial decomposition of Figure 2.1.	23
Figure 4.1. The SEAN provider roles are applied to Core Functions. Providers are Partners, I&M WASO, Data Manager, and Project Leaders.	25
Figure 5.1. Depiction of the SEAN's "database" in the broad sense. Also shown is the relationship of the customers to the database. Red indicates the Dissemination Core Function and yellow corresponds to the Repository Core Function.....	32
Figure 6.1. Level 1 data flow diagram illustrating participants, processes, and stores. Participating roles are shown in boxes. The specific data they generate or require are identified by labeled directional arrows. The circles denote high level processes that connect the consumer to the information. The “Data Stores” symbol in the center holds the SEAN broadly-defined database of Chapter 5.	41
Figure 6.2. Sequence of steps involved in the Certification Core Function.	46
Figure 10.1. The SEAN home page content mock-up, illustrating the planned scope.	59
Figure 10.2. Vital sign main page concept example.....	60
Figure 10.3. Example of an I&M enterprise application that may be used to provide web access to SEAN information and data: NPS Data Store.	61

Figure 10.4. Example of a possible partner’s repository for the SEAN information and data. Shown is a typical web dissemination page maintained by Western Region Climate Center.....	62
Figure 10.5. Results of a localized Google search for SEAN background documents.....	63
Figure 13.1. Estimated SEAN development skill set requirements.....	76
Figure 13.2. Estimated SEAN operations skill set requirements.....	76

Executive Summary

The Southeast Alaska Network (SEAN) conducts and coordinates long-term ecological monitoring for and with Glacier Bay National Park and Preserve (GLBA), Klondike Gold Rush National Historical Park (KLGO), and Sitka National Historical Park (SITK). The SEAN has made a fundamental commitment to rigorous management of monitoring data, so data are available and accessible to a wide array of park and non-park audiences. This Data Management Plan (DMP) is a companion to both the SEAN Vital Sign Monitoring Plan (VSMP [Moynahan and Johnson 2008]) and the Inventory & Monitoring (I&M) National Data Management Plan (National Park Service. 2008). The SEAN VSMP prescribes the approach to long-term monitoring of specific Vital Signs. The national DMP provides and justifies overarching strategy and direction for data management components of the Vital Signs Inventory and Monitoring Program. The SEAN DMP, including the standard operating procedures in Appendix B, provides detailed specifications for establishing and fulfilling the network's information management role.

Introduction

Chapter 1 provides the background of the SEAN, its data management needs, and its foundational model for information management. The plan calls for first determining exactly what data products are needed to meet NPS goals, and then developing processes to deliver those products. It is essential that project protocols specify the data deliverables in advance and in detail.

The SEAN's conceptual foundation for its data management responsibilities is an information model that identifies four Core Functions. Core Function 1, "Information Dissemination," provides all network products to all customers using Internet web servers. Core Function 2, "Information Repository," securely stores the data products defined in the protocols. Core Function 3, "Information Certification," assures repositories have the highest quality data, with sensitive data secured. Core Function 4, "Information Acquisition," is led by park staff and covers everything from designing data collection forms to producing talks on long-term trends. Two organizing principles for the DMP are (1) all certified vital signs data will be web-accessible, and (2) vital signs data will be managed under an "authoritative source" concept.

Infrastructure and Systems Architecture

Chapter 2 describes the process by which an operational architecture is mapped onto the SEAN's data management philosophy. The SEAN will conform to all applications software standards required by DOI and NPS. Information dissemination services are provided to all customers by Internet web servers maintained by Washington Support Office (WASO) and, for some vital signs, external partners. Information repositories are physically located at WASO I&M, outside agency partners, curatorial collections, and the SEAN itself. The Information Certification Core Function uses the SEAN file server facilities for staging data. Information Acquisition processes use park file servers, networked workstations, and field acquisition devices. The SEAN plan details exactly how these architectural elements interact.

Project Management and Data Processing Workflow

Chapter 3 describes a project management model with two-phases: Development and Operations. We describe distinct workflows for both development and operation of each vital sign monitoring project. Development begins at defining end-products and working back to data point observation. Conversely, operation begins with data acquisition and culminates in dissemination. Thoughtful execution of these distinct yet complementary phases will ensure end products meet initial expectations, and that data collection meets the pre-defined needs.

Data Management Roles and Responsibilities

Chapter 4 identifies the roles of personnel and entities involved in the SEAN's data management process. For dissemination and associated repository operations, both WASO I&M and selected partners are used. SEAN staff is responsible for the development side of dissemination, such as building project web pages. Most Information Certification Core Function tasks, including validating that products meet specifications and installing certified data products in repositories, are done by the SEAN Data Manager. Information acquisition is the responsibility of the Project Leader, as is issuing final certification. The Data Manager and Project Leader share responsibility for certain limited tasks.

Initially, more resources are required for development and fewer for operations. Once the initial vital signs projects are established, development needs drop substantially.

Databases

Chapter 5 identifies and describes the types and locations of anticipated SEAN databases. Database contents include both tabular and non-tabular data. The main I&M repository is used to house network data whenever possible. When appropriate, data deliverables should be designed to fit within the constraints of NPSpecies, NatureBib, and other enterprise applications. Tabular data, in particular, should have specifications that allow them to be saved in the I&M's generic data repository, NPS Data Store.

Repository files are periodically refreshed and version-upgraded, so they may be readable after new versions of software get introduced. To meet data accessibility requirements over an extended period, supported file types are limited in number. The I&M NPS Data Store is used to contain all SEAN metadata, regardless of where the authoritative source data are actually housed. Even data sets entrusted to partners have their metadata recorded in NPS Data Store. Tabular data are collected in MS Access databases designed specifically to support the products of the project. The SEAN develops the database objects based on the Natural Resource Database Template (NRDT). The SEAN also programs data entry and edit screens, based on the NRDT "front end builder" tool set. Actual operation of Access databases is the responsibility of the Project Leader.

Data Acquisition, Processing, and Reporting

Data acquisition, processing, and reporting procedures are described in Chapter 6. Data acquisition refers to the collection and organization of field data and the creation of deliverable data products from them. Processing refers to steps that formalize, validate,

redact and certify the proposed products. Reporting includes installation of certified products into repositories. The remaining Core Function, Dissemination, is described in detail in Chapter 10.

Quality Assurance and Quality Control

Chapter 7 summarizes quality assurance and quality control (QA/QC) procedures. Quality assurance steps are taken to encourage accurate and complete data acquisition. These steps include thorough training of collection staff, verification, careful selection of field devices, and others.

Quality control is accomplished by formally validating every data item submitted. All validation criteria are enumerated in detail in advance in the protocol, both for tabular and non-tabular deliverables. Predefined validation criteria include mandatory and optional items. Failure to meet any mandatory criterion, after automated analysis of a submission unit, causes the entire submission to be returned to the Project Leader for corrections. Where applications permit, every row of a repository update is stamped with its date, protocol version, and the user id of the person making the changes.

Dataset Documentation

Chapter 8 outlines documentation procedures intended to attach metadata to all certified vital signs data sets. All data capable of being documented using the NPS-modified 8-level Federal Geographic Data Committee (FGDC) metadata standard have their metadata stored in NPS Data Store. This occurs regardless of whether a project includes a GIS component, or whether the data are actually stored in NPS Data Store.

The network's Internet web site is intended to provide access to considerable documentation for each vital sign including tracked deliverables, exception logs, protocols, abstracts, etc. Ample documentation is provided so data can "stand on their own."

Data Ownership and Sharing

Chapter 9 presents SEAN's principles of data ownership that are integrated into Cooperative Agreements, contracts, and data dissemination processes. Network data are the property of NPS. NPS data are not copyrighted and are in the public domain. Park Superintendents and network staff collaborate on assigning sensitive designations to specific items. Partners do not house sensitive data. The network Data Manager responds to Freedom of Information Act (FOIA) requests for validated data as directed. FOIA requests for draft and incomplete products are serviced by the Project Leader responsible for those products.

Data Dissemination

Chapter 10 presents the steps that the SEAN takes to act on its philosophy that dissemination of high quality products drives all other data management tasks for the network. All deliverables are digital and accessible through Internet web browsing. Each vital sign has its own main page offering the deliverables specified in the protocol. The most useful vital sign information is directly provided by hyperlinks set up by SEAN staff. Customers may also use SEAN's search appliance to browse through historical or

background files to find details of specific individual interest. The “push model” was replaced with a “pull model” of data access.

Data are disseminated only from designated repositories. The repositories contain one and only one copy of a deliverable, which are loaded only after candidate items pass all validations. In this way, SEAN ensures that data always meet the measureable standards defined in their protocol.

Records Management, Data Maintenance, and Archiving

Chapter 11 describes SEAN’s approaches to the considerable challenges of managing and archiving data over the long term. SEAN repositories all have the capability to archive digital items. All deliverables must be provided as computer files of an acceptable format as specified in the protocols. Use of I&M enterprise application repositories are particularly attractive because their data are professionally and reliably archived. SEAN-held products are backed up using best practices, and comprehensive permanent-retention backups are made for archival purposes. Version obsolescence of digital data is continually monitored and remediated. In the future, using a comprehensive central database management system may improve the long-term viability of monitoring data.

Project Tracking and Documentation

In Chapter 12 we outline a method for using spreadsheets to track progress in building the specified products. Each project has a tracking sheet, and deliverable tracking is updated only by SEAN staff. The tracking sheet is always available for public viewing as a link on the corresponding project web page. In the future, a database approach may enhance project tracking.

Implementation

In Chapter 13, we propose a multi-step implementation that provides essential functionality first. Subsequent enhancements are added without impacting existing service. The first step develops substantial new processes and is the most time consuming. Operations begins after development is well underway. Operations will eventually become the largest consumer of resources.

The Data Manager is expected to be involved in protocol development to ensure that deliverable data products are articulated, described, and technically defined in detail. The Network Coordinator and Data Manager intend to achieve effectiveness by closely coordinating their efforts.

Acknowledgements

Of particular help to this plan were those very talented NPS data managers who developed the structure and content of what is generally known as the “National Data Management Plan.” The SEAN plan also owes particular thanks to the helpful assistance of Margaret Beer, Bill Eichenlaub, Scott Miller, and Dorothy Mortensen; four experts in the field.

Chapter 1. Introduction

1.1 Plan Context

The Southeast Alaska Network (SEAN) Data Management Plan serves as a companion to the NPS National Data Management Plan (National Park Service 2008). The national plan provides and justifies overarching strategy and direction. The SEAN plan, including the standard operating procedures in its appendix, provides detailed specifications for developing and fulfilling the network's information management role. Both documents share the same topic structure and chapter order, so, for example, Chapter 5 of this plan implicitly references Chapter 5 of the national plan. Those statutes, regulations, and agency policies that authorize and regulate network data management which are already comprehensively cited and referenced in the national plan will not be explicitly cited in this document.

The SEAN DMP also conforms with and is an attachment to the SEAN Vital Signs Monitoring Plan (VSMP, Moynahan and Johnson 2008).

If a specification made in the SEAN Data Management Plan conflicts with a specification in the National DMP or the VSMP, then precedence is given to the latter plans.

1.2 Plan Purpose

The SEAN Data Management Plan explains specific needs to be addressed by the SEAN's inventory and monitoring (I&M) data management system over the next five years. It then details the steps required to build and operate a system to meet those needs. The plan was formulated with the goal of making valuable information available to managers and researchers over a 100 year period. Therefore, particular emphasis has been given to the special needs of long-term monitoring data sets.

1.3 Plan Audience

The SEAN staff and management use this plan to direct and prioritize ongoing operations. Vital signs monitoring Project Leaders use the plan to understand their data management responsibilities as well as the responsibilities of other participants in the program. Partners and cooperators use the plan to understand the program context they are operating in, including specific expectations for the services they will deliver. I&M WASO management relies on the plan to ascertain the technical direction the SEAN is taking, the progress being made, and the specific services the SEAN is relying on WASO to provide. I&M WASO technical staff use the plan to learn who their customers are, and gauge the type and service level of infrastructure the SEAN needs from them. I&M WASO IT planners can learn from the plan specific enhancements needed, including capacity increments, in order to adequately support the SEAN.

1.4 Role of Inventory and Monitoring

“...a primary role of the Inventory and Monitoring (I&M) Program is to collect, organize, and make available natural resource data and to contribute to the Service's

institutional knowledge by facilitating the transformation of data into information through analysis, synthesis, and modeling.”¹

1.5 Mission of the SEAN

The SEAN operates under this mission as stated in its Vital Signs Management Plan (Moynahan et al. 2008):

The SEAN collects, manages, analyzes, and reports long-term ecological data for the primary purpose of informing the network parks of the status, condition, and trends of key natural resources. The SEAN effectively delivers information so that customers – park staff, partners, and the public – can readily locate and use it with minimal reliance on network staff.

Explicit emphasis is placed on delivering useful products to customers. This can be accomplished by leveraging modern technologies, especially the World Wide Web. The SEAN is turning the disadvantages of a late start in data management into the advantage of designing operations around newer mainstream IT capabilities.

1.6 The SEAN Domain

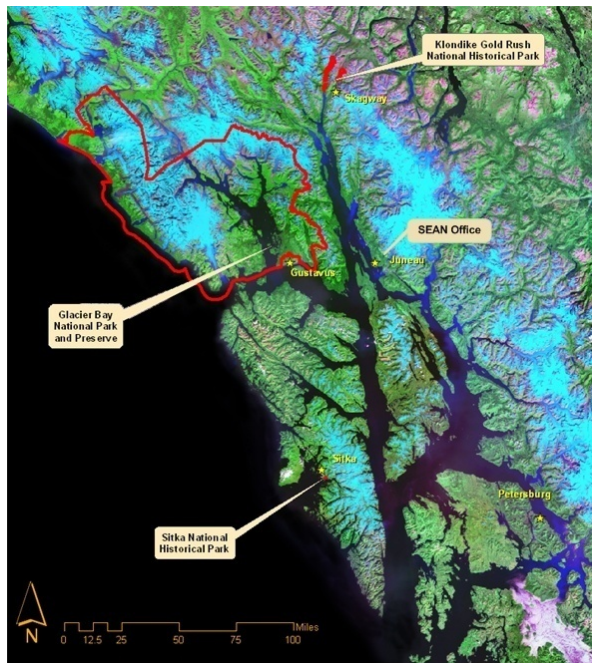


Figure 1.1. Satellite image illustrating SEAN’s geographic domain.

The SEAN services the three national parks of Southeast Alaska. Its office is centrally located in Juneau. No roads connect the three parks with each other or with the SEAN. Access is only by airplane or boat journey. The physical isolation of these units makes computer networking and advanced communications technology particularly valuable. The parks supported, depicted in Figure 1.1, are:

- Glacier Bay National Park and Preserve (GLBA)
- Klondike Gold Rush National Historical Park (KLGO)
- Sitka National Historical Park (SITK)

The SEAN is one of four I&M networks in the Alaska region. Its location in relation to the other networks is shown in Figure 1.2.

¹ National Park Service. 2008. NPS: Nature & Science » Inventory & Monitoring. <http://science.nature.nps.gov/im/index.cfm>.

1.7 The SEAN Customers

A first step in planning data management processes is identifying all the customers that will be served. The customers and their interest in the network are enumerated in Table 1.1. Note that all NPS staff, except for the two SEAN employees, are considered to be customers of the network.

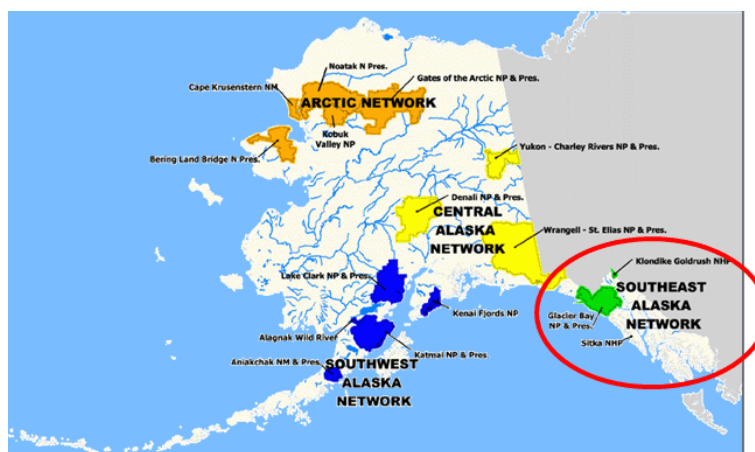


Figure 1.2. Map showing SEAN in relation to other networks in Alaska.

Table 1.1. The SEAN customers and their interests in using the network.

Customers	Use the SEAN Information To
Park managers	- Understand overall park ecology - Make management decisions based on sound science
NPS Resource Management staff	- Ascertain detailed information regarding their park and its ecology - Make sound recommendations to management
Non-NPS researchers and the general public	- Obtain data and methods for their own projects - Benefit from published protocols and other useful documents - Replicate NPS studies - Acquire background information and perspective
Libraries, museums, and archives	- Obtain materials for their customers - Assure longevity of park data, information, and specimens
Partner agencies	Offer to their customers combined data covering a wide range of temporal and spatial strata obtained from a variety of sources
Data acquisition project members	- Determine their exact responsibilities regarding vital signs monitoring projects - Understand deadlines in order to plan details - Follow project progress - Obtain useful related data from other SEAN projects - Get approval of and closure for their work - Have a source to refer colleagues to regarding their studies
External agencies	Support their research by sharing detailed data

Customers	Use the SEAN Information To
I&M program management	• Follow the status of the SEAN development and operations • Understand the strengths and weaknesses of the SEAN implementation • Make informed management decisions regarding the SEAN
Network and infrastructure management	• Know what infrastructure is being used • Plan for network bandwidth needs • Plan for future data storage capacity • Understand the service level required and attained • Prioritize emergency repairs

Figure 1.3 is a data flow diagram (DFD). It depicts the active consumers as well as the providers of the SEAN's information. Its scope is the top level of abstraction. This high level diagram is decomposed into detailed levels in Appendix A.

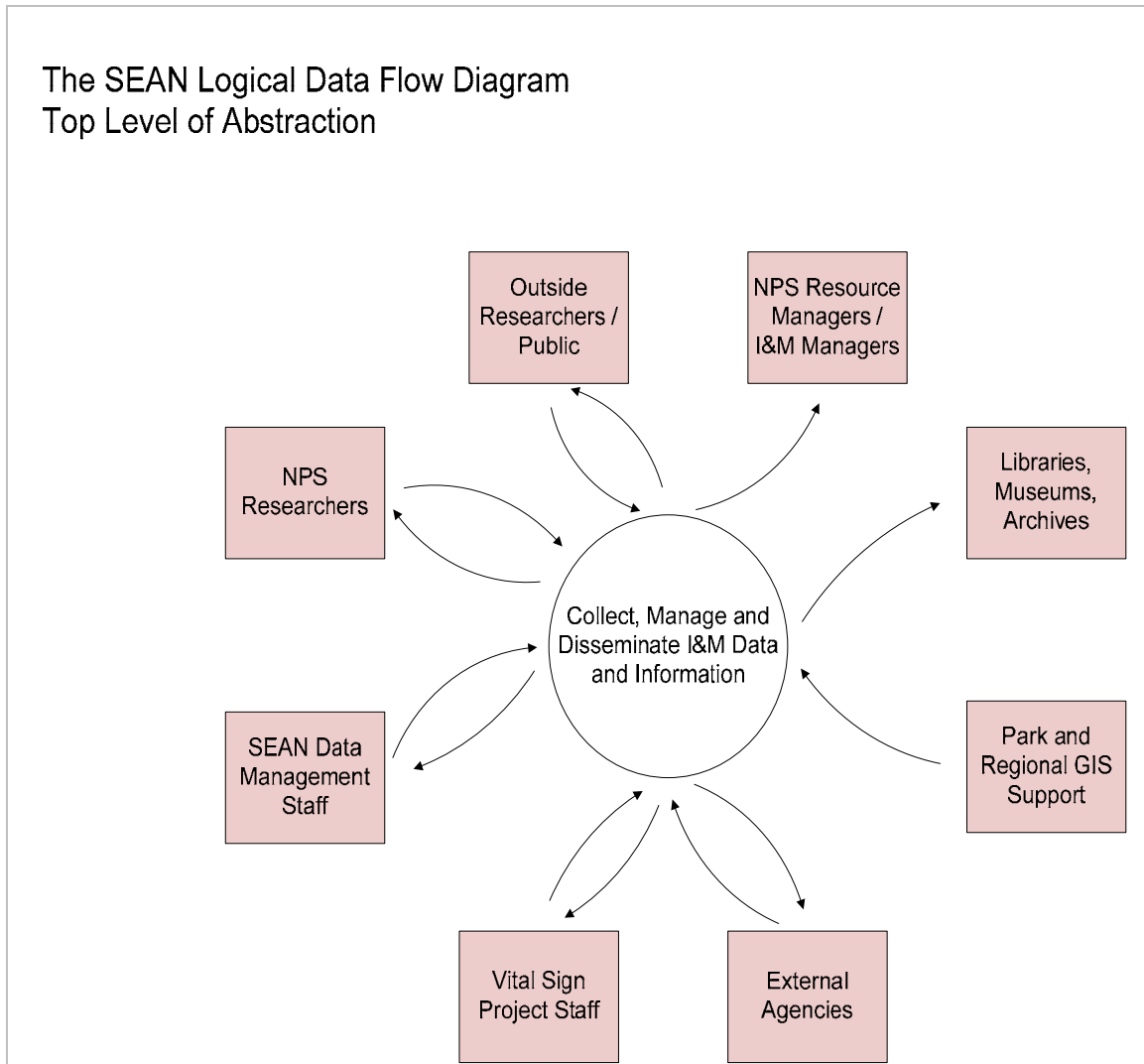


Figure 1.3. Context level data flow diagram. Arrows pointing into a box indicate that the group accepts information from the network. Arrows pointing out of a box indicate the group provides information to the network for subsequent storage and/or actions.

1.8 Data Management Network Design Philosophy

The SEAN is among the last of I&M's networks to be designed. This allows it to be forward-looking for two reasons. First, valuable groundwork and overall structure have already been created by other networks and the parks. Second, useful new technology has entered the mainstream over recent years, which the SEAN is in a position to take full advantage of.

Progressive design concepts were actively pursued in this plan, including:

- Emphasize the effective dissemination of information using web delivery technologies.
- Recognize the very different information management requirements of long-term monitoring projects compared to those that are ad hoc.
- Designate a single “authoritative source” for each of its information products in order to prevent multiple and divergent copies.
- Understand the need to disseminate a variety of information both inside and outside the NPS firewalls.
- Leverage the existing facilities of partners.

The SEAN provides direct support to the three network parks and makes information available to the NPS and the outside world. Its data management design is based on the Core Functional model depicted in Figure 1.4. This figure provides the foundation for the plan; each of the four nodes is referred to as a “Core Function” throughout the rest of this document. The model presented in Figure 1.4 is explicitly customer-centric; it is based on deciding and describing exactly what needs to be delivered to customers and then developing processes to properly provide these products.

The SEAN’s data management is based in these four primary Core Functions: 1) Information Dissemination, 2) Information Repository, 3) Information Certification, and 4) Information Acquisition. Each of the Core Functions must be addressed during the two phases of establishing a data management program that is fully integrated within the vital signs monitoring program: the Development Phase and the Operations Phase. Given the context of our customer-centric philosophy, we have structured the Development Phase to be built in a clockwise order illustrated in Figure 1.4 (i.e., 1-2-3-4). It is essential the SEAN clearly articulates the desired end-products *before* initiating data collection in the field. We recognize that this presentation is likely opposite the direction in which a monitoring program *operates*. Indeed, the Operations Phase (once the program is fully developed) does flow in the chronological and counter-clockwise order of 4-3-2-1 (beginning with data collection and ending in dissemination).

The purpose of our emphasis on dissemination is to underscore the critical importance of specifying desired data and products before constructing observation, certification, and repository processes. A useful metaphor: to achieve a desired objective, it is more effective to pick a destination and then drive to it, rather than to drive around and then see if the place one arrives at is useful. The SEAN DMP addresses both network *development* and network *operations*, which have significantly different characteristics and require differing skills. Please note that the color-coding for these Core Functions in Figure 1.4 is carried throughout other figures in this plan. The figure illustrates the linear operating process by which information is made available:

- Valid information is acquired from collectors.
- Certification is granted to valid information.
- Repositories are populated with certified information.

- Information is disseminated from repositories.
- Customers make use of disseminated information.

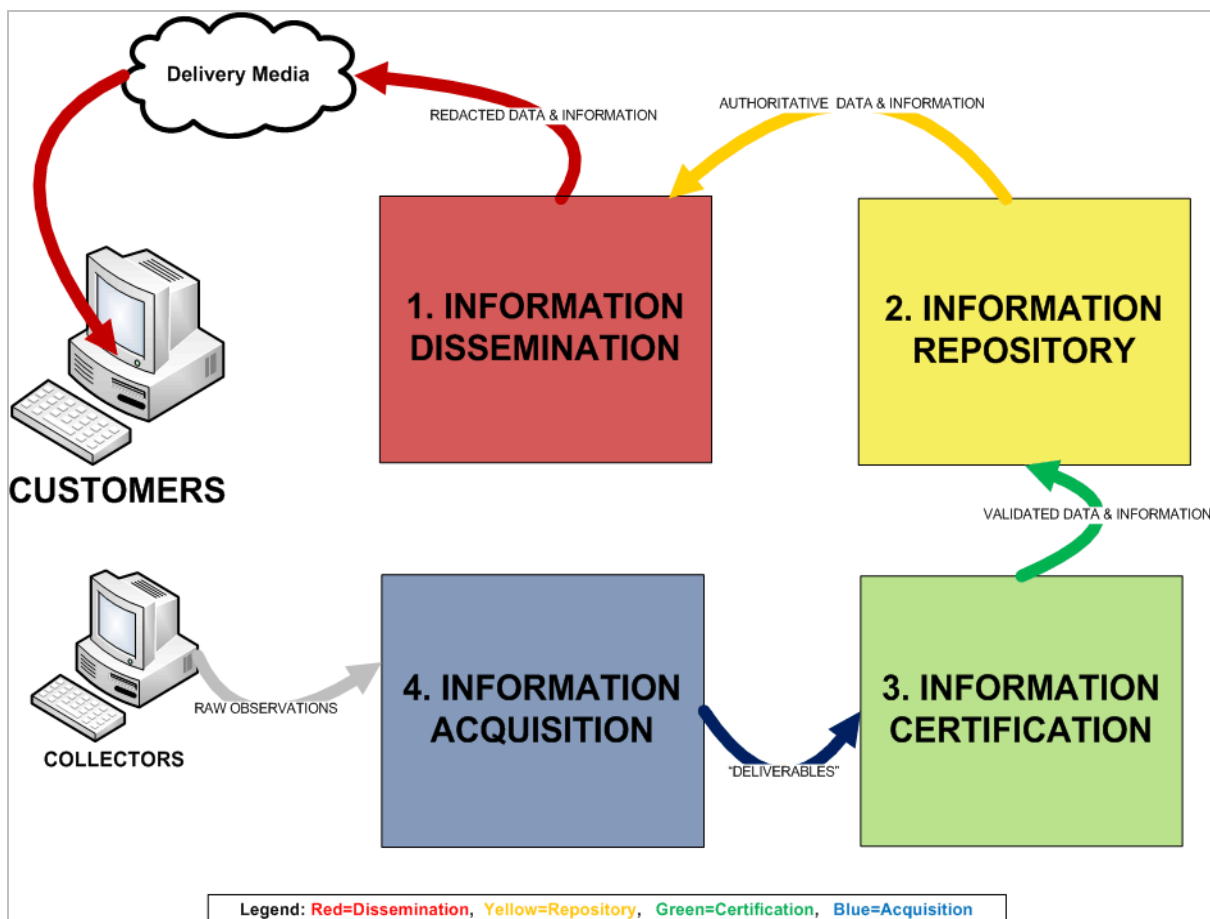


Figure 1.4. The SEAN Core Functional Model overview. Dissemination services (red) provide all data products to all customers using Internet web servers. Repositories (yellow) store the certified inventory and monitoring products called for in the protocols. Certification processes (green) assure repositories have the highest quality data and that sensitive items are restricted to authorized users. Data acquisition processes (blue) led by park staff and cooperators include a wide array of tasks, ranging from designing field forms to producing talks on long-term trends.

Long-term monitoring projects have fundamentally different data management requirements than one-time, stand-alone projects (e.g., research investigations or baseline inventories). The usefulness of time-series (i.e., monitoring) data over time is dependent on their consistency over time. Modification of data methods between collection periods produces discontinuities which result in opportunities for interpretation errors and may constrain retrospective analyses. Therefore, rigorous planning is needed to *a priori* select (for each vital sign) the stable data elements that are useful over the long run, and to establish broader priorities that identify the types of data management services to which

the SEAN will direct its attention. To this end, we identify four priorities for the network in pursuing its mission.

1. First and foremost, the SEAN will establish and execute projects that directly perform long-term monitoring related to the Core Program vital signs described in Chapter 3 of the VSMP. Data management work includes incorporating legacy data for a specific sign where practical.
2. The SEAN will contribute data management support for Vital Signs in the Secondary Program (VSMP Chapter 3).
3. We may assist the SEAN parks with technical data management support for Vital Signs that are not in the Core or Secondary Programs, but are included on the “parent list” in the VSMP, Appendix I.
4. Lowest priority is assigned to one-time measurements that may inform a vital sign monitoring process but that are not vital signs themselves. Development of data management processes for these efforts may cost as much as full operation of some long-term programs. We include this item in our plan because we recognize that a particularly important future data need could be identified that would require considerable effort and expense to obtain.

For vital sign data to be accessible across time and people, it must be readily locatable and accompanied with sufficient documentation (i.e., metadata) to enable accurate interpretation and reliable use by a customer. Data use should not be conditional on collaborating with specific staff of a network or park, because data should be longer lived than any employee’s tenure. Metadata – the specific documentation necessary to make time-series data useful – will vary by vital sign.

The World Wide Web (“web”) offers easy access to information, as well as the ability to clearly link data with its essential documentation. Computer networks at all SEAN parks have adequate bandwidth, latency, and reliability for staff to rely on the web for daily work. Also, customers generally do not have to purchase or install special software products to work with data. Browsers and plug-ins are free and readily available. Technologically, we can now leave behind the “push model”, where key research persons need to provide customers with assistance. The “pull model” lets current and future customers make good use of vital signs data on their own.

The SEAN is committed to making all of its data products and supporting information available through the web. The SEAN internet home page will deliver program information and have a link for each vital sign. Once a vital sign monitoring protocol is designed and implemented, the SEAN Data Manager will create a main page for it. As deliverables and documentation become certified, links on the vital sign main page will provide access to them. The SEAN requires that all project data deliverables assume a form that may be directly disseminated over the web.

NPS emphasizes that data quality is extremely important to maintain long-term integrity. The SEAN will enforce an “authoritative source” concept, whereby one and only one copy of each data deliverable will be stored in one and only one repository. After vital

sign datasets are acquired, they will be validated by the Data Manager. The Project Leader will next certify completeness. The certified data deliverable will then be installed in one and only one designated repository, fully replacing any earlier version of that deliverable. Adherence to this approach is essential because if multiple copies of data exist, those copies will eventually diverge. Otherwise, given the 100-year horizon for I&M vital signs data, the probability of two copies of any dataset remaining identical over time is low. When multiple divergent datasets are known to exist, the accuracy of none can be assured. For this reason, SEAN data management will maintain only one authoritative copy of any product, and only that one copy will be disseminated.

1.9 Plan Revisions

Because this document is based on the National Data Management Plan, any significant update to the national plan should trigger a revision of the SEAN plan. Future revised DMPs shall each be issued as a complete unit, not as piecemeal sections.

Because SEAN data management has not yet been fully implemented, defects in the plan may not yet have become visible. Defects are expected to surface as network development progresses. Also, rapid evolution of technology regularly provides better ways to accomplish tasks. For these reasons, annual plan reviews and updates are recommended. A static plan is likely to become obsolete.

As new processes are honed, their workings should be documented in new and revised standard operating procedures. These may be added to SOP Appendix B by the Data Manager as needed. It is not necessary to formally revise the entire SEAN data management plan when incrementally improving Appendix B.

The version of a network plan document shall be denoted using its specific issue date. Revision numbers will not be assigned.

SOPS Relevant to this Chapter

101 - Review Process for Network Guidance Documents

Chapter 2. Infrastructure and Systems Architecture

For the purposes of this plan “cooperators” are non-park entities that perform only the Acquisition Core Function; “partners” are outside entities that only perform Repository and Dissemination Core Functions.

2.1 The SEAN Overall Architecture

Based on the Core Functional diagram in Figure 1.4, the SEAN developed an architecture to service it (Figure 2.1).

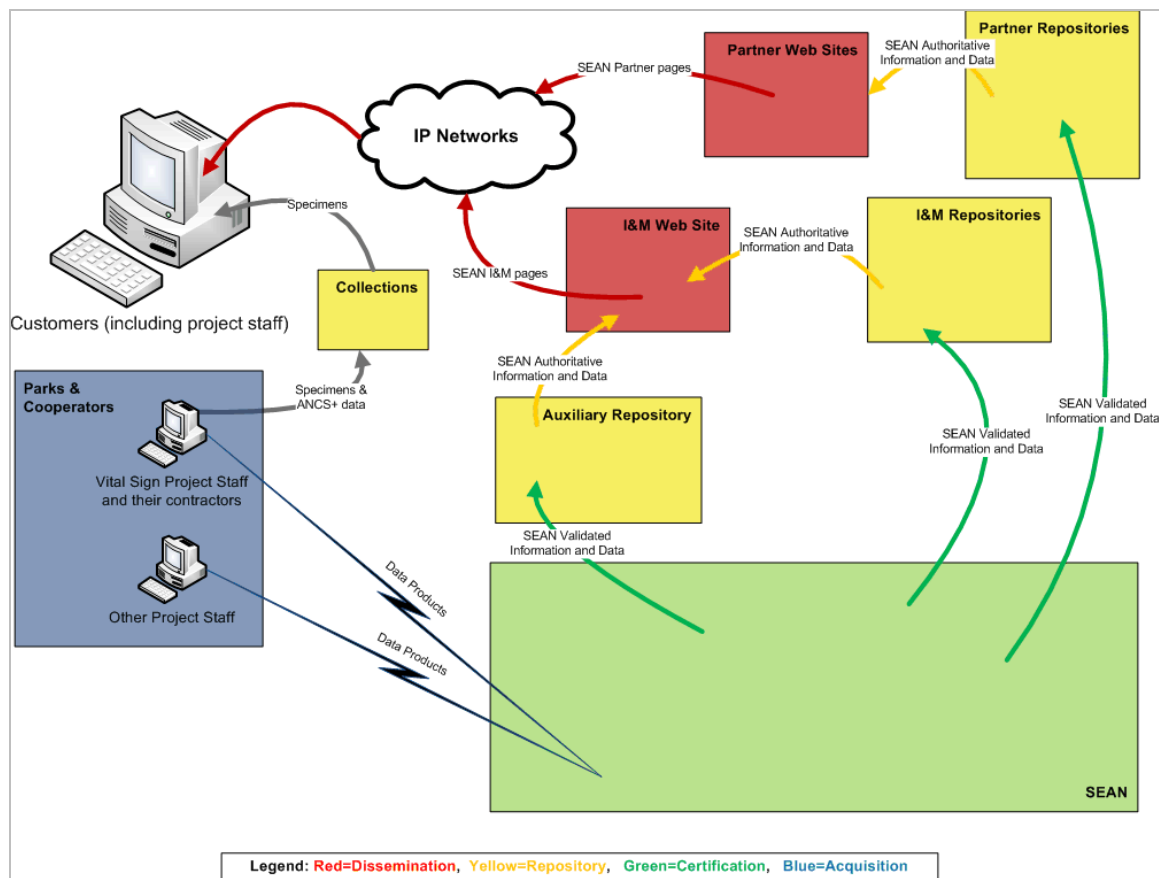


Figure 2.1. The SEAN overall architecture. In operating order, data acquisition (blue) provides deliverables to validation (green). Certified products are installed in the appropriate repository (yellow). Customers find the data using web browsers and servers (red).

Looked at in order of development, the red dissemination services are provided by Internet web servers maintained by WASO and SEAN partners (e.g., Western Regional Climate Center). Yellow repositories are located at WASO I&M, outside agency partners, collection curators, and the SEAN itself. The I&M repositories include the stores for

enterprise applications like NPS Data Store and NPSpecies. Specimen collections in various locations also service the Repository Core Function. The certification processes, in green, rely on IT facilities in the Juneau office. Blue represents the data acquisition processes led by park staff and cooperators using local IT resources. The data Acquisition Core Function creates all the deliverables, including non-tabular items like reports, posters, and media files. Technical details of the associated data flows are diagrammed in Appendix A. Facilities required to implement the service architecture are presented in Table 2.1.

Table 2.1. Facilities required to support the system architecture, by location.

Location	Facilities Used
I&M Fort Collins	• Main repository: database servers that store enterprise application data and possible future SEAN-specific products • Public-facing Internet web servers
SEAN Juneau	• Auxiliary repository: file servers storing data products that don't fit the main repository • File servers holding data deliverable staging areas and folders of working papers • Development web server • Project management tools and documents • Possible future development database server
GLBA, KLON, and SITK	• File servers holding local project work areas, NRDT MS Access project databases, and GIS facilities • Collections that house physical specimens
Cooperators	• File servers holding local project work areas and NRDT MS Access databases
Partners	• Partner repositories: as defined by the partner • Public-facing Internet web servers

2.2 Standard Commercial Applications

Application software standards required by DOI and NPS must be observed. Data products cannot be specified in the protocol as a deliverable if they are based in software that attempts to bypass required standards. The following items are included in the standards:

- Microsoft Office including Microsoft Access
- ESRI geographic information system tools
- Microsoft SQL Server database management system

The SEAN also limits acceptable products in order to keep the number of formats manageable, as shown in Table 5.3. Long-term data cast in certain file formats must be periodically converted to match commercial software version upgrades. Otherwise, the data will eventually become unusable to customers. The high cost of regularly version-converting, auditing, and retesting data becomes unsupportable unless the number of allowable software product formats is limited. The SEAN concentrates on well-supported software having high market share, which usually exhibit the greatest longevity.

2.3 Enterprise Applications

The Service maintains several enterprise-scale applications capable of storing and disseminating information often associated with vital signs. The SEAN can achieve economy and consistency by taking advantage of them to the fullest extent practical. These applications are listed in Table 2.2.

Table 2.2. NPS applications available to store and disseminate vital sign-related information.

Application	Uses
NPS Data Store	• Metadata repository capable of documenting all SEAN datasets, regardless of their ultimate repository • General data repository capable of storing tabular data
NPSpecies, centralized version	• Tabular data detailing biological taxonomy in NPS-managed areas
NatureBib	• Repository for citable documents
NPS Storet	• Agency staging area for preparing data submissions to the national EPA Storet water quality database
ANCS+	• Accessioning database used in curating NPS collections
RPRS	• NPS research permit database • “Investigator Annual Report” repository
NPS Focus	• Repository for photo sets

The national I&M program is currently developing additional information storage and retrieval capability. The project is designated Integration of Resource Management Applications (IRMA). It is based in service-oriented architecture. The processes defined in this DMP are compatible with IRMA. We expect to migrate to IRMA services as they become available.

2.4 IP Network and Web Server Infrastructure

The SEAN office is physically located outside the parks it services. Because roads do not connect the parks with the SEAN, effective network connectivity is essential in order to get work done. The network and park offices all have dedicated circuits with adequate

bandwidth, latency, and reliability. This infrastructure is maintained by IT technicians in the parks, region, and WASO.

The SEAN relies on the Internet's World Wide Web to disseminate information to all customers. The SEAN does not overextend its resources by building a separate but essentially duplicated intranet site. Web sites used by the SEAN are public-facing. This approach not only allows external customers the ability to work with SEAN data products, but it also lets all NPS staff access them, regardless of whether or not they can connect to the internal WAN facilities from their location of the moment.

The SEAN's primary production web site is maintained by I&M in Fort Collins. Dissemination partners maintain their own public facing web servers for distributing the SEAN contributions.

If any sensitive data are identified within dissemination product candidates, the IRMA identity management service will be employed to assure data are made available only to those specifically authorized to see it. For additional security, authentication will be allowed only for browsers inside the NPS internal Active Directory domain. Since partner facilities are not under direct NPS control, sensitive data must never be stored in partner repositories.

Dissemination is enhanced by employing a Google search appliance dedicated to SEAN products, background files, and documentation. The appliance indexes the SEAN's web pages. It also indexes and directly serves out many SEAN Word documents, PDFs, and other files. Firewalls currently block all access to the Google appliance from outside NPS's internal network.

The SEAN employs a basic set of internal equipment and software facilities to support its functionality (Figure 2.2). Additional equipment must be used by parks, partners, and cooperators. The SEAN does not dictate infrastructure details to them. Instead, the SEAN provides specific detailed requirements of the products and services it needs. It is then up to the parks, partners, and cooperators to obtain and manage the technology required to deliver those products and services.

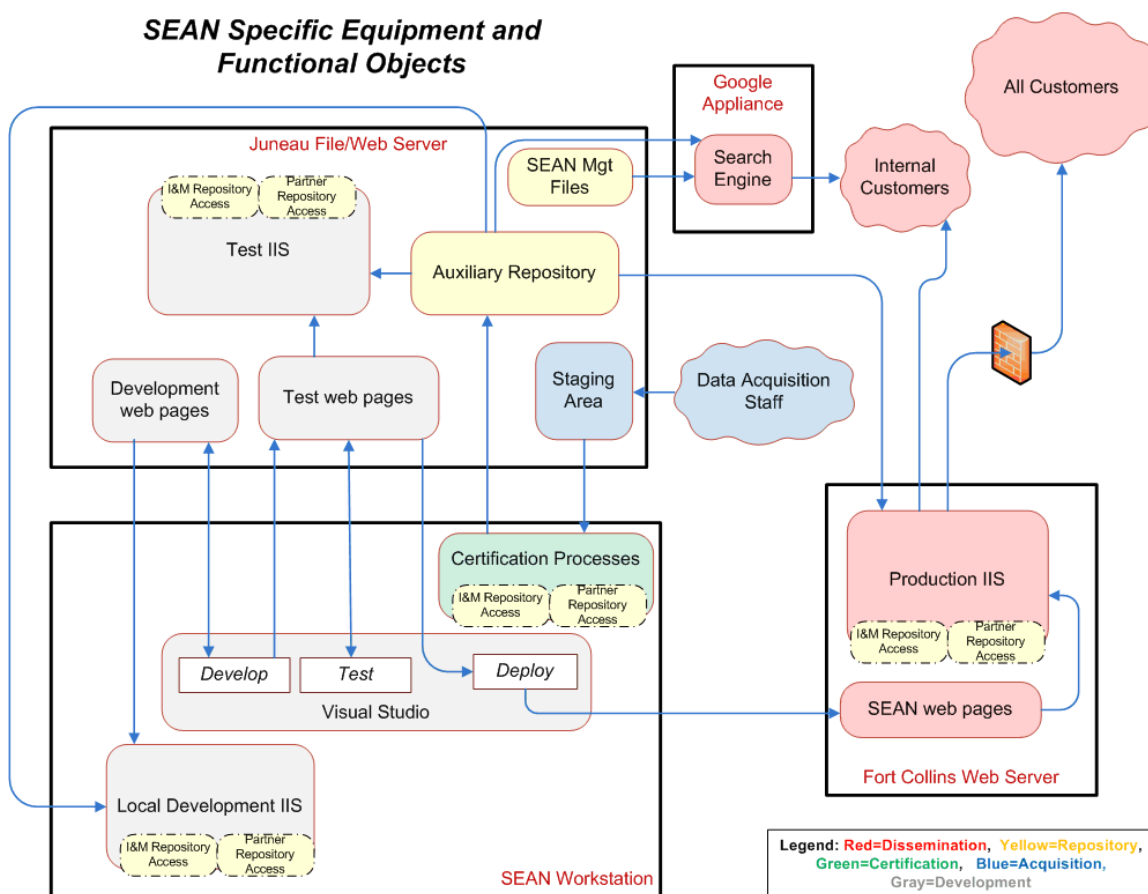


Figure 2.2. Details of the specific equipment and objects the SEAN directly employs to manage its data. Associated functional areas are denoted using the four colors consistently employed throughout the plan. Items in gray, however, have been added to represent development tools. Development tools are essential but are not associated with one single functional area.

While the production web facilities are at I&M, a development environment is also required for building and testing the SEAN's latest web pages. The development web site is maintained on a local SEAN file server in Juneau. Security measures prevent it from being altered except by the SEAN staff.

Data products from cooperators as well as NPS staff are typically submitted to the Data Manager as email attachments. This affords a substantial audit trail of deliverables, and requires no new infrastructure. Some files may exceed Lotus Notes' capabilities. NPS FTP services will be used to transfer items when email cannot.

The Data Manager will use NPS applications over the WAN to install deliverable data products in the I&M repository. Directly accessing files on remote servers through "Windows shares" may also be used under certain circumstances. The Data Manager will supply data to partner repositories using the protocols required by the partner. The Data Manager will install items in the auxiliary repository using Windows shares.

2.5 Database Servers

The I&M repository resides in database servers at Fort Collins. These support the production stores of NPSpecies, NPS Data Store, and other existing applications. Relying on the I&M production database servers offers significant advantages over keeping local data. Among them:

- Fulltime, experienced staff maintain the databases.
- Comprehensive backup, restore, and archiving processes exist.
- The physical proximity to the production web servers allows for very fast connections and improved customer response times.
- Business continuity measures are established.

The SEAN initially uses a local “auxiliary” repository to house many data products. It is likely the SEAN will eventually start installing items directly in production I&M databases. We will then need a local development SQL Server DBMS for testing and staging. It will be deployed on a suitable hardware platform.

2.6 Mid-tier

The SEAN has no direct requirements to maintain its own application servers or middleware at this time. As technology progresses and service-oriented architecture is deployed, a middle tier may eventually become essential.

2.7 File Storage

Data storage is provided by the identified repositories using their own production facilities. One file staging area used for validation exists as directories on the SEAN file server. The auxiliary repository is also stored locally. No data deliverables are ever placed on SEAN workstations, as workstations are not comprehensively backed up. Data acquisition work in progress should be stored, whenever possible, on IT-maintained park file servers. Because field computers are not as robust as park file servers, Project Leaders should back up field data to servers as often as is practical. Handhelds and data loggers, when in the field, should be periodically dumped to flash memory for backup purposes.

2.8 Inter-network Shared Resources

As I&M networks grow, there is much opportunity for inter-network cooperation. This is particularly so in Alaska, whose networks face some similar challenges. All networks would benefit from setting some joint standard operating procedures. In particular, standardization of lookup table objects and contents would increase the ability to share data sets among the networks.

SOPS Relevant to this Chapter

- 201 – Management of Network Internet and Intranet Sites
- 202 – Infrastructure Organization and Maintenance
- 203 – Project Level Directory Structure Recommendation
- 204 – Backup and Restore Routines for I&M Network Data
- 205 – SEAN Internal Directory and File Naming

Chapter 3. Project Management and Data Processing Workflow

This chapter describes the processes used to manage generation of data products by individual vital sign monitoring projects.

3.1 Project Management

The SEAN follows the four-stage project management model illustrated in Figure 3.1. The model comprises two phases: Development and Operations. Development includes the planning/approval stage as well as the design/develop/test stage. The Operations Phase comprises the implement-execute stage and the evaluate-close stage. Because the SEAN has very limited staff resources, the model must be simple and conventional.

The two phases are distinguished by the generally distinct skill sets required to accomplish them. There is skill overlap between planning and designing. There is also overlap between execution and evaluation. But the gulf between design and execution is quite distinct. Ongoing monitoring projects are cyclic and periodically reopen the Development phase. The two phases also occupy different temporal realms, with Development predominating at first but eventually giving way to ongoing Operations.

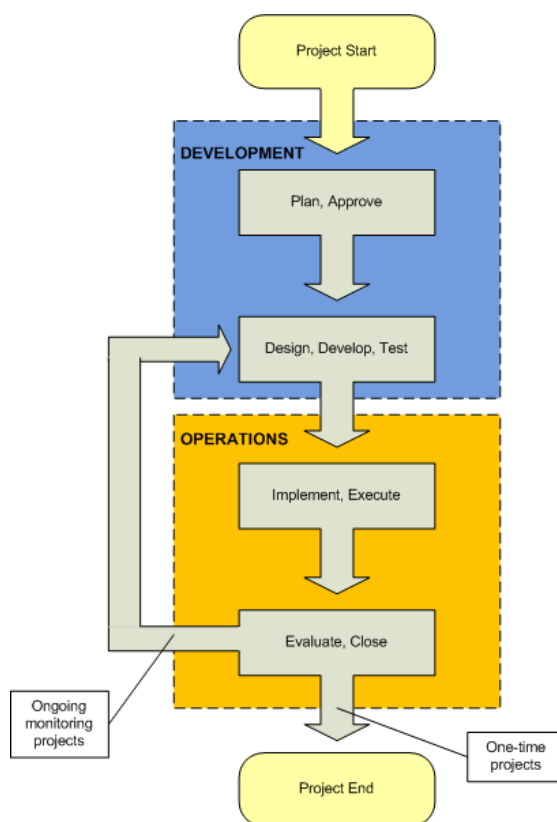


Figure 3.1. The four stages of SEAN project management within the context of two phases: development and operations.

The first stage of the Development Phase is *Planning and Approval*, whereby project goals are framed by articulation of the problem or issue. Once the problem is defined, one can identify measureable objectives that support solutions. These objectives are used to develop a proposal that will meet them. After a proposal is approved and funded, a principle investigator writes a detailed project management plan that identifies deliverables.

The second stage of the Development Phase is *Design, Develop, Test*; an iterative process with multiple tasks. Project management skills are particularly useful here, as very specific products result from this stage. A working protocol should be designed early in development. Chapters 6 and 10 detail the facilities needed, required tasks, and responsible persons encountered in the data side of SEAN project development – once the protocol has been set. It is important that the testing tasks be completed before entering the execution stage, as prescribed in the VSMP. Failure to adequately test proposed approaches typically results in higher costs and lower quality of final work products.

The Operations Phase includes two different stages. The first, *Implement and Execute*, employs the processes delivered in the Development Phase. Components, staff, and tasks commonly required are detailed under “Operations Phase” in Chapters 6 and 10.

Evaluation is the second stage of the Operations Phase, and must follow each execution effort. Responsible persons assure that data collection expectations have been met. Receipt of contracted deliverables must be formally approved and the project tracking log must be closed. For iterative vital sign monitoring procedures, a year-end debriefing identifies improvements to protocols, designs, and policies. These improvements then are addressed by cycling back to the *Design, Develop, Test* stage.

3.2 Data Processing Workflow Concepts

Before detailing the data processing workflow, we define key concepts.

Customers are the users of the network’s data products. Customers include park and non-park users who employ the SEAN products to build other products (e.g., detailed analyses or management recommendations). Customers’ use of the data and related products is what ultimately gives the data collection effort its value.

Project Leaders are those responsible for the Data Acquisition Core Function. Project Leaders may also be their own customers by using their early deliverables as a basis for subsequent products.

Information Dissemination in the SEAN context equates to Internet web delivery. The web supports near-instantaneous distribution of data and information. Web accessibility is inexorably perceived as more important as time progresses.

Repositories are specific facilities that store data and information. Their contents are certified to meet certain standards. Repositories contain final, verified products; the SEAN repositories will never contain informal, draft, or intermediate products.

Repositories have processes in place to guarantee adequate backup, business continuity, and service delivery levels.

An *Authoritative Source* is the one certified copy of data or information stored in a repository and disseminated over the web. It is important to designate a single authoritative source in order to prevent the divergence inherent in maintaining multiple copies of anything on computer facilities. Such divergence may occur without being known, thereby causing different customers to use different data values in their studies. Restoration of a dataset may not be possible once a divergence occurs, because it may not be possible to determine which copy (if any) contains the true project data. The authoritative sources for a particular project may be accessible across several repositories for technical or management reasons. But the same data/information product is never placed in more than one authoritative source, and that authoritative source must reside in one distinct repository.

Information Acquisition is the set of processes required to generate the specific data and information products defined as deliverables in the protocols. Data collection is designed to meet specific pre-defined purposes. It is the responsibility of the Project Leader.

Vital Sign Monitoring Projects measure and report parameters of park vital signs at specified locations and intervals over a span of time. (I&M recommends a planning horizon of 100 years). All data acquisition and certification tasks are done under auspices of a project which is fully defined in a protocol.

Time-series Inconsistency is the result of changing the scope, form, or context of data attributes at particular times. It may prevent data of different periods from being directly comparable. If it goes unnoticed, time-series inconsistency can result in spurious conclusions or may ill-inform resource management decisions.

Partners are defined as non-NPS entities that perform selected Repository and Dissemination Core Functions only.

Cooperators, as distinct from Partners, are non-NPS entities that execute selected roles within the Acquisition Core Function.

3.3 Workflows Differ by Phase

The workflow for the Development Phase centers on designing items to fit the Core Functional model illustrated in Figure 1.4. In designing the data management processes, the first Core Function (Dissemination) is addressed. After the exact products to be disseminated are identified, the Data Manager determines the necessary repository structure and processes (Core Function two). From these structures and processes, the forms of data that will be accepted and maintained in Core Function three (Certification), are ascertained. Once data validation rules are stated, the data collection details may be defined for the Acquisition Core Function four. After all design elements are finalized, then software and hardware are assembled to support the design. The Data Manager then oversees proscribed application building and testing processes.

During the Operations Phase, the Core Functional workflow is followed in the opposite direction. Data are collected, certified, installed in repositories, and, ultimately, disseminated to users. The processes carried out in Operations are those established in the vital sign monitoring protocol and its supporting data management system. They typically have no correspondence with the development processes used to set up the project.

3.4 The SEAN Delivery Workflow

Figure 3.2 details the workflow on the delivery side of the SEAN. (This figure, along with Figure 3.3, depicts workflow from a Development Phase viewpoint, following the functional order 1-2-3-4. If it were considered as a guide to Operations, the workflow would be read from the bottom up.)

Figure 3.2 highlights the Dissemination and Repository Core Functions. Dissemination is performed using Internet web facilities from two sources. The most prominent providers are the NPS's public-facing sites maintained by I&M. They serve customers with deliverables housed in the NPS main and auxiliary repositories. The second type of dissemination providers are the partner web sites. Partners maintain their own repositories, in which the SEAN project deliverables are installed. The repositories are fed from the Certification Core Function. Certification obtains its products from the Acquisition Core Function.

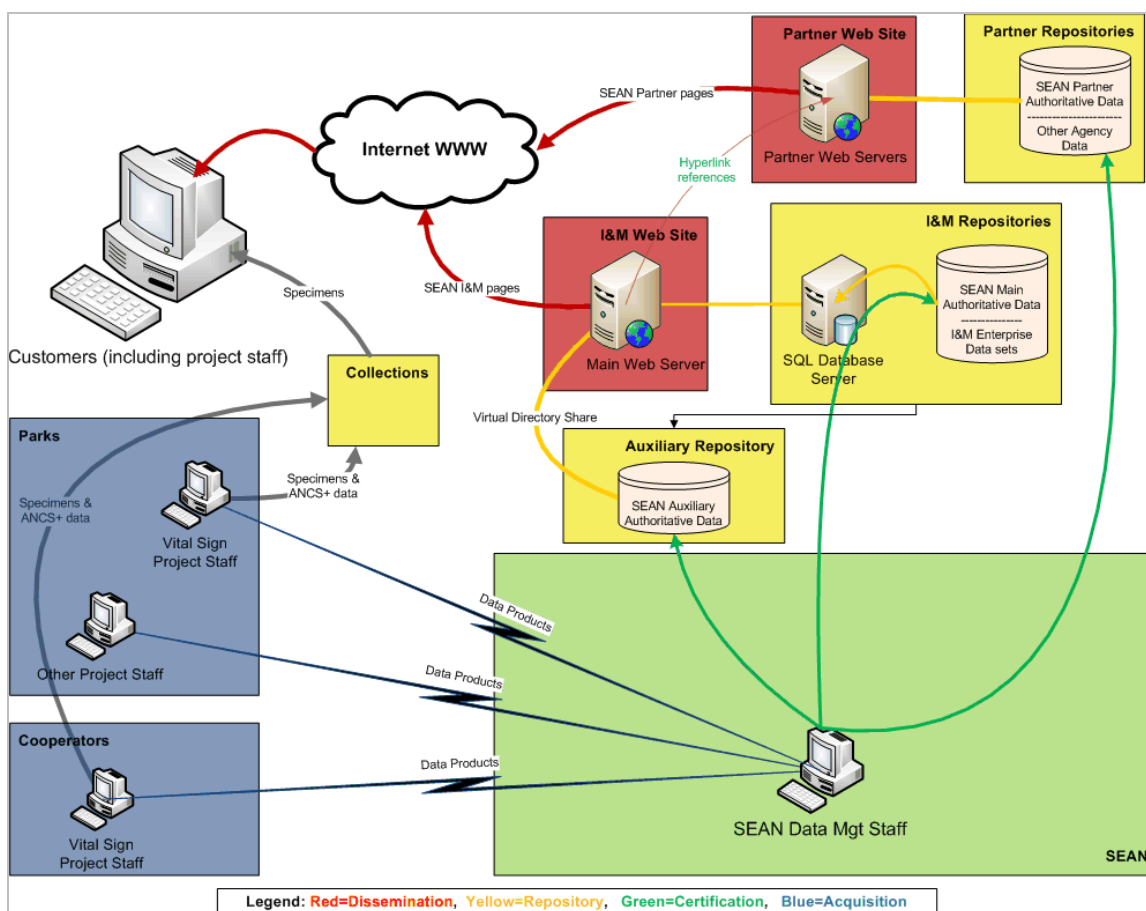


Figure 3.2. The SEAN delivery architecture detail; a partial decomposition of Figure 2.1.

3.5 The SEAN Acquisition Workflow

Figure 3.3 details the acquisition workflow, again in Development Phase order. It reflects the Certification and Acquisition Core Functions, respectively. Candidate data are staged in the Certification Core Function. Only after the products pass rigorous automated validation are they installed in the proper repositories, which were identified in Figure 3.2. Certification is fed with deliverables from the Acquisition Core Function. Acquisition may be performed by park staff and/or cooperators. Project Leaders locally maintain various incomplete and tentative files. The SEAN only accepts and manages deliverable products that are identified in detail in the protocols. The SEAN does not house any other data.

A special direct workflow runs between Project Leaders and Curators of collections; the SEAN does not participate in the management of physical specimen collections.

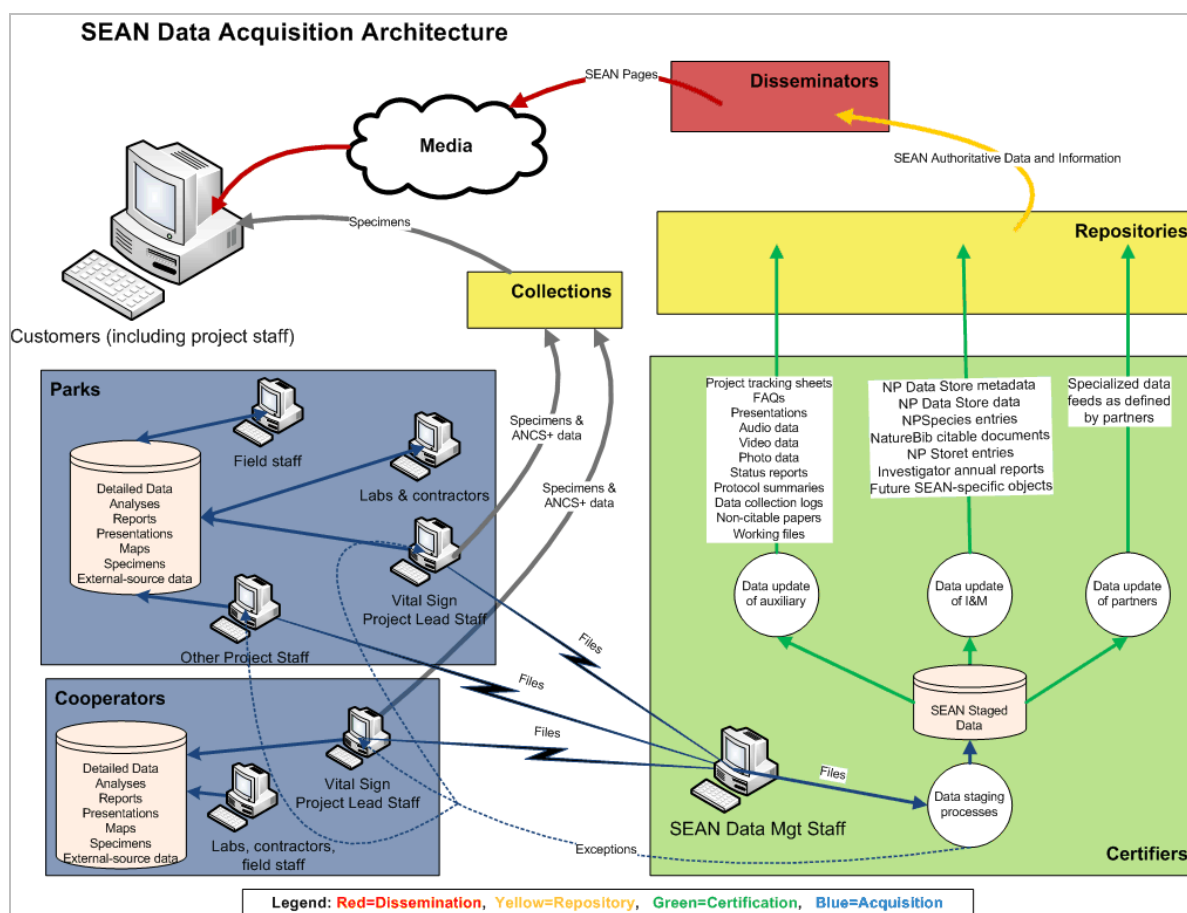


Figure 3.3. The SEAN acquisition architecture detail; a partial decomposition of Figure 2.1.

SOPS Relevant to this Chapter

301 – Checklist for New Data Management Projects

302 – Data Management Considerations in Protocol Development

Chapter 4. Data Management Roles and Responsibilities

4.1 Group Roles

The four Core Functions are serviced by four main provider groups (Figure 4.1). The topmost service provider listed is the group of partners that maintains web dissemination facilities and the repositories from which web sites draw data. The partners independently structure their operations to meet their own objectives. Partners are used only when their objectives and resulting facilities match those useful to NPS projects.

The next group on Figure 4.1 is I&M's national data management staff who operate the public-facing web servers on which the SEAN depends for most of its dissemination. I&M maintains programs on their web servers that serve out information from enterprise applications (e.g., NPS Data Store and NPSpecies). I&M also maintains the repositories for these applications.

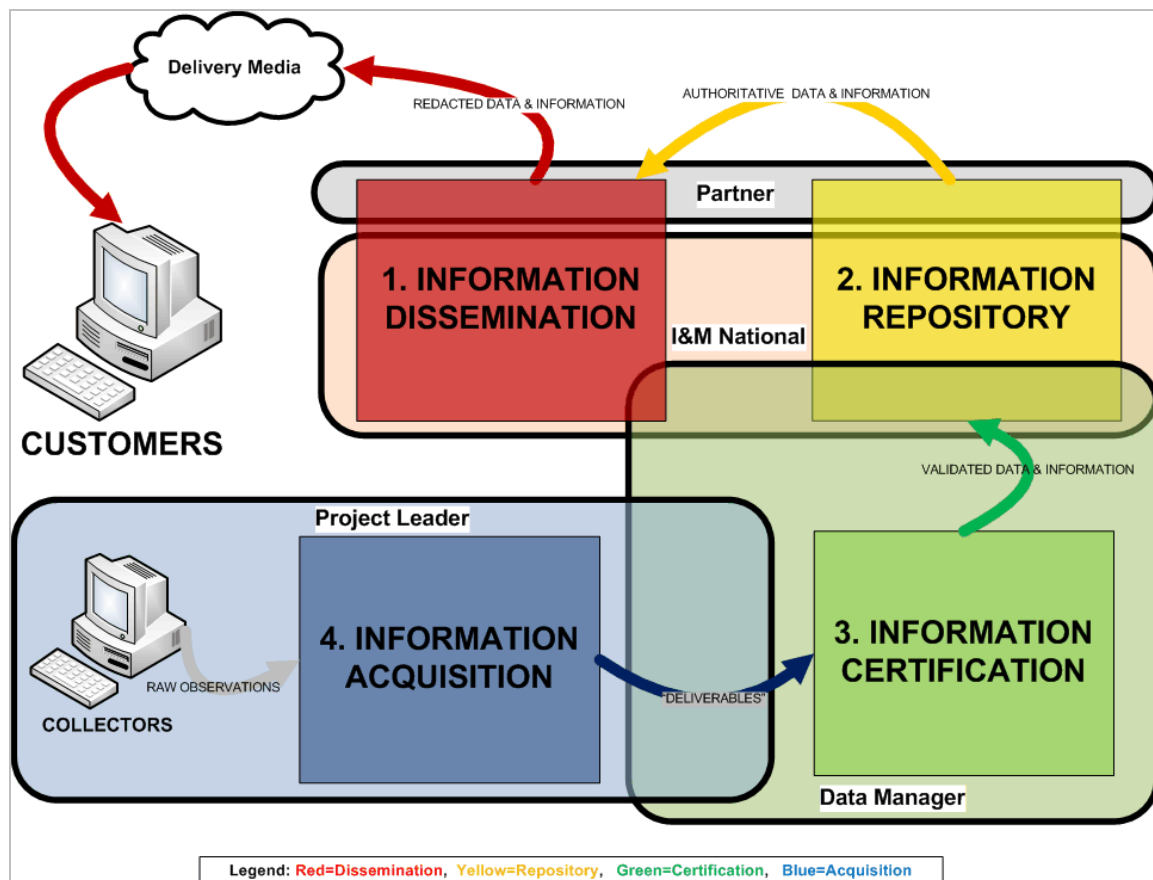


Figure 4.1. The SEAN provider roles are applied to Core Functions. Providers are Partners, I&M WASO, Data Manager, and Project Leaders.

The third group in Figure 4.1 includes the SEAN staff, particularly the Data Manager. The SEAN staff intersects with the Dissemination Core Function because it maintains web pages for the overall SEAN program and for individual vital signs. For two reasons, the SEAN Data Manager also overlaps the Repository Core Function. First, the auxiliary repository of data products is maintained solely by the Data Manager. Second, the Data Manager interacts with the other repositories when managing the installation of deliverables into them. The data staging and validation tasks of the Certification Core Function are completed entirely by the Data Manager.

The final provider group is the Project Leaders, including both park staff and cooperators. They are responsible for creating the deliverables that the SEAN disseminates.

4.2 Development Roles Differ Markedly from Operational Roles

The above mapping of providers to Core Functions (Figure 4.1) is valid for the ongoing operations of the network's data management program. However, a significant development effort must be undertaken to build the systems that the operations staff use (e.g., the SEAN website). Development may be phased such that operations lag only a few steps behind development.

4.3 Specific Staff Roles

While not explicitly defined in the four operating Core Functions, two additional responsibility classes are required to build and run the network: program management and system development. These two critical functions are directed by the Network Coordinator and Data Manager, respectively.

Program management will be a constant and level need. However, development is necessarily front-loaded. Initially, more resources will be required for development, and fewer for operations. Once the initial vital sign programs have been established, development needs will drop substantially. Information management is a joint responsibility among network participants. Specific positions needed to accomplish the Core Functions are shown in Table 4.1.

Table 4.1. Staff requirements by Core Function. Besides the four Core Functions identified in the SEAN's underlying model, management and development functions are also included.

Role	Network Program Management	Information Dissemination	Information Repository	Information Certification	Information Acquisition	System Development
Network Coordinator	■					■
Technical Committee	■					
National I&M Data Manager	■					
Network Data Manager		■	■	■		■
I&M It Staff and Management		■	■			■
Partner Staff		■	■			
Park Superintendent				■		
Project Leader				■	■	■
Park and Regional It Technical Staff					■	■
Curator			■			
Subject Matter Experts				■		■
Park and Regional Gis Specialists					■	■
Subcontractors, labs, etc.					■	■
Park Staff					■	
Cooperators					■	
Web Developers						■
Database Developers						■

4.4 Specific Responsibilities of Participants

Table 4.2 lists the primary responsibilities required of the positions supporting the network. Flexibility and overlap of responsibilities is needed for accomplishing the network's objectives. In order to be successful, some roles need to share tasks that may not be listed as their primary duties. As positions work together, it is not always clear-cut where one person's responsibility ends and another begins. For example, Project Leaders and the Data Manager must coordinate considerably when integrating data management

practices into long-term monitoring protocols. In actual network operations, however, it is desirable to have one specific person bear ultimate responsibility for every task. Otherwise, accountability suffers.

Table 4.2. Task responsibilities for each required staff role.

Role	Representative Data-related Tasks
Network Coordinator	• Provide data management oversight • Prioritize work for the SEAN staff • Allocate funds for development and operations • Manage protocol development • Negotiate partner agreements • Supervise Data Manager • Mediate any disagreements among participants • Evaluate contractor and staff performance
Technical Committee	• Determine scope and priorities of vital sign coverage • Review management plans
National I&M Data Manager	• Recommend enterprise-wide standards • Approve data management plans • Liaise between SEAN and WASO technical staff • Facilitate inter-network communications
Network Data Manager	• Design network information architecture • Compose data management plans • Set technical data requirements in protocols • Model vital signs data • Manage system development • Validate deliverable submissions • Assist Project Leader in resolving data errors • Install final work products in repositories • Maintain the auxiliary repository • Set web content • Tag sensitive data for special security • Track and document deliverable production • Enter selected CSDGM metadata • Audit metadata submissions from others • Audit outside repositories • Monitor dissemination functionality and performance • Coordinate standards with other networks • Implement projects to upgrade repository data to newer versions and formats

Role	Representative Data-related Tasks
Project Leader	• Oversee creation of all deliverables, including non-tabular products • Manage field data collection, entry, editing • Merge into field databases external data from other agencies, laboratories, and contractors • Maintain log of data exceptions and corrective actions • Be sure draft project products are properly backed up • Supervise data entry and verification • Submit deliverables to Data Manager for validation • Correct validation violations • Enter selected CSDGM metadata • Certify valid data products • Prepare specimens for and deliver them to collections • Maintain and archive physical forms and records • Review deliverable schedule progress with Data Manager
Park Superintendent	• Determine which data items are sensitive
I&M IT Staff and Management	• Manage web servers • Provide database servers • Support enterprise applications • Back up and archive repositories
Partner Staff	• Maintain repository servers • Manage web servers • Merge submissions into repositories
Park and Regional IT Technical Staff	• Maintain Internet connectivity • Upgrade software periodically • Repair infrastructure • Provide file server backup and restore
Curator	• Accession specimens
Subject Matter Experts	• Validate specialized data • Contribute specifics to protocols

Role	Representative Data-related Tasks
Park and Regional GIS Specialists	• Advise Project Leader on best ways to use GIS in presenting vital signs • Provide standardized spatial coverages • Develop spatial components and integrate them into deliverable work products • Enter selected CSDGM metadata • Update GIS data products on regional X: drives
Subcontractors, labs, etc.	• Ascertain and report specialized data
Park Staff	• Contribute to producing deliverables as required by Project Leader
Web Developer	• Draft and refine the SEAN home page according to I&M standards • Maintain a web development environment • Post pages to I&M web servers • Develop template for vital signs main pages • Build individual vital signs web pages
Database Developer	• Create data acquisition databases and tools • Maintain a database development environment • Develop programs for validating submissions, building repository feeds, redacting data • Extend repository functionality by developing databases in MS SQL Server environment • Build methods for storing and retrieving binary-large-objects in SQL databases

SOPS Relevant to this Chapter

401 – Operational Roles and Responsibilities

Chapter 5. Databases

5.1 Defining “Database” from a Holistic Perspective

The SEAN defines its database broadly as the comprehensive collection of containers holding its data products, as illustrated in Figure 5.1. The SEAN has only one database, consisting of three repository sets. Each repository houses data products in the form of objects. Objects include photos, reports, sound-file collections, etc. Objects in the SEAN’s database also include tabular data such as MS Access MDB files. MDB files are typically referred to as databases, too. So, the SEAN’s database contains databases, among other things.

Customers access the SEAN database through the SEAN home web page. From there they may retrieve all authoritative SEAN products to work with on their local desktop.

Three repository sets are used by the SEAN database. The first is designated “auxiliary” and is maintained by the SEAN. It holds products that the other repositories are not designed to accept, such as status reports and video files. The second repository is designated “I&M.” This houses data products maintained by enterprise applications such as NPSpecies, NPS Data Store, and NatureBib. The third repository is designated “partners.” These include stores supporting specialized information, for which dissemination is well developed by a third party. NOAA’s Western Regional Climate Center, for example, is a partner repository employed for weather and climate tabular data.

Technically, there is a fourth repository set: “collections” is used for retained curatorial specimens obtained by certain projects. We do not address collections in this section because other service providers – not the SEAN – accept, catalog, and maintain specimen collections.

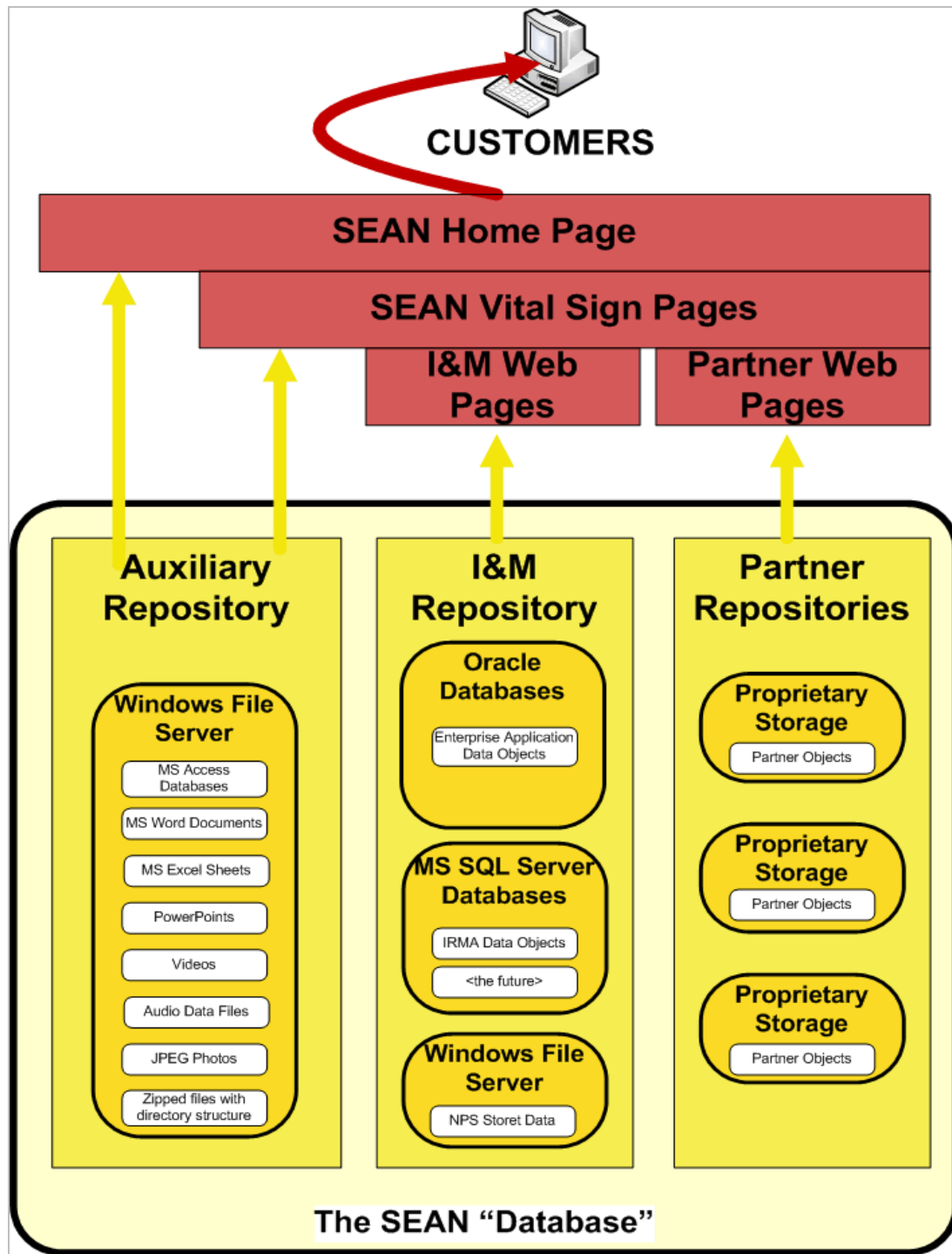


Figure 5.1. Depiction of the SEAN's "database" in the broad sense. Also shown is the relationship of the customers to the database. Red indicates the Dissemination Core Function and yellow corresponds to the Repository Core Function.

5.2 Database Content Is Stored in Specific Repositories

Various object types are used to house particular data products. For example, the deliverable “Protocol Summary” fits in the object type MS Word Document. Word document objects are only stored in the file server used by the Auxiliary Repository. Table 5.1 shows how actual project deliverables are distributed among the repositories.

Table 5.1. Target repositories for specific deliverables.

Auxiliary Repository	I&M Repository	Partner Repository
• Project tracking sheets • Protocol summaries • Project abstracts • Status reports • Data collection logs • FAQs • Presentations • Audio data • Video data • Photo data • Access-type databases • GIS file geodatabases • Non-citable papers • Working files	• NPS Data Store metadata • NPS Data Store data • NPSpecies entries • NatureBib citable documents • NPS Storet entries • Investigator annual reports • NPS Focus photos • Future SEAN-specific objects	• Specialized data feeds as defined by partners

5.3 Implementing the Auxiliary Repository

Two of the data repositories (I&M and Partner) are in operation and are professionally managed without requiring the SEAN staff resources. The remaining one – the auxiliary repository – is built and maintained by the network. Five possible implementation architectures were considered, as shown in Figure 5.2. Architecture selection was not simple because integrating the auxiliary repository contents with the web servers raises substantial organizational and technical issues. The selected architecture houses the auxiliary repository on a Windows file server located in Juneau. This is the simplest alternative and demands fewer up-front development resources than the others. As the network matures, the SEAN expects the auxiliary repository will migrate to a Microsoft SQL Server facility in Fort Collins.

Table 5.2. Possible auxiliary repository architectures considered, in order of desirability.

Possible Auxiliary Repository Architecture	Advantages	Disadvantages
Use MS SQL Server at WASO (with or without mid-tier services) Preferred long-term solution	• IT resources are in place to guarantee ongoing availability, backups, and upgrades • Has very fast connection to NPS disseminating web sites • DBMS facility offers much greater functionality over file-based systems • Database server already in place • May migrate individual items into this gradually and transparently, without impairing existing Dissemination Core Functions	• Development requires long lead time • Storage space requirements for products may be very significant • Development costs are relatively high • Requires maintaining a separate development database server
Use auxiliary repository of directories and files on local SEAN Windows file server Selected initial solution	• Easily controlled by the SEAN staff • File server facility is already in place • Less IT support needed to keep data available and intact, when compared to SQL database alternatives • Overall lower cost than SQL databases	• Has relatively slow connection to NPS disseminating web sites and to NPS users • Requires WASO to be willing to set up a virtual directory on their web server that references a Windows share on a remote server within the AD domain • Requires local staff resources to maintain availability, backups, and upgrades • Marginal robustness for a production facility • Does not have the powerful management and automation capabilities of a SQL database

Possible Auxiliary Repository Architecture	Advantages	Disadvantages
Use auxiliary repository of directories and files on an existing WASO Windows file server	• IT resources are in place to guarantee ongoing availability, backups, and upgrades • Has very fast connection to NPS disseminating web sites • File server facility is already in place • Less IT support needed to keep data available and intact, when compared to SQL database alternative • Lower cost than SQL database	• Requires WASO to be willing to set up a virtual directory on their web server that references a Windows share within their building • Storage space requirements for products may be significant • Unknown robustness for a production facility • Does not have the powerful management and automation capabilities of a SQL database
Use auxiliary repository of directories and files on WASO Web Server	• Does not require virtual directory to another computer • Fastest web site dissemination of content – no network connection required • IT resources are in place to guarantee ongoing availability, backups, and upgrades	• Storage space requirements for products may be very significant for a web server • Does not have the powerful management and automation capabilities of a SQL database • Not a best practice
Use MS SQL Server locally at the SEAN	• DBMS facility offers much greater functionality over file-based systems	• Very limited local IT resources available to maintain availability, backups, tuning, and upgrades of a DBMS • Has relatively slow connection to NPS disseminating web sites • Significant equipment and software costs

5.4 Repository Issues Faced in Each Core Function

5.4.1 Core Function 1: Information Dissemination

Fully functional web dissemination facilities currently exist for the I&M and Partner repositories. These are regularly maintained and enhanced. The auxiliary repository, however, does not yet offer interactive server-based query, extraction, and analysis tools. The primary way products in the auxiliary repository are used is by customers

downloading complete files to their workstations, and manipulating them locally. Downloading is done using web browser / web server facilities only.

5.4.2 Core Function 2: Information Repository

Partners are responsible for managing their own repositories. We provide them with specific submissions, but have no control over actual detailed operations. For this reason, sensitive products may never be given to a partner.

As part of any partner agreement, we will require partners commit to holding the data products we entrust to them using practices that assure the accuracy and long-term preservation of the items. However, because the network is ultimately responsible for the vital signs data, it will permanently maintain safekeeping copies of all submissions made to partners. These copies, however, are never available for dissemination; otherwise the “authoritative source” policy would be violated. They are solely kept for internal business continuity processes, allowing restoration of data to a failed partner repository.

Through the Data Manager, the network will maintain regular business contact with each partner to ensure that the SEAN is aware of any plans partners have for redefining or terminating services, and to ensure that their specifications do not change without the SEAN’s awareness. Such changes in partner processes or relationships could compromise the availability or accuracy of NPS data. Because SEAN will periodically deliver products to partners, contact with partners is somewhat automatic.

The main NPS repository, on the other hand, is professionally managed by I&M. We count on them to properly preserve the SEAN’s submissions. Regular business contact between the SEAN and Fort Collins is assured. So changes in repository management processes will be known well in advance.

It is desirable to reduce the SEAN’s data management effort by deliberately defining final network products to fit existing I&M repository enterprise application databases. For example, formal reports can be specified in a manner that assures they are citable and so may be loaded into NatureBib. Similarly, Microsoft Access database tables used for data acquisition may be laid out so they are assured of fitting within NPS Data Store.

Data products will be housed in NPS Data Store whenever practical. Practicality will be determined on a case-by-case basis, in consultation between the SEAN and the management of NPS Data Store.

The I&M NPS Data Store will be used for all SEAN metadata, regardless of where the authoritative source data are actually housed. Even data sets entrusted to partners, as opposed to being housed in an NPS repository, will have their metadata recorded in NPS Data Store.

Non-tabular data are typically stored as files in the auxiliary repository, within directory structures defined by the SEAN. As the I&M IRMA initiative develops, more types of data products may be accepted by the I&M repository. The SEAN expects that most

auxiliary repository data products may eventually be seamlessly migrated to I&M WASO.

Repository of physical specimens is managed by the park curator or, in certain cases, an outside collection curator. The Project Leader is responsible for delivering specimens to the curator with proper identification and descriptive data in a form acceptable to the curator. In NPS, the standard descriptive data is defined by the ANCS+ record keeping system. The SEAN does not participate in the collector-curator relationship.

5.4.3 Core Function 3: Information Certification

The certification processes validate data and are detailed in Chapter 6. They ensure that all repository data meet the integrity standards set in the protocol and that sensitive items are secured. They ensure that both the Project Leader and the Data Manager agree each deliverable is accurate and ready for dissemination and that there is only one authoritative repository copy of each deliverable.

Files in the auxiliary repository are targeted to be kept available for 100 years. This will require regular ongoing conversion of files to new formats that are usable by current IT tools, as detailed in Chapter 11. Therefore, there is a practical need to restrict the file types that may be specified and submitted to the auxiliary repository. Table 5.3 lists the SEAN's currently accepted file types. The Certification Core Function will not accept file forms outside this scope.

Table 5.3. Specific file types acceptable to the SEAN auxiliary repository.

File Extension	Contents Supported
.DOC/.DOCX	Microsoft Word document
.PDF	Adobe Portable Document Format document
.MDB/.ACCDB	Microsoft Access database
.XLS/.XLSX	Microsoft Excel worksheet
.MP3	Audio
.MP4 ²	Video
.JPG	Image (compressed)
.GIF	Image (simple)
.TIF ³	Image (lossless)
.PPT/.PPTX	Microsoft PowerPoint presentation
.CSV	Comma-separated ASCII text data

² MP4 refers to the MPEG-4/H.264 group of specifications. MP4 content may also be “wrapped” in AVI or WMV containers.

³ TIFF version 6.0 or greater should be used; earlier versions may face software compatibility issues.

File Extension	Contents Supported
.XML	Extensible markup language data
.VSD	Microsoft Visio drawing
.ZIP	LVL-compressed files with directory structure maintained (such as ESRI GDB file geodatabases)

5.4.4 Core Function 4: Information Acquisition

Data acquisition processes store all project data locally in a standard folder structure defined in the protocol. Standard folders greatly facilitate discovery and review by all involved in the project.

Tabular data are collected in local MS Access databases (compliant with the Natural Resources Database Template [NRDT]) that are designed specifically to support the products of the project. The databases and associated tools are always developed and tested in advance of field operations.

The SEAN will not accept tabular vital sign data in the form of spreadsheets, and the protocols will reflect this. A tabular data deliverable might, at times, originally come from a single tab of an Excel spreadsheet. In such cases, the Access database may have a single table which is populated by importing the spreadsheet. Such trivial databases are totally acceptable by the SEAN as a source of raw data. Spreadsheets are not, for two reasons. First, the SEAN does not have the resources to develop expertise in programming complex validation and submission packages for both Access databases and spreadsheets. Second, Access offers some level of quality assurance methods to minimize data corruption. Spreadsheets offer none whatsoever. (Note this does not preclude using spreadsheets to present analyses, as opposed to raw tabular data.)

Depending on proposed analyses, a tabular database may require additional data from outside sources to be merged with project data. It is important this be identified at the protocol development stage so the SEAN has enough information to properly develop Access objects and processes for effectively pulling in the external data. In particular, the keys for joining outside data to the project must be very accurately defined. Performing the merge operation, where actual rows are imported and joined, is the responsibility of the Project Leader.

MS Access databases are not engineered to be distributed databases. The SEAN will provide a single Access database to a Project Leader for capturing the vital sign values. Data entry is to take place on that one database. A Project Leader may, however, try to make multiple copies of the database, distribute them to multiple locations, have multiple entry sessions take place, and then merge the data back into one database. The SEAN does not support and actively discourages this distributed approach because it is a common source of lost and invalid data. If a Project Leader uses such a process, the Project Leader bears full responsibility for attaining data quality. Tabular data products must be submitted to the SEAN for validation as one complete and comprehensive MS

Access file covering the protocol-defined submission unit. Partial submissions are not accepted.

5.5 Database Creation Workflow

Database content creation follows the 1-2-3-4 functional workflow shown in Figure 1.4. Protocol developers define end products. The Data Manager models the data needed to create the end products. Network staff locate the best available dissemination points. The Data Manager locates connected repositories to support the deliverables, and creates any needed containers. He or she outlines methods to be used for validating, securing, and submitting data to repositories. Lastly, the network builds for the Project Manager any necessary acquisition databases and software tools needed for field collection.

During this workflow it is valuable to make overall data designs consistent and comprehensive. There are not enough SEAN staff to build a unique individual data architecture for every project; design elements must be reused whenever practical across the many vital sign projects. Poor execution in laying out the pioneer database structures may compromise subsequent projects. So particular attention must be paid to the first projects.

The basic steps required to build a project's database elements and associated applications are detailed in the development sections of Chapters 6 and 10.

5.6 Inter-Alaska Database Cooperation Offers Benefits

Regional economies in database operations can be attained by pooling the abilities of the four Alaska networks. The most obvious area is in skill sharing, where mutual technical expertise may be provided among network data managers. Another benefit of regional cooperation may be realized by establishing and leveraging standards. If networks agreed on standard attributes to include with, say, freshwater quality measurements, then it would be possible to join datasets into regional reports. If we all used the same coding for tide stage and moon phase and gender and species, direct comparisons could be made on a regional level. On the other hand, if everyone names, stratifies, and aggregates the same data elements differently, then regional analyses will be – and are – burdensome.

SOPS Relevant to this Chapter

- 501 – Managing Spatial Data Products
- 502 – Managing Photographic Data Products

Chapter 6. Data Acquisition, Processing, and Reporting

Data Acquisition refers to the collection and organization of field data, and the creation of deliverable data products from them (Core Function 4 in Figure 1.4). Data processing refers to the processes that formalize, validate, redact and certify the proposed products (Core Function 3). Reporting describes installation of certified products into repositories (Core Function 2). Dissemination (Core Function 1) is detailed in Chapter 10.

6.1 The SEAN's Data Flow

Figure 6.1 further develops the context data flow previously illustrated in Figure 1.3. Additional levels of data flow are provided in Appendix A.

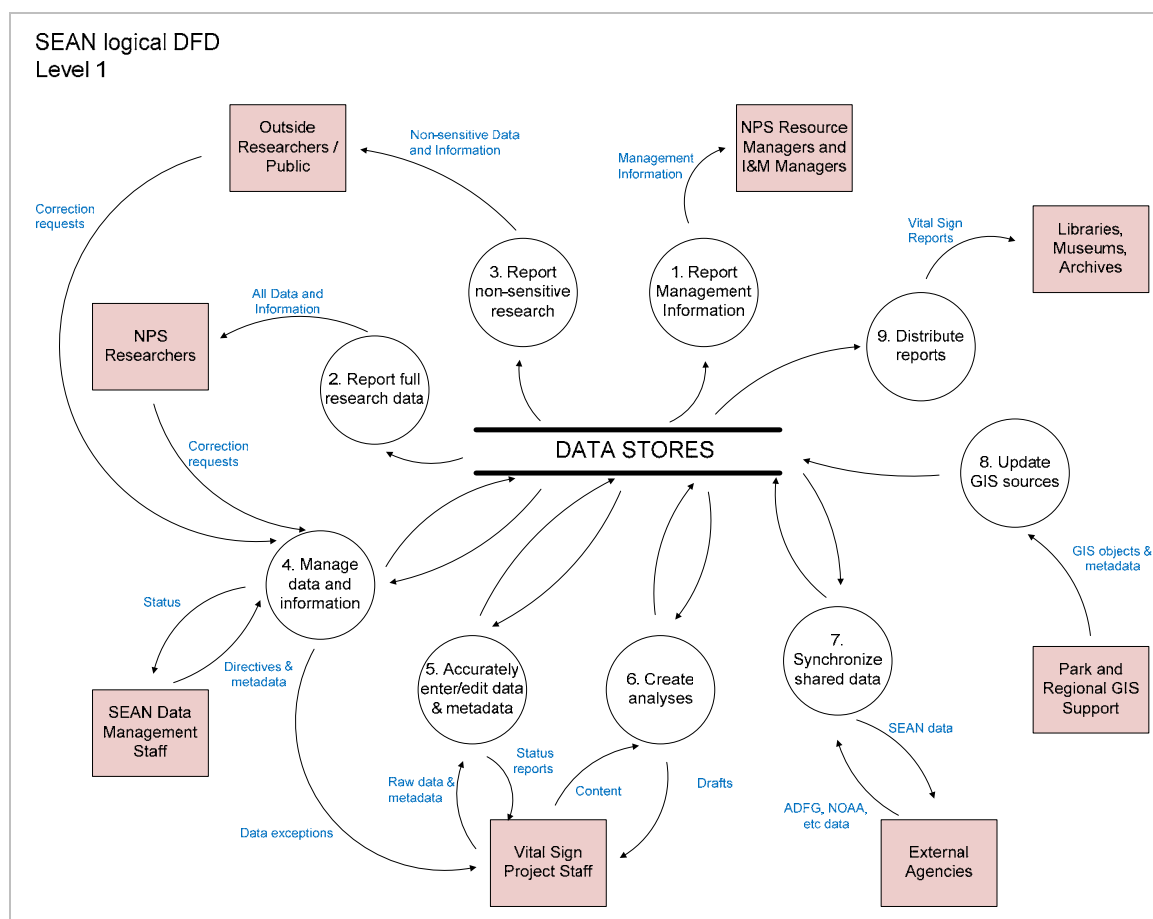


Figure 6.1. Level 1 data flow diagram illustrating participants, processes, and stores. Participating roles are shown in boxes. The specific data they generate or require are identified by labeled directional arrows. The circles denote high level processes that connect the consumer to the information. The “Data Stores” symbol in the center holds the SEAN broadly-defined database of Chapter 5.

6.2 A Distinct Development Phase Precedes the Operations Phase

As the project management section summarized in Chapter 3, the three Core Functions addressed in this section (Repository, Certification, and Acquisition) each have two phases: Development and Operations. It is essential that time-series data be consistently collected and managed over time. Because altering collection rules invites error at a fundamental level, development should be completed, and the data rigorously defined, before any operations begin. As an analogy, consider that the automobile industry finishes building pickup trucks before selling them; they do not need technicians bolting on overlooked parts as the trucks speed down freeways. Likewise with building a comprehensive data management system, if the original development is done well, then operations may continue full speed for many years with only an occasional reopening of development.

When addressing a vital sign for the first time, there may be details that need to be resolved by pilot studies. In that case, the pilot study perpetually resides in the Development Phase. Real operations that generate reliable, consistent time-series only begin after pilots are completed and final protocols are defined. It is desirable to limit use of pilot studies because their data products are sometimes inconsistent with the final data series, and, therefore, may not be useful.

6.3 The Development Phase

Development is executed in the 1-2-3-4 order described in the Functional Model. The tasks involved in development that satisfy Core Functions 2, 3, and 4 are listed in order in Table 6.1 (again, Dissemination, Core Function 1, is addressed in Chapter 10).

Table 6.1. Development tasks necessary to fulfill information repository, certification, and acquisition for each project.

Function	Tasks
Prerequisites	• Specify desired final deliverable data products. • Draft data model that supports creating the deliverables. • Identify any needed NPS application or service interfaces. • Identify any needed partner repository interfaces. • Identify outside raw data sources needed to complete products. • Identify any needed GIS functionality. • Develop detailed entity-relationship diagrams. • Identify business rule constraints. • Define data validation criteria for all data, including for non-tabular objects. • Specify the formats and versions of deliverables. • Draft project deliverable tracking spreadsheets. • Freeze the data products in the protocols. • Submit protocol to I&M clearinghouse with proper versioning notations.

Function	Tasks
Reporting to repositories (Core Function 2)	• Define and execute any partner agreements. • Verify that appropriate backup, restore, security, and accessibility are available for the repositories employed. • Create auxiliary repository containers. • Create and test extract-transform-load processes used to accurately install project data in particular repositories.
Validation/certification processing (Core Function 3)	• Park Superintendents or their designees determine which elements of the proposed data sets are sensitive. • Create project data staging area on the SEAN file server. • Create and test validation processes. • Create and test processes that restrict sensitive data access to NPS staff. • Create final project deliverable tracking spreadsheet. • Clarify progress reporting requirements and schedules for each deliverable between Project Leader and Data Manager.
Data acquisition (Core Function 4)	• Create and test NRDT-compliant database tables and objects for local data collection. • Create and test entry/edit programs using “front-end builder” platform. • Build required GIS components and interfaces. • Build tools required for importing and verifying external data. • Build database reports for predefined analyses. • Perform an end-to-end system test of developed processes. • Design and print any necessary sampling forms. • Obtain any data loggers, tablets, and handhelds. • Create standardized local directory structures on data acquisition computers. • Distribute software to field computers and test. • Review field computer backup procedures. • Initiate a project log for recording collection events and exceptions. • Train data collection staff. • Record initial metadata in NPS Data Store.

6.4 The Operations Phase

Using the two-phase model of project management (Chapter 3), Operations must follow the functional order 4-3-2. Deliverables are collected, certified, and then reported to repositories (Table 6.2).

Table 6.2. Operations tasks needed to fulfill information Acquisition, Certification, and Repository Core Functions.

Function	Tasks
Data acquisition (Core Function 4)	• Project Leader originates certain metadata. • Import any needed data from external groups into database. • Discover, interpret, format any needed legacy data. • Oversee entry/verification of raw field data. • Import any data-logger or PDA captured data into database. • Backup as necessary to prevent catastrophic loss. • Update the project log as significant events occur. • Submit raw data products to validation via email or ftp. • Revise products until validation is passed. • Certify valid products for installation in repositories. • Build any other deliverables based on valid detailed data. • Submit additional deliverables for validation, correction, and certification. • Send specimens to curators. • Archive any original collection forms, tapes, paper photos. • Update deliverable tracking system. • Update metadata.
Validation/certification processing (Core Function 3)	• Data Manager accepts submissions and attempts to validate them. • Report errors and outlier analyses to Project Leader for resolution. • Originate certain metadata. • Accept Project Leader certifications. • Review metadata for completeness. • Update deliverable tracking system.
Reporting to repositories (Core Function 2)	• Apply security tags to sensitive data products. • Once certified, transmit products to repositories of partners, I&M, and the SEAN. • Verify each repository has correctly and completely accepted data. • Update deliverable tracking system.

The subset of above tasks that result in certified products crisscrosses Core Functions 3 and 4, which are managed by different persons. Because this certification process is critical to the integrity of the repositories, its specific sequences are detailed in Figure 6.2.

After the Project Leader submits a candidate deliverable, custom programs analyze all tables in the database to validate them against the mandatory criteria defined in the

protocol. Programs also report outlier analyses and optional criteria exceptions to the Project Leaders. Non-tabular data generally cannot be validated as rigorously. Validation of all data includes determining the file can be correctly opened by current versions of the associated software.

For all file types, validation is withheld until the deliverables successfully match all detailed specifications in the protocols, and until the metadata are complete. Once the data are stamped valid by the Data Manager, the Project Leader certifies that they may be disseminated. Then the deliverables are installed in their appropriate repository by the Data Manager. This is the only route used for populating data repositories. It assures there is only one version of each data product maintained, and it meets the accuracy levels defined in the protocols. This is process by which SEAN realizes the authoritative source concept.

Before certified products are sent to the proper repository, sensitive items are identified and flagged to prevent their unauthorized use. Sensitivity is determined by the park Superintendents or their designees.

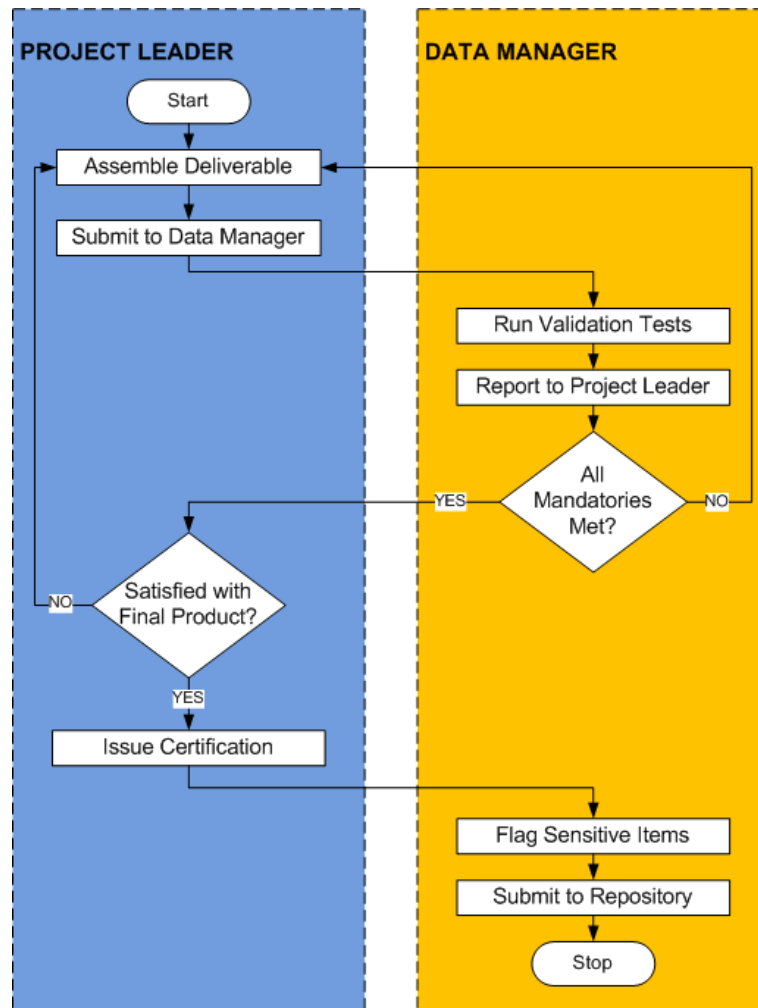


Figure 6.2. Sequence of steps involved in the Certification Core Function.

6.5 Legacy Data Acquisition

Parks contain collections of data that have been historically gathered for various purposes. They are of varying value and quality. The process of locating legacy data to meet current purposes is called data discovery or, sometimes, data mining.

The mission of SEAN is to disseminate meaningful vital signs data and information. To meet this, each vital sign has a detailed protocol that specifies exactly what deliverable products will be produced and which data elements are required to build them. The protocol developer, after diligent research, may decide valuable legacy data are available to base a product in. Procedures to obtain and incorporate these data then get written into the protocol for the vital sign.

The project leader is responsible for executing the protocol, which may include gathering legacy data as well as new field data. Providing products based in legacy data is solely

contained within the data acquisition function. No special handling of legacy data vs. newly collected data is required at the certification, repository, or dissemination level.

6.6 Data Analysis and Reporting

Primary vital signs information products must be detailed up front in the protocol. They should be defined by experts who understand what items are of highest value to the customers. To accurately perform analyses, underlying data needed to produce them should be specified completely. If it is desired to disseminate detailed data, those particularly should be fully defined as deliverables in the protocol. Certification can only be performed on completely specified deliverables.

All information products, including summary tabulations and graphs, need to be generated and delivered to the certification function by the Project Leader. In the event the SEAN has sufficient staff resources, it may be able to detail persons to temporarily work in the data acquisition function in order to assist the Project Leader in basic tabulations and graphing.

SOPS Relevant to this Chapter

- 601 – Procedures for Certifying Project Data
- 602 – Version Control
- 603 – Documenting Product Provenance

Chapter 7. Quality Assurance and Quality Control

7.1 Practical Quality Assurance

Quality Assurance (QA) comprises steps that encourage accurate and complete data acquisition.

7.1.1 *Quality Assurance for Paper-Based Tabular Data Acquisition*

Staff will be thoroughly trained in the proper way to complete collection forms. It is important to stress legible handwriting. To minimize transposition errors during data entry, the screen programmed for entry should match the layout of any paper form as closely as possible. Data entry staff must be thoroughly familiar with edit criteria so they can recognize erroneous entries.

Verification is a process sometimes used to improve the quality of data entry. Most simply, it involves visually checking input against the original document. The easiest verification is having the entry person visually review the screen against the paper form before saving each record. This is simple, but not very effective for two reasons. First, the original entry person who misinterprets a field at entry time is likely to make the same misinterpretation when sight-verifying. Second, entry staff often become mesmerized and simply do not see errors.

Verification results improve when two persons are employed. One reads the form fields aloud and another checks the screen for mismatches. This is staff-intensive and may not be economical. The gold standard of verification is double-blind entry, whereby data are entered separately into two databases by two staff. Programs are executed that indicate which records don't match, and staff take corrective action. This costs more than twice as much as simple entry: entry is done twice; plus custom programs must be built to deal with the comparisons and corrections.

Cost-benefit ratios should be reviewed before committing to expensive forms of verification. The verification cost may exceed the likely benefits realized in data correctness. This is particularly true for double-blind verification, where small record counts often can't justify the large fixed cost of application development.

7.1.2 *Quality Assurance for Computer-based Data Acquisition*

Data that are directly acquired on computer without producing an audit trail of paper forms are susceptible to additional issues. Because no paper records exist, miskeyed data are likely to become permanent errors. Without a document trail, *post facto* corrections are usually impossible. Advantages, however, include some immediate validation at the point of entry and no paper transcription errors. Regardless, direct data acquisition is rapidly becoming the method of choice over paper forms. When the SEAN relies on data from direct acquisition, four steps can be taken to encourage quality.

First, data acquisition applications running on devices must be thoroughly tested. If it is discovered part way through the season that data collection devices are making errors in recording, then all data collected up to that point are suspect. Equipment should be of the

most popular brands and models, because they've had the most field testing. Software must come from reputable developers who do thorough testing in advance of deployment.

Second, in order to catch errors at the point of entry, some degree of collection-time data validation will be employed. However, programmed validation rules cannot be very tight, because overly restrictive rules may cause data to be lost instantaneously. If data cannot be recorded due to a rigid rule, it is rarely possible to have the rule be reprogrammed in the field and allow the data in.

Third, for some vital signs and when possible, the SEAN will use data-loggers with predefined standard solutions built into them. For example, NOAA automated weather stations and industry-leading oceanographic probes do an acceptable job of capturing data accurately and reliably.

Fourth, frequent backups of field computers are essential. Without a paper trail, data accidentally lost are data permanently lost – unless there is a recent backup. Also, field acquisition devices commonly experience battery failure, physical damage, and misplacement. For these reasons, making backups to non-volatile external media in the field is particularly important for these units.

7.1.3 Quality Assurance Considerations for Using Data from Multiple Sources in One Project Database

A vital sign project acquisition tabular database may need to incorporate external data, such as laboratory findings for particular specimens. This is best done by keeping a separate table for each set of external data obtained, as opposed to trying to merge external data into rows of existing observations. For reporting and analyses, the tables may later be logically joined to produce results. This is much simpler to accomplish and less error-prone than fetching existing records, matching them to external values, and then merging the new data columns into the original observation table. Transaction auditing and secure archiving are also preserved by maintaining separate relations.

Joining observations with external data requires well-defined primary and foreign keys. For example, a specimen number may be kept with an internal observation record. A matching specimen number should tag a lab specimen as it goes outside, and the same number should appear on results sent back to the project. The keys should be among the objects specifically identified in the original data model. (In actual practice keys may technically be GUIDs, but the attributes for joining must still be well-defined.)

Regardless of the particular method used to incorporate external data into a project database, the external data's metadata must be reported to NPS Data Store.

7.2 Quality Control

7.2.1 Effective Quality Control is Based in Validation Criteria

Quality control (QC) is most effectively enforced by the data validation processes within the certification Core Function. All validation criteria must be enumerated in advance in

the protocols. The criteria should be derived from the data model's identified business rule constraints. Criteria are of two classes: mandatory and optional.

A data product's mandatory criteria, once codified and defined in the protocols, must be met. Example: all water temperatures, if recorded, are Celsius values ranging from -2 degrees to 35 degrees. Mandatory criteria are also used to restrict values to certain data types, to values in associated lookup tables, to values that are unique, and to whether nullity is allowed. Exceptions to mandatory criteria are not allowed. It is the compliance with mandatory criteria that assures the validity of the repositories. If exceptional data were allowed to enter the repository, the accuracy of all repository data would be called into question.

Optional criteria are used to identify outliers. Violations of optional criteria are reported to Project Leaders for consideration and possible correction. Optional criteria exceptions need not, however, prevent a data product from being incorporated into a repository. An example optional criterion: all water temperatures, if recorded, are Celsius values ranging from 0 degrees to 15 degrees.

Both categories of criteria may also use cross-attribute validation. For example, if the observation time of day is present, then the date cannot be null.

The results of each validation attempt by the Data Manager should immediately be reported to the Project Leader. They should also be recorded in the Submission Log, as defined in SOP 603 – Documenting Product Provenance. Validating staff are never allowed to correct data: they are neither qualified nor responsible for candidate data. The Project Leader accomplishes all corrections.

7.2.2 Quality Control for Tabular Time-Series Data Products

Time-series data sets necessarily are partially updated as new and corrected data become available; most existing data are not disturbed. Only data of a specific scope are updated at each occasion. In order to reliably perform selective updates, each data set needs a "submission unit" defined. That is, the scope of update must be specified as, say, "calendar year" or "calendar year for a specific park." Data attributes used to define the submission unit, of course, must be unambiguous and may never be null.

A global authoritative data set gets updated by deleting all existing data that fits the scope of the submission unit and then inserting the complete validated data for that submission unit provided by the Project Leader. This approach simplifies time-series management. Other techniques require complex methods for identifying exactly which data get deleted, inserted, or changed. Their complexity offers new avenues of error, which is precisely not wanted for something as important as updating the single authoritative master data set for a vital sign.

To improve the audit trail, deleted submission units should be recorded in appropriate audit tables with timestamps. When tabular rows are inserted into the repository, they should include the timestamp, userid, and protocol version number involved – where it is practical.

To improve data completeness, submission deadlines should not be allowed to fall far behind. Over time, observers forget individual details and staff may change positions. If submissions are not timely, data may be lost.

7.2.3 Quality Control for Non-tabular Products

Non-tabular products (e.g., narrative reports or audio collections), are subject to minimal validation. The Data Manager ensures the delivered files are of an allowed type. The basic format is validated by successfully opening the file in the software intended for this purpose. The Data Manager also ensures associated metadata in NPS Data Store are adequate. If any additional validation rules are specified in the protocol, they are also tested.

Chapter 8. Dataset Documentation

8.1 Tabular Data Will Comply with the 8-Level Metadata Standard

Without documentation, data lose their usefulness over time. The metadata reporting standard details are fully explained in Chapter 8 of the national data management plan, which includes the Federal Geographic Data Committee's (FGDC) 7-level Content Standard for Digital Geospatial Metadata (CSDGM) specification. NPS adds one additional level of park-specific documentation. Depending on the nature of particular data, only a subset of the full metadata structure may be appropriate or available.

All SEAN data capable of being documented with the FGDC/NPS metadata standard will have their metadata stored in NPS Data Store.

CSDGM Section 0 (NPS Information) and Section 2 (Data Quality), shall be entered by each Project Leader. CSDGM Sections 3 and 5 regarding spatial data shall be entered by the Park or Regional GIS specialist who consulted in determining how to represent spatial values. The remaining sections are the responsibility of the Data Manager.

8.2 External Data Considerations

External data may be merged into project database tables for a variety of purposes. The corresponding external metadata must also be collected from the originating entity. Network metadata will be generated by merging NPS metadata with outside metadata.

8.3 Data Definitions May Be Proscribed by a Partner

Data housed in a partner repository will be submitted to the partner with metadata in the NPS-required format. It is desired the partner always make metadata available to customers through their web sites. However, a partner may have an established web infrastructure that will not readily accept NPS standard metadata. In that case, a method for providing NPS metadata in the partner's format should be found.

A partner who is ultimately incapable of disseminating adequate metadata when providing SEAN's data to customers faces a serious limitation to their effectiveness. When negotiating a partnership agreement, SEAN should consider inadequate metadata support to be a major deficiency. Unless other very high-value benefits are offered by the partner, pursuing the partnership may not be in NPS's interest.

8.4 Metadata for Non-tabular Products

Some non-tabular data don't have recognized metadata standards. For example, photos may exhibit a wide variety of attributes, but no single definition of what should be recorded as metadata is available from a single authoritative body. Instead, one must chose among IPTC, EXIF, XMP and other competitors.

Some non-tabular data typically have attributes of value beyond the FGDC/NPS scope. For example, Word .DOCX documents are designed to hold keywords, word counts, update dates, and title.

Where protocols define non-tabular deliverables, it is essential the precise metadata content and form be specified. In the absence of recognized standards, the metadata content written into the protocol will be based on guidance from section 8.3 of the national plan.

Where practical, files that support internal attributes will have these recorded in the files by the originators. JPEG photo and MP3 audio are examples of file types that can carry detailed metadata inside them.

The network's Internet web site is intended to afford access to considerable documentation for each vital sign including tracked deliverables, protocols, exception logs, abstracts, etc. One of the Dissemination Core Function's goals is to provide enough documentation alongside the data so it may stand-alone and be self-explanatory. The SEAN vital sign monitoring protocols will rigorously document data in its auxiliary repositories, and this information will be web-accessible.

SOPS Relevant to this Chapter

- 501 – Managing Spatial Data Products
- 502 – Managing Photographic Data Products

Chapter 9. Data Ownership and Sharing

Network data – both certified products as well as all levels of draft – are the property of NPS. Data ownership includes data directly originated by NPS as well as data collected by cooperators. Cooperative agreements must explicitly make this ownership clear. Network data that employ a partner for repository and dissemination services are still property of NPS.

NPS data are in the public domain. Sensitive data, however, are restricted according to law and policy, as outlined in the national data management plan. The park Superintendents' designees and network staff will collaborate on assigning sensitive designations to specific items. Before data are submitted to a repository, the Data Manager will mark sensitive rows or files. Applications that access repositories will never show sensitive items unless proper authorization credentials have been established. Partner repositories, whose access is not totally under control of NPS, shall never have sensitive data submitted to them.

Maintenance of draft work copies of data in various stages of completion are the responsibility of the Project Leader. This includes tabular data as well as reports, analyses, presentations, and other non-tabular data. Once a deliverable is properly validated it becomes the responsibility of the network Data Manager. The Data Manager will distribute products to repositories as soon as they are certified by the Project Leader. However, if a Project Leader fails to certify validated data within the time frame specified in the project protocol, the Data Manager may submit data to repositories unilaterally when it is in the best interest of the Service.

The Project Leader is steward for all data products up to the point they meet all mandatory validation criteria. The network Data Manager is steward once they are validated. Freedom of Information Act (FOIA) requests for validated data will be serviced by the network Data Manager, as directed by management. FOIA requests for draft and incomplete data will be serviced by the Project Leader who is responsible for them.

Data sharing – dissemination – is a key goal of this SEAN DMP and is detailed in Chapter 10.

Chapter 10. Data Dissemination

“Documents prepared without carefully considering how people access or read information inevitably end up as digital landfill: either inaccessible or, if inadequately networked, forgotten.” (Martin. 2002)

10.1 Dissemination Is a High Priority Core Function

Dissemination, being the first Core Function considered in the SEAN’s design, drives all data management tasks in the network. All deliverables are digital and available through Internet web browsing.

The main entry point to the SEAN is the network’s home page. A conceptual mock-up illustrates the home page’s purpose in Figure 10.1. Each vital sign has a link on the home page that connects to the vital sign’s (i.e., project’s) own main page. Figure 10.2 illustrates a mock-up of a hypothetical project main page (these figures are for illustration only; functioning web pages will be reviewed for conformance to NPS standards).

Each vital sign or project has its own main page, offering documentation that enables customers to use the data and information. Such documentation may include (among other things) a project abstract, formal protocol documents, deliverable tracking sheets, permit details, and contact information. Alongside the documentation references are links to the actual deliverable products. These may include Access databases, raw ASCII data, GIS file sets, and others. Links to I&M connect customers to web-enabled enterprise applications like NPS Data Store, as shown in Figure 10.3. Partner data source links connect customers to third-party web sites that specialize in housing and analyzing certain types of data. Figure 10.4 illustrates how this might work for the climate data stewarded by Western Region Climate Center, a group now undergoing partnership consideration.

Project main pages guide customers to the most useful products. The SEAN has also installed a network search appliance (Google Mini®) that indexes and delivers selected background documents, correspondence, and ancillary files. Customers may use the search appliance to browse the program’s full offerings. Figure 10.5 illustrates typical information that may be obtained from a local Google search.

The Dissemination Core Function provides the single authoritative copy of data products from designated repositories; leaving no question as to which of several competing sources may be accurate. Repositories are never loaded with project data until all candidate items in a discrete submission pass all validations; so data meet the measureable standards defined in the protocol. Repositories get populated in blocks of defined submission units; so partial update problems do not arise. Sensitive data are protected through formal user authentication.

Because dissemination is essentially a communication process, it should be bidirectional. The SEAN invites customer dialog. Each of its web pages includes a link for contacting appropriate scientific and support staff.

In order to provide dissemination services at the level listed above, some technical criteria must be met. First, the site has to be public facing. That is, it is an Internet rather than an intranet application. The SEAN does not have enough resources to maintain two entire sites. The Internet design is capable of servicing the widest audience both inside and outside NPS, in the office and on the road. The only technical issue is establishing an authorization process for allowing access to sensitive data, should any be collected. There are a number of standard methods available to accomplish this, such as the IRMA identity management service.

Another technical item is that all our non-sensitive repository data must be delivered through the public World Wide Web over the Internet. This will be a requirement when considering partners. I&M repository data also have to be accessible through the Internet. It is expected the IRMA project will further Internet accessibility of I&M data.

A third technical issue is that I&M's public web servers have to be able to deliver content from the auxiliary repository located in Juneau. A typical practice for accomplishing this is for technicians to establish what's called, in IIS parlance, a virtual directory on a remote server.

Finally, to use the SEAN's facilities, one has to find them. Advertising where and how to find network products may help visibility, at least within the immediate work groups. But the key to success in discovery is being indexed and highly ranked by the leading search engines. The ability to be indexed requires that web pages be "crawlable". The network must not block public search engines from the web server. The high ranking requires careful web page construction, and perhaps the enlistment of other web sites to link to the SEAN. Unless the SEAN's data products hit the search engines, they will remain obscure to the broad scientific community.



Figure 10.1. The SEAN home page content mock-up, illustrating the planned scope.

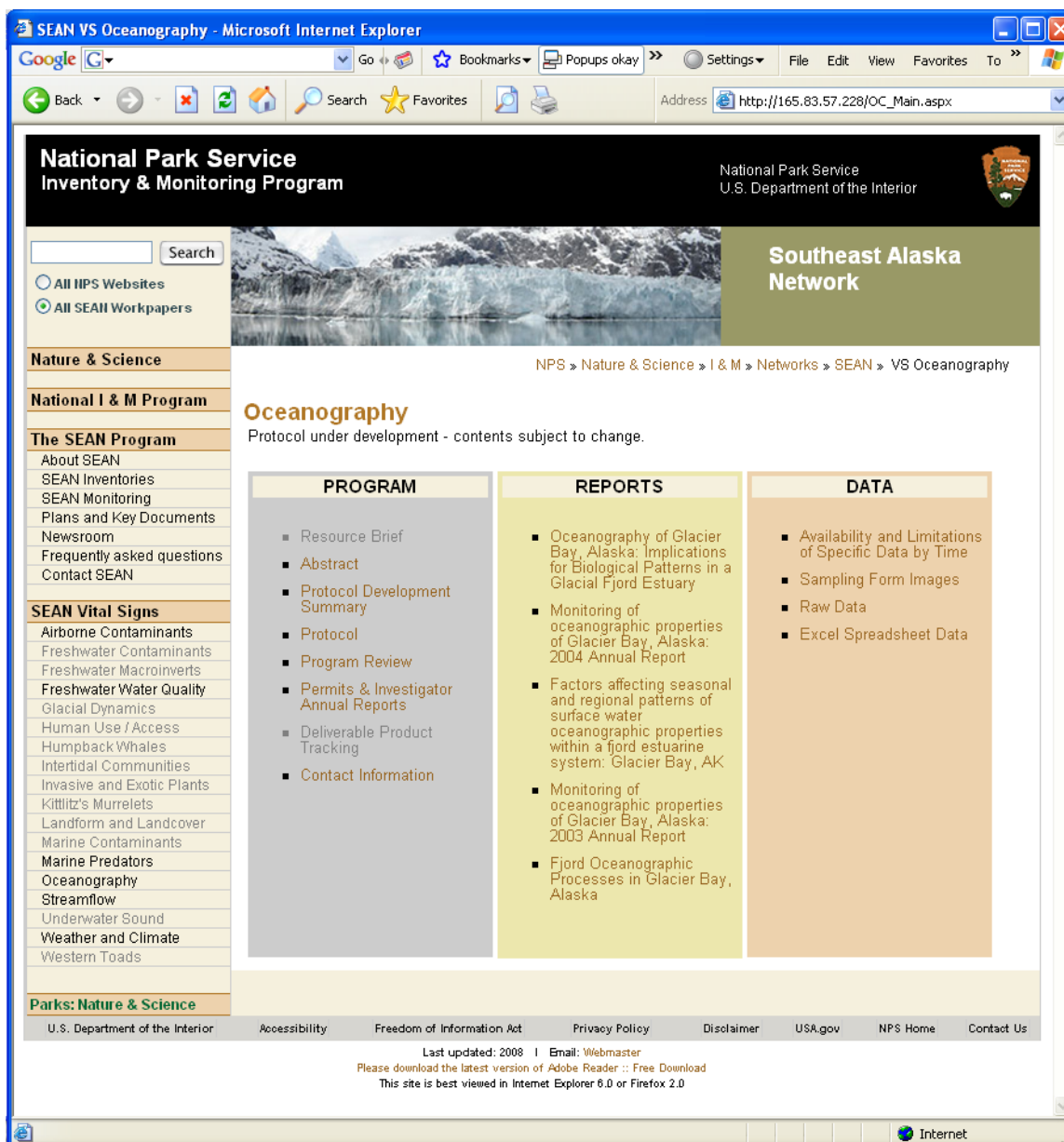


Figure 10.2. Vital sign main page concept example.

NPS Data Store
National Park Service
U.S. Department of the Interior

Application Home »
Login

Search
Select Search Type:
I&M Monitoring
and NPS Unit:
Glacier Bay National Park
Category:
Boundary
Word Search:
Keyword ☐
Title ☐
Abstract ☐
Search
Advanced Search

Help
Instructions
About
News
Data Server
More Info
Data Store FAQs
Data Standards FAQs
Metadata Standards
NPS Metadata Profile

Links
Metadata Tools & Editor
GIS/Metadata Tools
NPS GIS Clearinghouse
(search by map)
Legacy NR FTP Search
NPSFocus
NPS I&M Program
NPS Research Permits
NPS Park Planning

SEARCH CRITERIA: 135 RECORDS RETURNED
NPS UNIT: GLACIER BAY NATIONAL PARK AND PRESERVE (GLBA)
Click on column headings to sort records.
Natural Resource Inventory & Monitoring Program: Official Inventory(I) or Monitoring(M) Records

DATASET CITATION TITLE	TIME PERIOD	SUBJECT/CATEGORY	METADATA PURPOSE
▶ 1994 Elections Districts - Final Reapportionment Plan	//1994	OTHER	CSDGM, NPS, ONLINEDATA
▶ 2005 Alaska Fire Management Options	05/01/2005	FIRE	CSDGM, NPS, ONLINEDATA
▶ 60 Meter Hillshade of the NED Clipped to Alaska	20020615	HILLSHADE	CSDGM, NPS, ONLINEDATA
▶ 60 Meter Hillshade of the NED for Glacier Bay National Park and Preserve	20020615	HILLSHADE	CSDGM, NPS, ONLINEDATA
▶ 60 Meter Hillshade of the NED for Southeast Alaska	20020615	HILLSHADE	CSDGM, NPS, ONLINEDATA
▶ 60 Meter Hillshade of the NED for Southeast Alaska	20020615	HILLSHADE	CSDGM, NPS, ONLINEDATA
▶ ACPM Coastal Zone Management Boundaries	03/01/1994 07/31/2001	BOUNDARY, GENERAL MANAGEMENT	CSDGM, NPS, ONLINEDATA
▶ ADEC's Contingency Planning Regions	//1996	BOUNDARY, RESOURCE MANAGEMENT	CSDGM, NPS, ONLINEDATA
▶ ANCSA Corporation Boundaries	//1990 //1995	BOUNDARY, LAND STATUS / OWNERSHIP	CSDGM, NPS, ONLINEDATA
▶ Administrative Large Parcel Boundaries	12/19/2001	BOUNDARY	CSDGM, NPS, ONLINEDATA
▶ Airports and Runways (FAA and DNR)	12/04/1994	TRANSPORTATION	CSDGM, NPS, ONLINEDATA
▶ Alaska 90 Meter Hillshade	1950 1959	HILLSHADE	CSDGM, NPS, ONLINEDATA
▶ Alaska Coast, 1 to 1,000,000	1991	PHYSICAL	CSDGM, NPS, ONLINEDATA
▶ Alaska Coastline 1 to 63,360	//1997	HYDROGRAPHY, WATER	CSDGM, NPS, ONLINEDATA
▶ Alaska Coastline 1:250,000	//1950 //1980	HYDROGRAPHY,	CSDGM,

Figure 10.3. Example of an I&M enterprise application that may be used to provide web access to SEAN information and data: NPS Data Store.

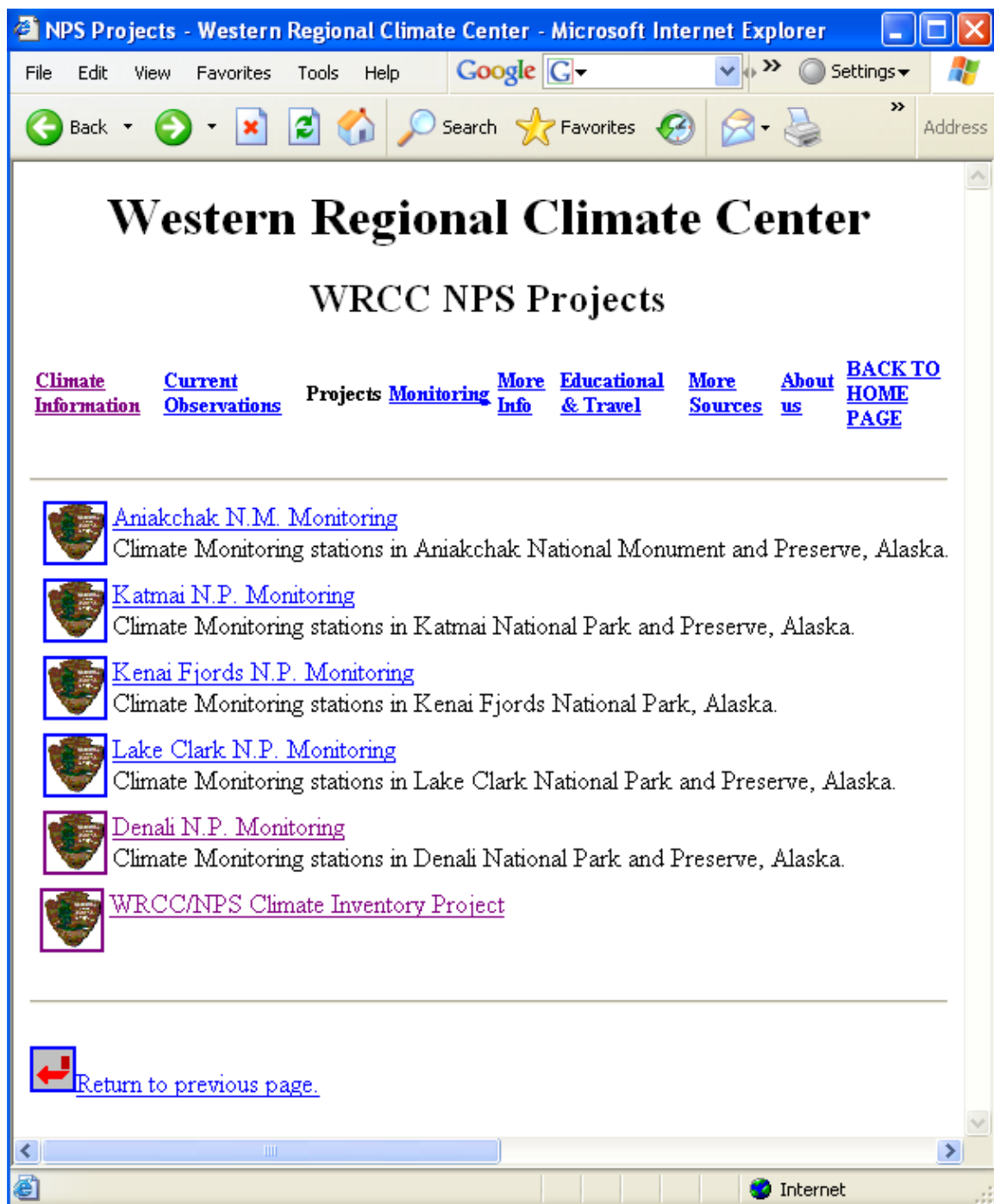


Figure 10.4. Example of a possible partner's repository for the SEAN information and data. Shown is a typical web dissemination page maintained by Western Region Climate Center.

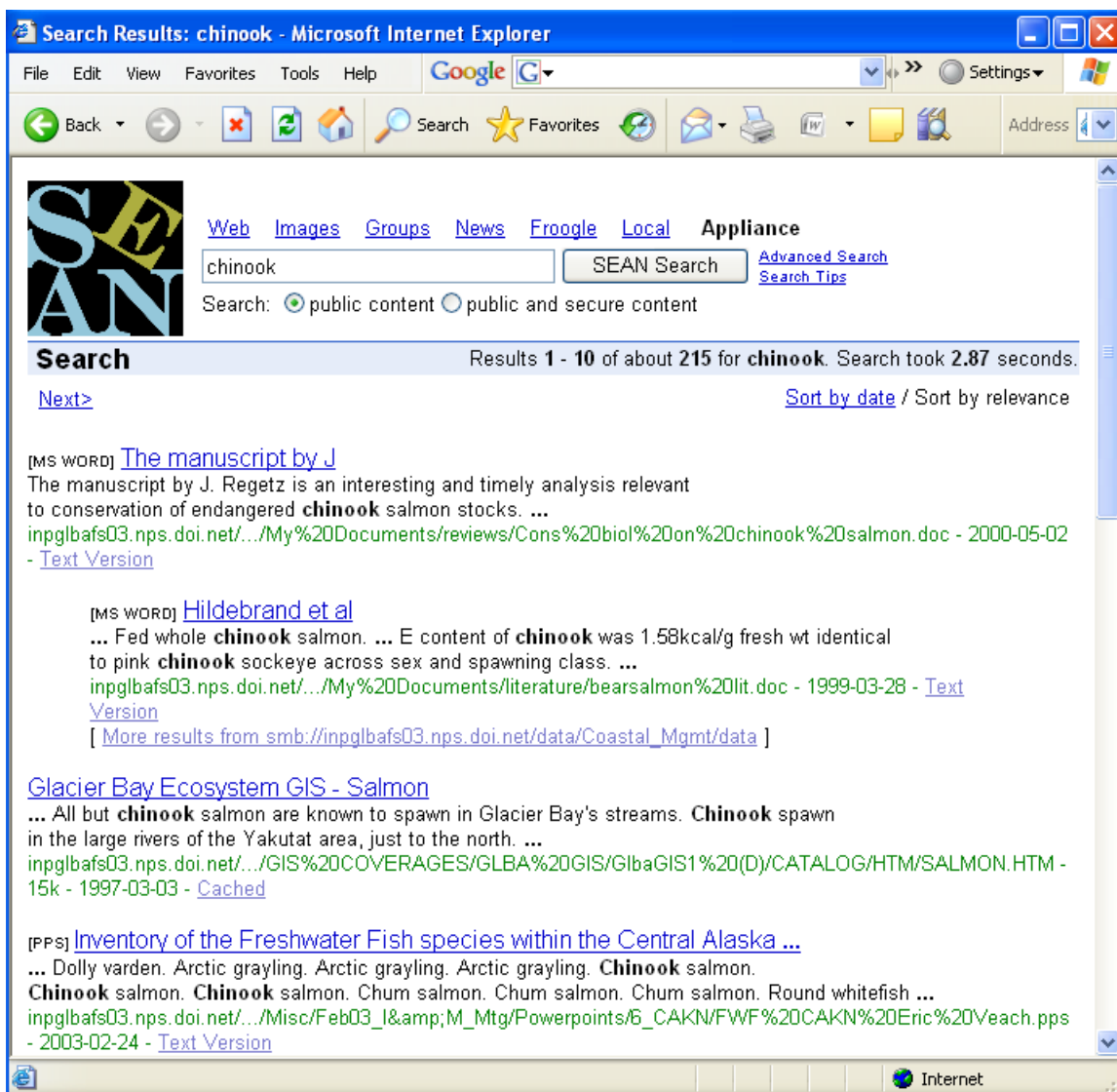


Figure 10.5. Results of a localized Google search for SEAN background documents.

10.2 The Development Phase

10.2.1 Create Web Site Objects for Basic Information Delivery

Dissemination processes are known to be developed in advance of the other Core Functions (which are addressed in Chapter 6).

The SEAN's web site will be hosted by I&M's server in Fort Collins. A virtual directory needs to be established to a share on the auxiliary repository host (currently physically located in Alaska). The capacity and speed of the web server and network will be evaluated. They should be augmented if necessary to provide acceptable performance. In conjunction with security managers, an acceptable form of authorization will be determined for distinguishing NPS and specific park users from public users.

Web development is done on the SEAN's local server in Juneau. Development is never done directly on the production I&M server. Code will be managed by a source code control system compatible with Microsoft Visual Studio.

The SEAN home page will be constructed according to agency design standards. The preferred environment is ASPX under Microsoft IIS programmed in C#. The home page will initially link to frequently-asked-questions, contact information, the three park home pages, and similar basic documentation. A link to the SEAN's dedicated search engine will be established, having security as described in section 2.4. As additional products are developed, links will be added as the products are delivered.

The search engine configuration should be reviewed to ensure it is providing the best service. Establishing search appliance "collections" may be useful for this. The working-paper store should be evaluated for clear organization and value of content.

A general project main page template will be developed, on which each vital sign's main page will be based. Basic project documentation content needs to be created, linked to and stored, typically in the auxiliary repository. These include such items as abstracts, protocol documents, resource briefs, contact information, deliverable tracking.

Using the template, main pages of each of the six initial vital signs defined in the Vital Signs Management Plan will be created. Partner links will be added to the project main page, where and when appropriate.

10.2.2 Deploy Objects to Production Web Site

New web site components must be fully tested in the development environment before being deployed. Deployment is done using the standard Visual Studio tools for copying items onto the production web site. The production site must be re-tested to ensure all links operate correctly with the new host using automatic link-checking software. This should be done from outside the firewalls.

Usability should be reviewed after each major redeployment. Of particular interest are response times faced by customers. The page content should also be examined to ensure items are self-explanatory and adequate documentation is present for customers to successfully use the site without NPS staff assistance. If any project work products have been tagged as sensitive, it must be demonstrated that they may only be accessed after supplying adequate credentials to the system.

Each new deployment must be recorded in a site management log. After updates have been posted for at least one week, search engine presence will be evaluated. If the new pages are not found at all, then the technical infrastructure will be reviewed for access problems. If new pages are indexed, but are ranked very poorly, the Data Manager will consider special efforts to raise their visibility.

10.3 The Operations Phase

The Dissemination Core Function is actually operated by the customer.

To support this, the SEAN's data products will be accessible from all modern workstations. A customer's web browser will be able to successfully render the SEAN's basic web page HTML files, though some variations in look may exist among browser brands and versions. The customer should also be able to work with the SEAN's limited set of work product file types, as identified in Table 5.3. All browsers inherently handle JPG and GIF images. They inherently interpret web objects delivered by I&M applications like NPS Data Store. Customers may need to obtain commonly-available "plug-ins" to render PDF documents and multimedia files of type MP3 and MP4. The Microsoft Office based files are best worked with by having Office installed on the customer's workstation. Customers without Office may still view the content by downloading readily-available free "viewer" software. Collections of proprietary ESRI GIS files cannot be viewed directly in browsers. The customer will have to download them through the browser to the workstation in order to work with GIS products like ArcMap, which must be installed on the local machine.

Chapter 11. Records Management, Data Maintenance, and Archiving

The SEAN will permanently retain the contents of its repositories, as well as the programming code developed to manage the network. Due to rapid developments to the information technology environment, this presents both challenges and opportunities.

11.1 Deliverable Data Products Must Be in Digital Form

All the SEAN deliverables must be provided to the Data Manager as computer files in a protocol-specified digital format. This requirement will not be burdensome for new material; most deliverables are now “born digital.”

11.2 Directives Designate Some Project Records as Mission-Critical

Mission-critical records need to be retained permanently in archives, as detailed in section 11.2 of the national data management plan. The SEAN’s repositories are capable of archiving digital products only. The SEAN building does not have facilities capable of properly supporting physical collections or archives. Any records that are to be retained by SEAN must be designated as formal deliverables in the protocol, so that they may be subsumed into repositories for permanent retention.

The SEAN will assist Project Leaders in finding park or other agency archives capable of cataloging and storing any physical items requiring permanent archiving. Physical collections are essential for items like biological or mineral specimens. Their usefulness is realized through physical examination.

The SEAN will also assist Project Leaders in converting paper field forms into PDF files and in converting paper photographs into JPG files so they may be archived digitally. All items capable of being stored digitally should be, because it makes them much more available. Five years from now an external researcher will be able to readily download a PowerPoint presentation that explains data acquisition details used by a particular SEAN project. Obtaining physical photos, slides, or data sheets to accomplish the same thing could be problematic.

11.3 Preserving the SEAN Repositories

All work products reside in either the auxiliary, I&M, or partners’ repositories. I&M enterprise application repositories are professionally and reliably archived by central NPS staff. The SEAN does not perform any extraordinary measures to duplicate I&M’s regular work.

Partner repository archiving, however, is less certain. While we entrust partners with our data, we have no control over their backup and archiving performance. Therefore, a copy of each partner submission will be kept at the SEAN for NPS archiving, to prevent catastrophic loss due to partner negligence. These copies will only be used to restore lost repositories and will never be directly accessed by customers.

SEAN-held products in the auxiliary repository will be backed up using best practices. Products are never stored on individual workstations. They are always stored on a file

server with redundant disks that is backed up at least once per day. The backup history's granularity covers multiple generations. Copies are kept on-site and off-site.

11.4 Local Physical Archiving

All local data, including web pages, staged validations and certified submission copies, are stored alongside the auxiliary repository. The auxiliary repository and local data are competently backed up. But archiving needs extend beyond that level of coverage. Periodic “permanent retention” copies will be made of the entire repository. The media will be set to read-only. It is noted, digital media do have finite physical and logical life spans. Processes will be implemented that periodically renew the “permanent” media by making new physical copies.

11.5 “Version Obsolescence” Must be Continually Guarded Against

As opposed to the physical life of data, maintaining their logical life is more difficult. For example, the physical medium may be available, but a ten-year old file may not be accessible because its file format is no longer recognized. No one can easily deal with a Quattro-Pro spreadsheet or WordPerfect 2.0 report now; in twenty years, current versions of Excel or Word are unlikely to be useable.

Resources must be periodically expended to bulk upgrade repository files to recent formats. As long as a software product is commercially popular, there will always be a data upgrade path for a period of time – the length of which cannot be predetermined. The SEAN will logically refresh its files periodically so that their format is never more than one version earlier than the currently released product format, as detailed in SOP 1101. Refreshing can be done laboriously by opening files in the latest software and saving them as new files in the current format. When this becomes burdensome, the SEAN may invest in automated methods for accomplishing this. In order to make version upgrades manageable, the SEAN restricts the number of file formats it accepts to the fourteen listed in Table 5.3.

11.6 Archiving Technologies Are Undergoing Rapid Changes

The field of information technology is expanding rapidly. Likewise, the field of archiving technology struggles to maintain pace with that expansion. The “gold standard” archive medium of paper is showing weaknesses. Regardless of paper's physical stability, it is not practical to store an MP3 audio file of bird songs on paper. And, while it could be possible to dump the contents of a large Access database to a printer, it would not be practical to actually use the resulting reams of paper as the basis for data analysis.

Beyond the obvious issue of how to permanently save data is the problem of how customers can find and use it. Web technology offers great facilities for locating data. It allows instantaneous retrieval of information, which is now expected by customers. But the web only works with digital data, not paper catalogs and warehouses of boxes. Until methods are developed to give digital information the same longevity as paper, archiving will be problematic. The state of archive technology should be periodically reviewed with a mind toward finding better archive processes for the network in the future.

11.7 The Value of Monolithic Databases for Housing Repositories

While not currently implemented, the plan looks forward to eventually taking advantage of SQL database technology as a comprehensive platform for repositories. While that technology obviously supports tabular data, current database software also supports storing as database objects all the fourteen binary file types used by the SEAN. A file may be stored as a field in a record that can be retrieved with a query. This capability is rapidly developing and improving.

There are many benefits to such an approach.

- Database management systems (DBMS) that offer binary object storage also provide enterprise-level reliability and functionality.
- The ability to store all network data in a single location using a single access method considerably simplifies data updating, auditing, disseminating, and status monitoring.
- Binary objects may be stored in records along with category attributes – permitting retrieval of primary data as well as “affiliated products” tagged with the same values.
- There is a single platform to use, which vastly simplifies processes for backup, restore, and archiving.
- By using a monolithic database, the periodic DBMS version upgrade process is simplified and is automatically applied to all projects in the database at once.
- When multiple entities share a database platform, they may define and benefit from authoritative standard lookup tables at the park, network, region, and national levels.
- By using shared DBMS technology, it is possible to directly produce composite reports using information from multiple databases of multiple agencies in multiple locations.
- DBMSs support an array of sophisticated management tools to define, monitor, update, and version protect deliverables.

A DBMS solution may be realized either through applications that interact directly with the database tier, or through applications that consume web services fronting the database.

For these reasons, SEAN will regularly assess opportunities for migrating auxiliary data storage and management from the system proposed in this plan for network start-up to a comprehensive DBMS, as posited in Table 5.2.

SOPS Relevant to this Chapter

1101 – Network Archiving Procedures

Chapter 12. Project Tracking and Documentation

Project tracking consists of documenting the progress of delivering the specific products called for in the protocols, and making this information widely available.

12.1 Scope of Project Tracking Responsibility

SEAN tracks its progress in delivering certified work products to customers. Tracking is initiated beginning the moment a project is approved for development. Various basic items, such as completion of a data model and freezing of the protocols, are tracked before any operations begin. Because vital signs projects are typically long-term, tracking tables are created for years and are permanently retained. Progress is recorded exclusively by the data manager and is made widely available over the web.

Each specific operational task must be clearly defined at the outset as the management responsibility of either the project leader or the data manager. The SEAN data management staff will only manage tasks for which it has authority and direct responsibility.

12.2 Tracking is by Spreadsheet

SEAN project operations progress uses a standard spreadsheet form to track progress toward meeting objectives and producing deliverables. While the same form is used for each vital sign, the exact deliverables and milestones vary.

Each tracking spreadsheet is stored in the auxiliary repository. The sheet is globally accessible from the vital sign's web page, alongside all the completed deliverables. Updates to the tracking record are solely the responsibility of the SEAN staff.

While it is desirable to comprehensively track and report on projects using a sophisticated project management system and database, the network cannot currently justify the overhead these require. The project tracking model should be reviewed at the start of each calendar year to determine what improvements are worth pursuing, given available staff and funding.

12.3 Current Tracking Sheet Format

SOP 1201 shows an example tracking sheet, and details its contents and how it is to be maintained.

SOPS Relevant to this Chapter

1201 – Project Tracking

Chapter 13. Implementation

Building the data management operation defined in this plan should be done in four steps that build on one another.

13.1 Implementation Requires Multiple Passes to Complete

The SEAN design specified in this plan is a comprehensive solution to the network's problem of getting accurate data to customers so they may put it to good use, now and in the future. Building the data management facility is too complex to complete in a single, quick, linear process, unless there is an increase to the very limited resources currently allotted the network.

The first pass at SEAN monitoring will focus on the “top six” vital signs identified by the Technical Committee and Network Coordinator: Airborne Contaminants, Freshwater Physicochemical Water Quality, Marine Predators, Oceanography, Streamflow, and Weather and Climate. Initial data management will be implemented exclusively for those signs. For development of some vital signs (particularly Oceanography and Weather and Climate), the SEAN will emphasize establishing partnerships. Partnerships offer a chance to free up scarce staff resources from some dissemination and repository development tasks.

Because the network is being built from scratch, considerable development commitments – both in terms of staff and funding – must be made in the first several years of program implementation. Significant one-time costs will be encountered in setting up basic technical infrastructure and processes. Following basic infrastructure, the next large resource demand will be in implementing the first six vital signs, for which much original staff work will be required. Subsequent vital signs will cost less, as the management framework built for the first six can probably be duplicated and adjusted to accommodate the others with much less effort. The marginal cost for setting up vital sign eighteen will be considerably lower than it was for vital sign one.

The following engagement scheme for implementing the data management plan consists of a series of four steps. Within each step, tasks are listed in expected chronological order.

The time span for implementation is very much a function of the staff and funding resources that will be directed at the detailed tasks. Because those allocations are not yet fully set, no attempt has been made here to establish rigid implementation timelines. Generally, though, the schedule for Vital Signs Monitoring Plan calls for all seventeen protocols to be initiated or completed within seven years.

13.2 Step One – Establish the Core Structure using Six Vital Signs

- Complete and obtain approval of the data management plan.
- Contribute to the development of each vital sign monitoring protocol; identify the detailed technical specifications for all required deliverable data products; identify enough business rules to set final product validation criteria.

- Once protocol is frozen, create a data model for each vital sign; record in NPS Data Store the initial metadata that was developed through modeling.
- Determine information architecture; select which repository each product will reside in; determine dissemination scheme; define data submission unit.
- Create the production auxiliary repository; interface it to the production web server using virtual directories or a similar construct; build and test backup/restore/archive processes.
- Establish partnerships, starting by exploring AOOS for oceanographic data and WRCC for weather and climate.
- Build a development web server.
- Create a development auxiliary repository; interface development auxiliary repository to the development web server.
- Build an application development environment; build and test library and backup/restore/archive processes for source code.
- Create the SEAN home page; design a template for vital sign pages; create vital sign main pages. Determine best methods for invoking I&M enterprise applications from web links.
- Initialize any repository containers that will hold the data products for the initial six signs; generate basic auxiliary repository content for each web page including contact info, FAQs, abstracts, etc.
- Develop programs and processes for automated validation testing of candidate deliverable data products and for extract-transform-loading data products into I&M and partner repositories.
- Build MS Access acquisition database(s) for Project Leader to use in collecting data for each vital sign.
- Build data entry/edit/validation applications for acquisition database. Build routine data summarization/analysis applications for acquisition database.
- Operate a data acquisition cycle for each vital sign.
- Publicize availability of new products to parks and research community.
- Evaluate data acquisition cycle for each vital sign and document issues for next protocol revision.

13.3 Step Two –Establish Mature Data Management Structure

- Revise the DMP based on lessons learned in step one.
- Update the protocols for the first six vital signs based on issues documented in step one. Evaluate work products in light of revised protocols; sequester any step-one deliverables that do not adequately conform to the detailed long-term specifications in the current protocol.

- Re-evaluate the performance realized from partnerships; research new partners.
- Build protocols and data management processes for additional vital signs.
- Operate a data acquisition cycle for each new vital sign.
- Evaluate data acquisition cycle for each new vital sign and document issues for their next protocol revision cycles.

13.4 Step Three –Enhance Repository and Dissemination Core Functions

- Revise the DMP based on lessons learned in step two.
- Build a SQL DBMS-based database to be used as a new development auxiliary repository and/or locate suitable I&M web services that can do the jobs.
- Transition selected existing vital sign products from the file-based auxiliary repository to the development SQL repository as a pilot project; create “binary large object” (BLOB) management services for storing file objects into SQL repository and retrieving them to web pages.
- Build a SQL DBMS-based database to be used as a new production auxiliary repository, or find production I&M services that offer adequate functionality. Move the selected vital sign products used in the pilot to the new production repository; revise web pages so they properly access the new production repository; optimize server/network capacity and performance.
- Transition remaining products from old file-based production auxiliary repository to new DBMS-based repository (with or without web services); abandon file-based repository when transition has completed.
- Review the current GIS tools and technologies available to support web dissemination; determine if it is practical to launch a project for integrating GIS deliverables directly into the SEAN web pages.

13.5 Step Four - Ongoing Operations

- Revise the DMP based on lessons learned in step three.
- Perform the standard data management processing cycles.
- Attempt rehabilitation of vital-sign related legacy data.

13.6 Estimated Workforce Composition for Implementation

The skill sets required for the implementation initially emphasize technical IT development talent in step one, which primarily covers development phase tasks. The need for IT expertise steadily diminishes until Step Four, where requirements are almost entirely for data management and operations skills. The expected distribution of workforce development skills is depicted in Figure 13.1. Operations skills requirements are estimated in Figure 13.2.

For the program to succeed in a reasonable timeframe, the SEAN needs more technical resources at the very beginning. It should investigate how to satisfy this temporary, but essential, need.

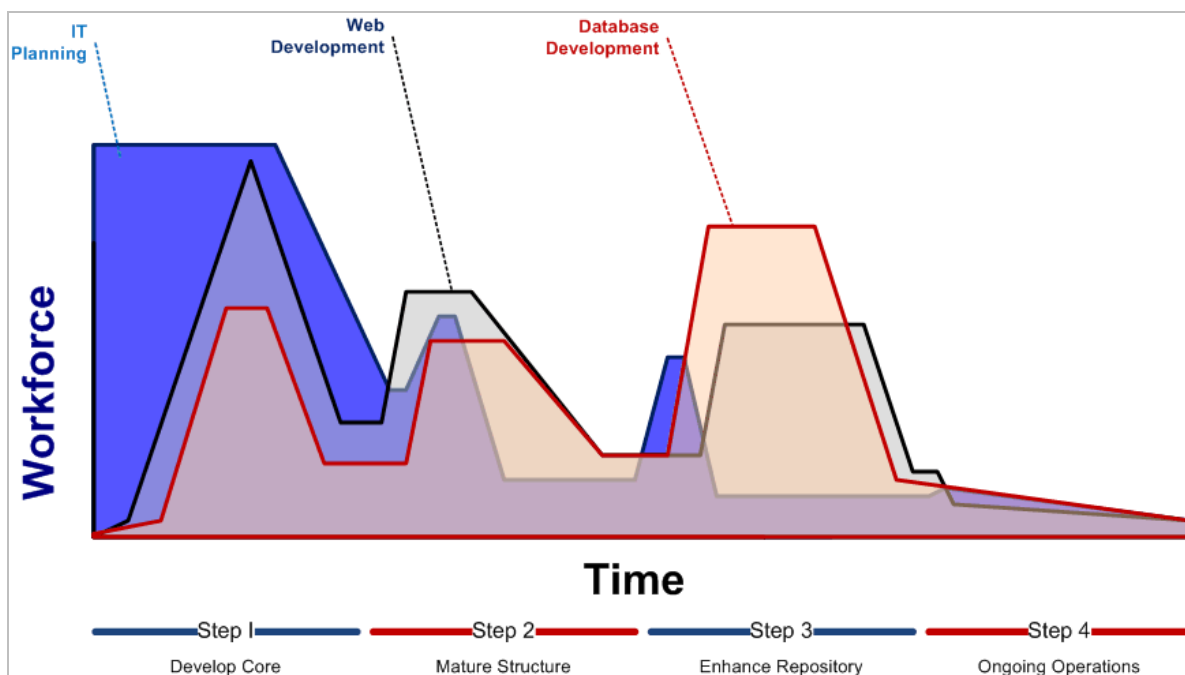


Figure 13.1. Estimated SEAN development skill set requirements.

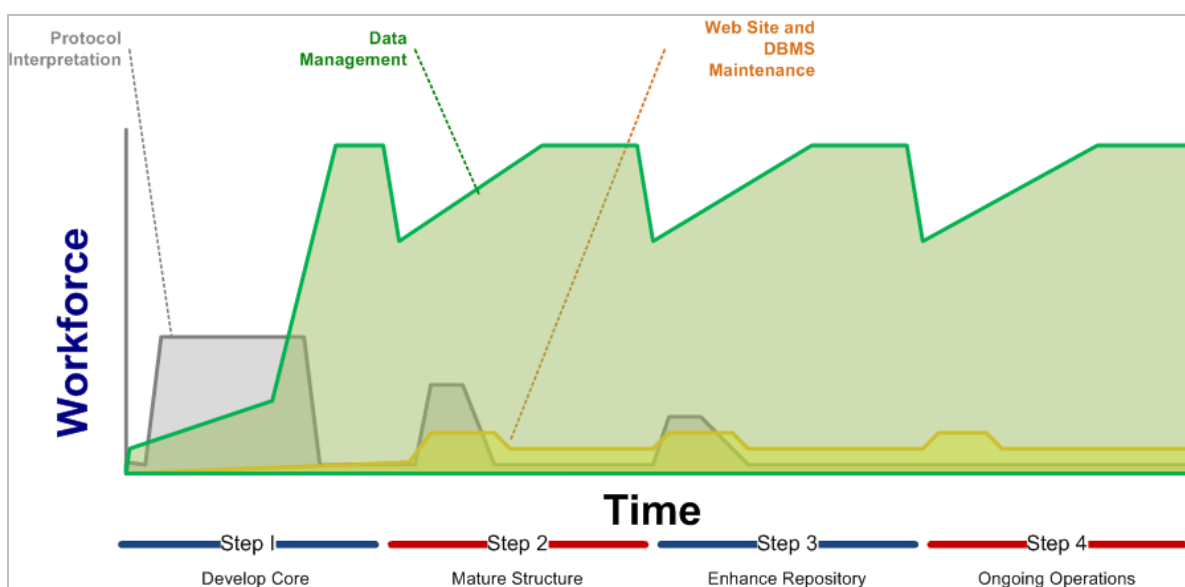


Figure 13.2. Estimated SEAN operations skill set requirements.

Literature Cited

Martin, J. and D. Coleman. 2002. The Archive as an Ecosystem. *Journal of Electronic Publishing*. 7:3.

Moynahan, B. J. and W. F. Johnson. 2008. Vital Sign Monitoring Plan: Southeast Alaska Network. Natural Resource Report NPS/SEAN/NRR—2008/059. National Park Service, Fort Collins, Colorado.

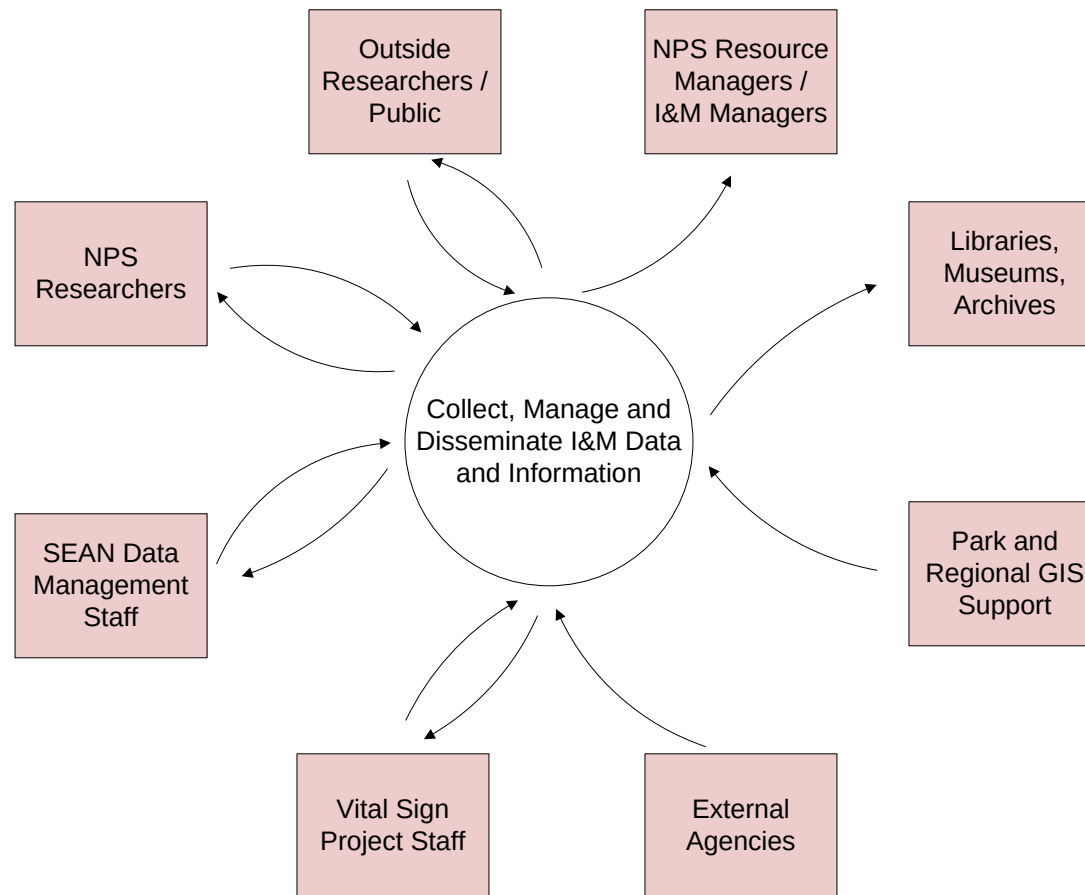
National Park Service. 2008. Data Management Guidelines for Inventory and Monitoring Networks. Natural Resources Report NPS/NRPC/NRR—2008/035. National Park Service, Fort Collins, CO.

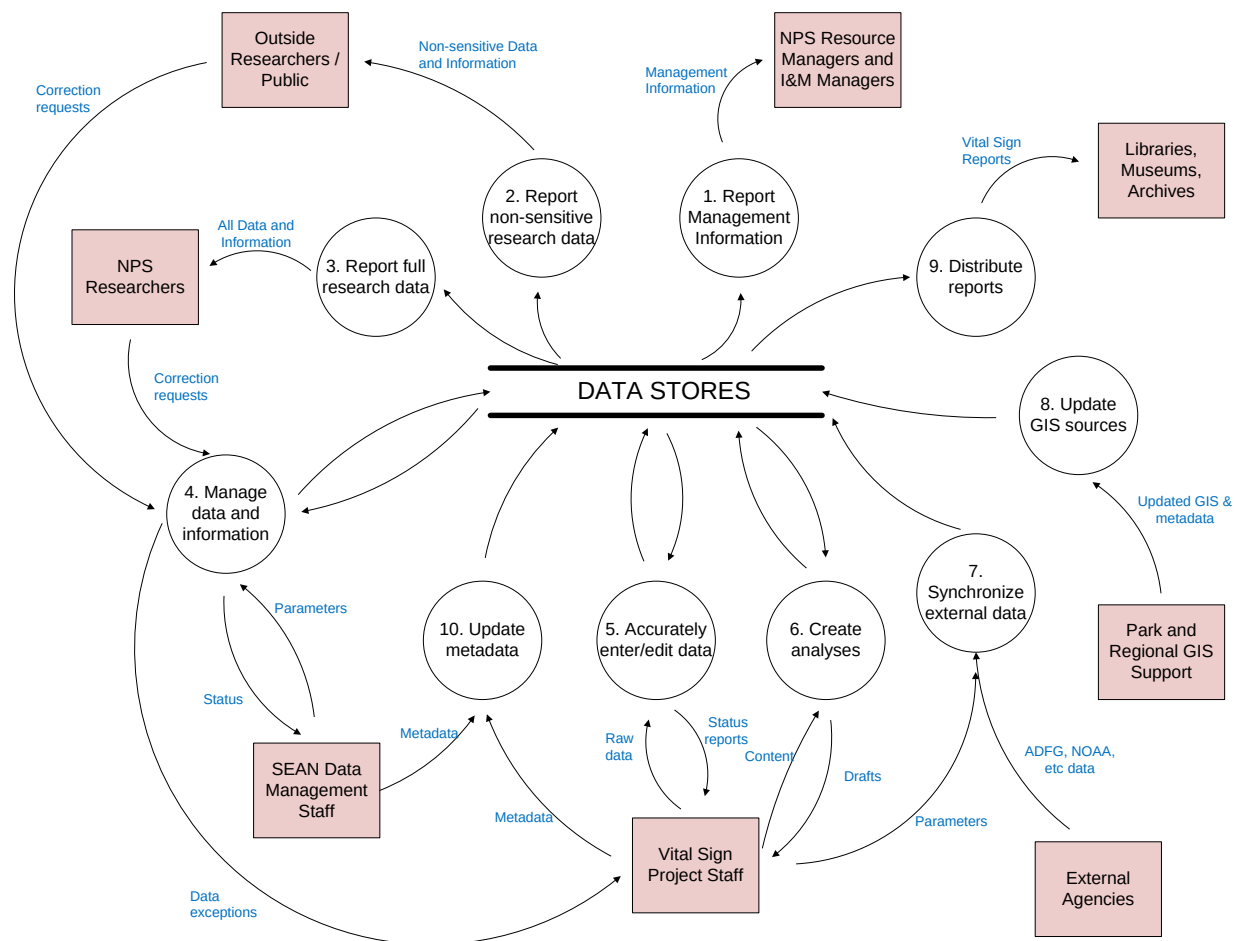
Appendix A: Dataflow Diagrams

Dataflow diagrams illustrate the data that each participant consumes, as well as data they provide. It also identifies the processes that are used to transform data as they move among participants. Finally, it indicates where specific data used for specific purposes are stored.

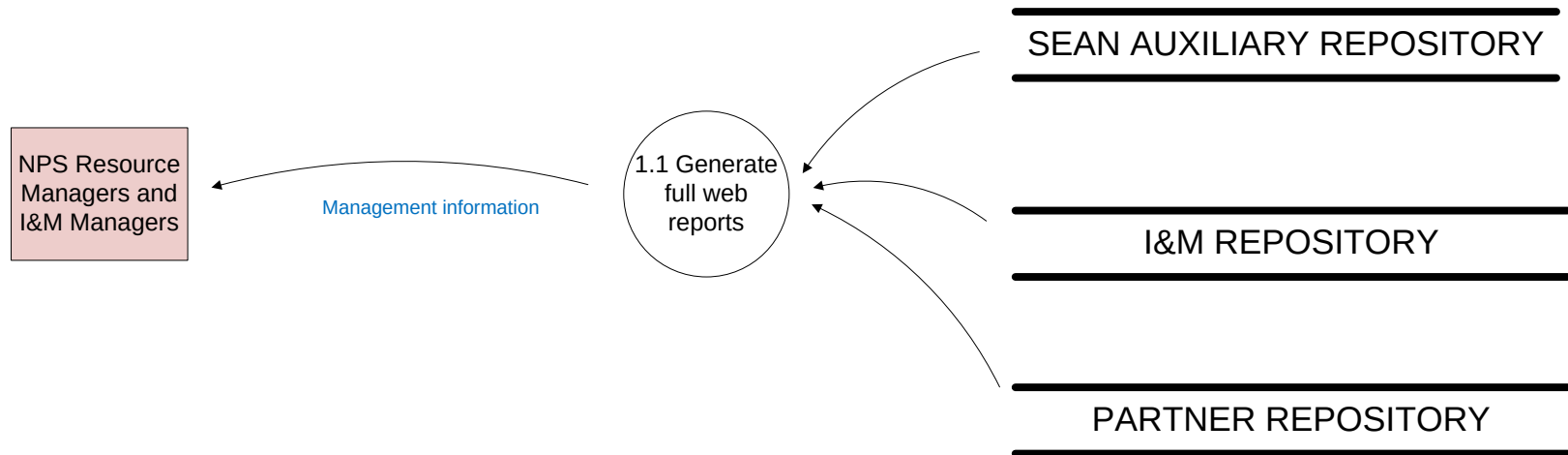
This appendix shows basic data flows within the Southeast Alaska Network.

SEAN logical Dataflow Diagram Level 0

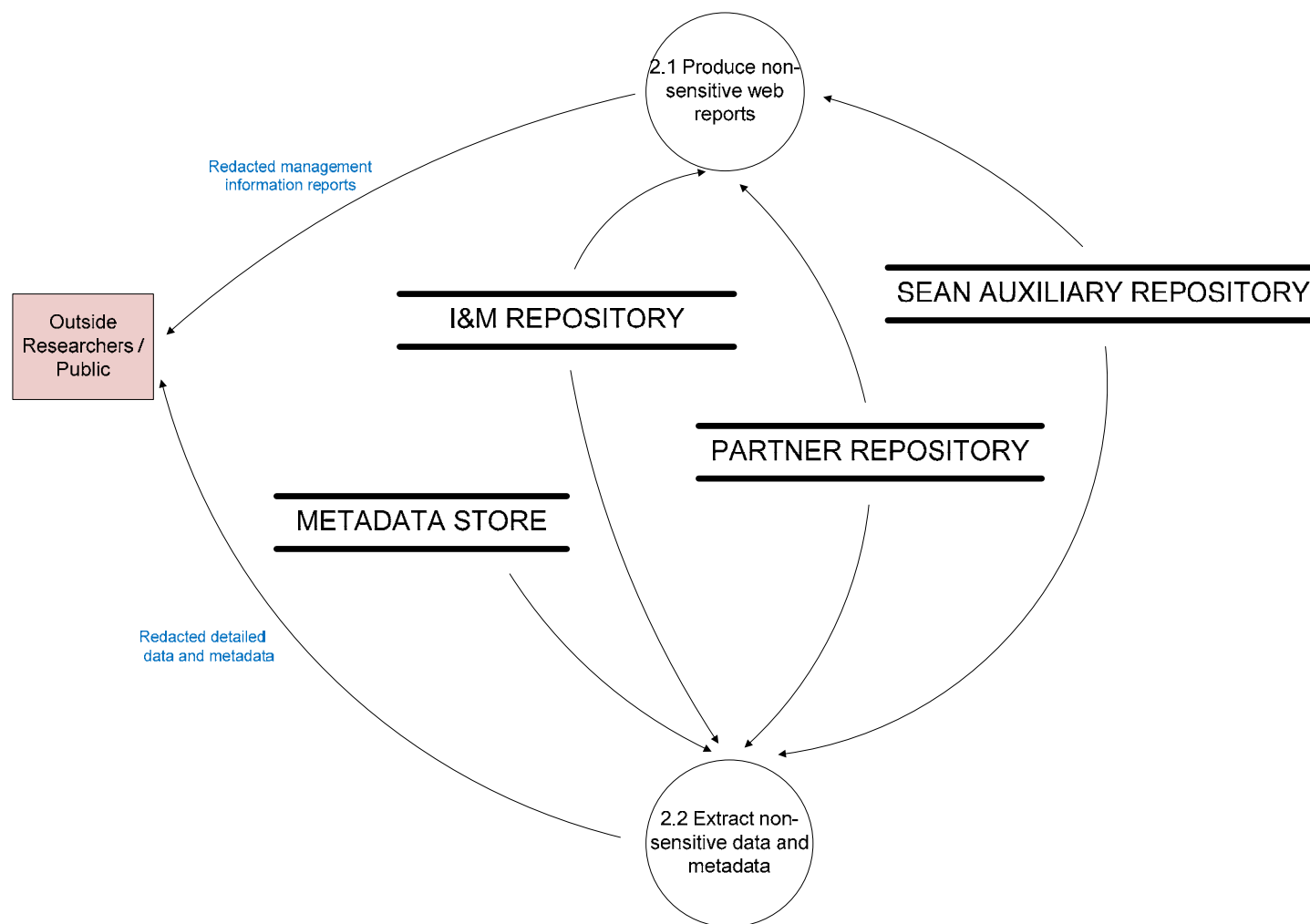


SEAN logical DFD
Level 1

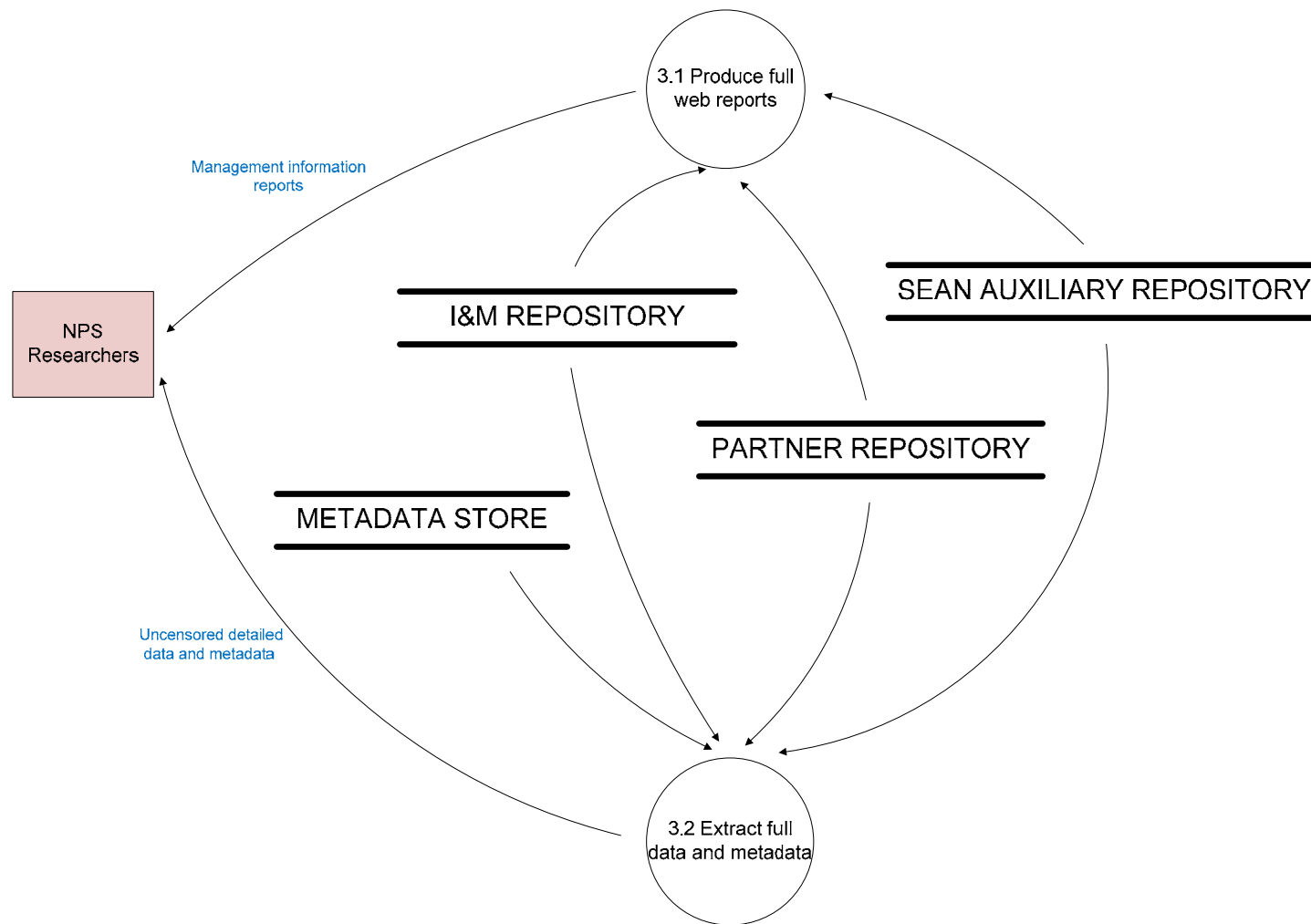
SEAN logical DFD
1 - Report management information



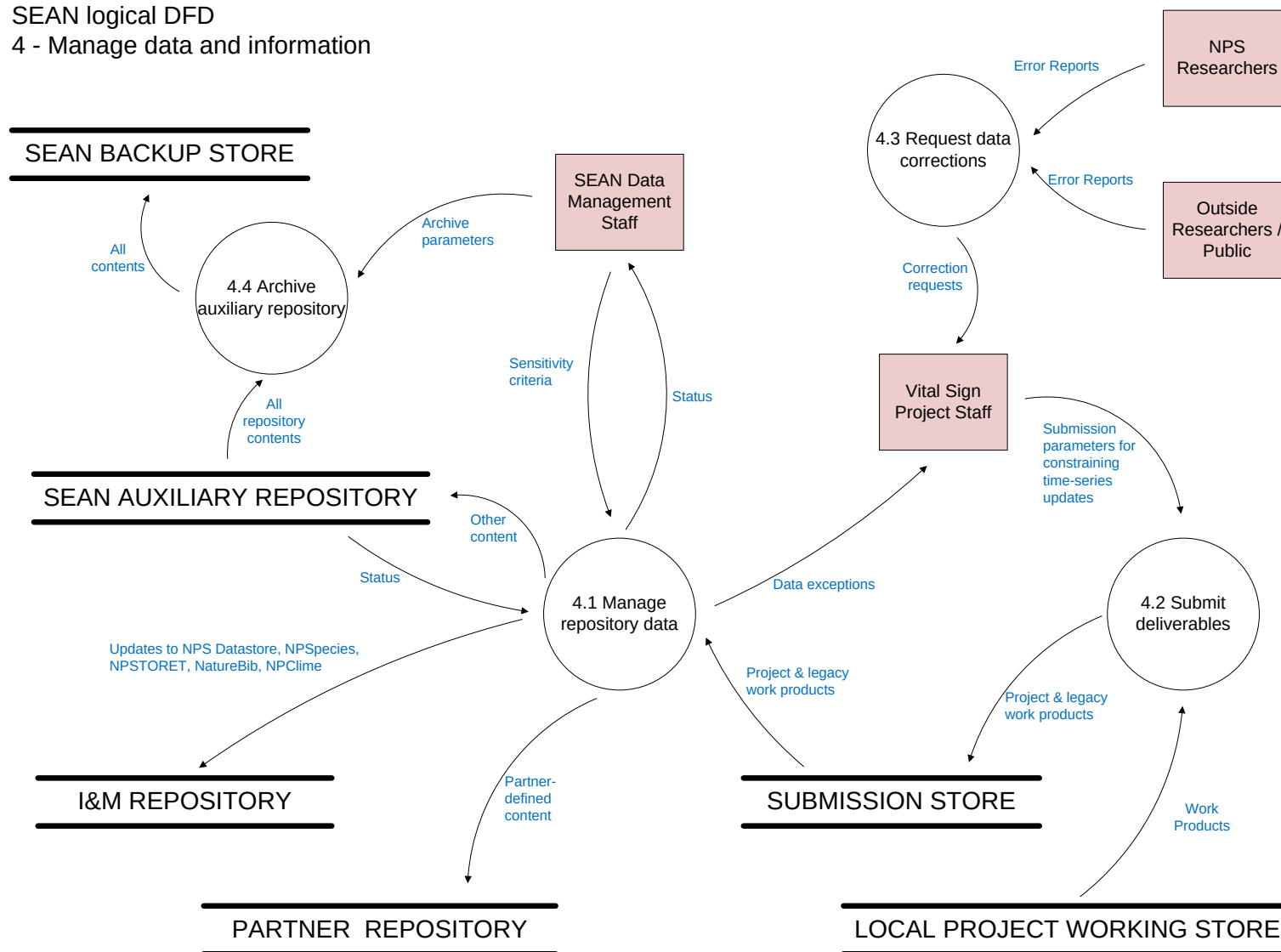
SEAN logical DFD
2 - Report non-sensitive information and data



SEAN logical DFD
3 - Report full research



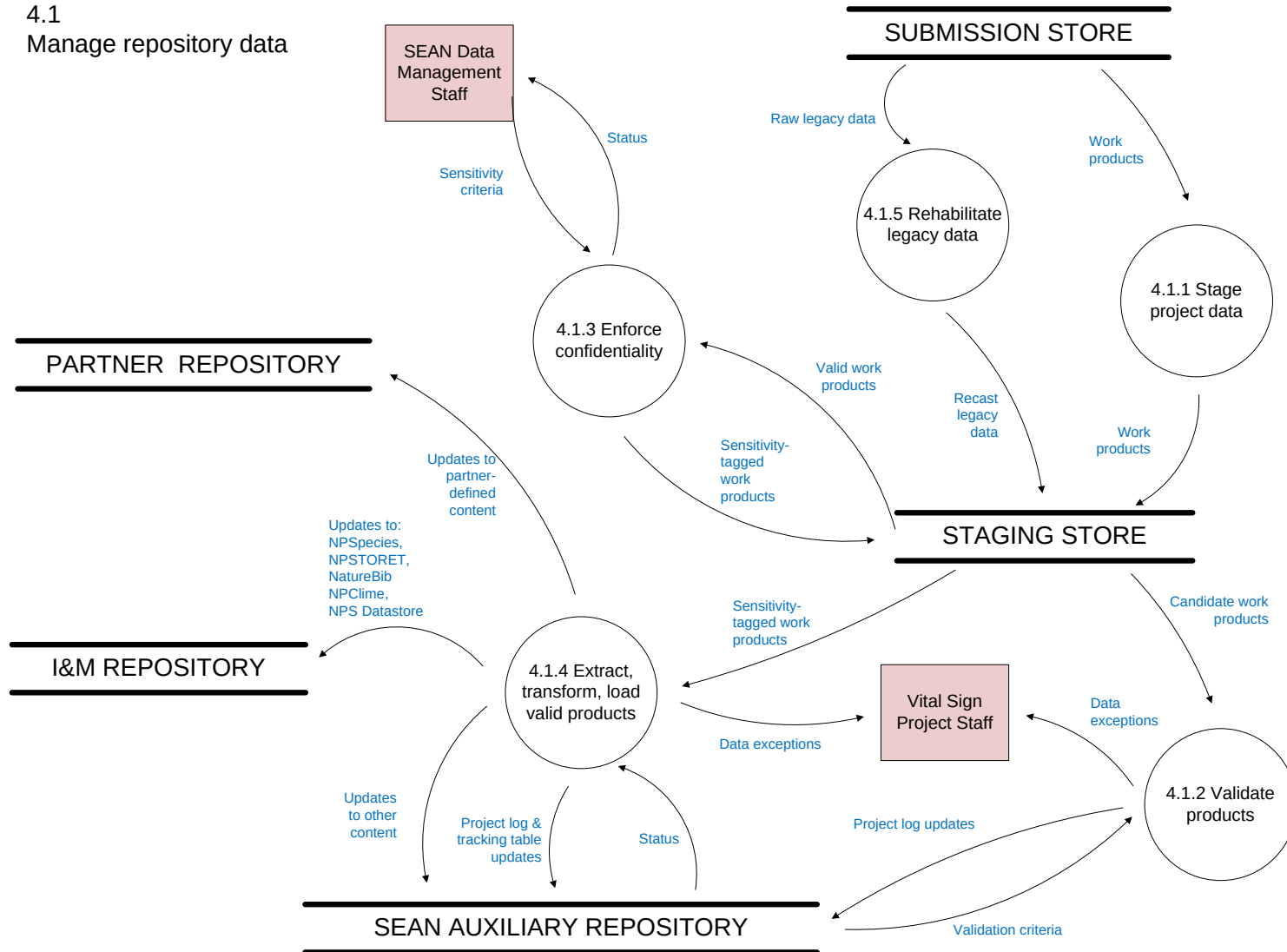
SEAN logical DFD
4 - Manage data and information



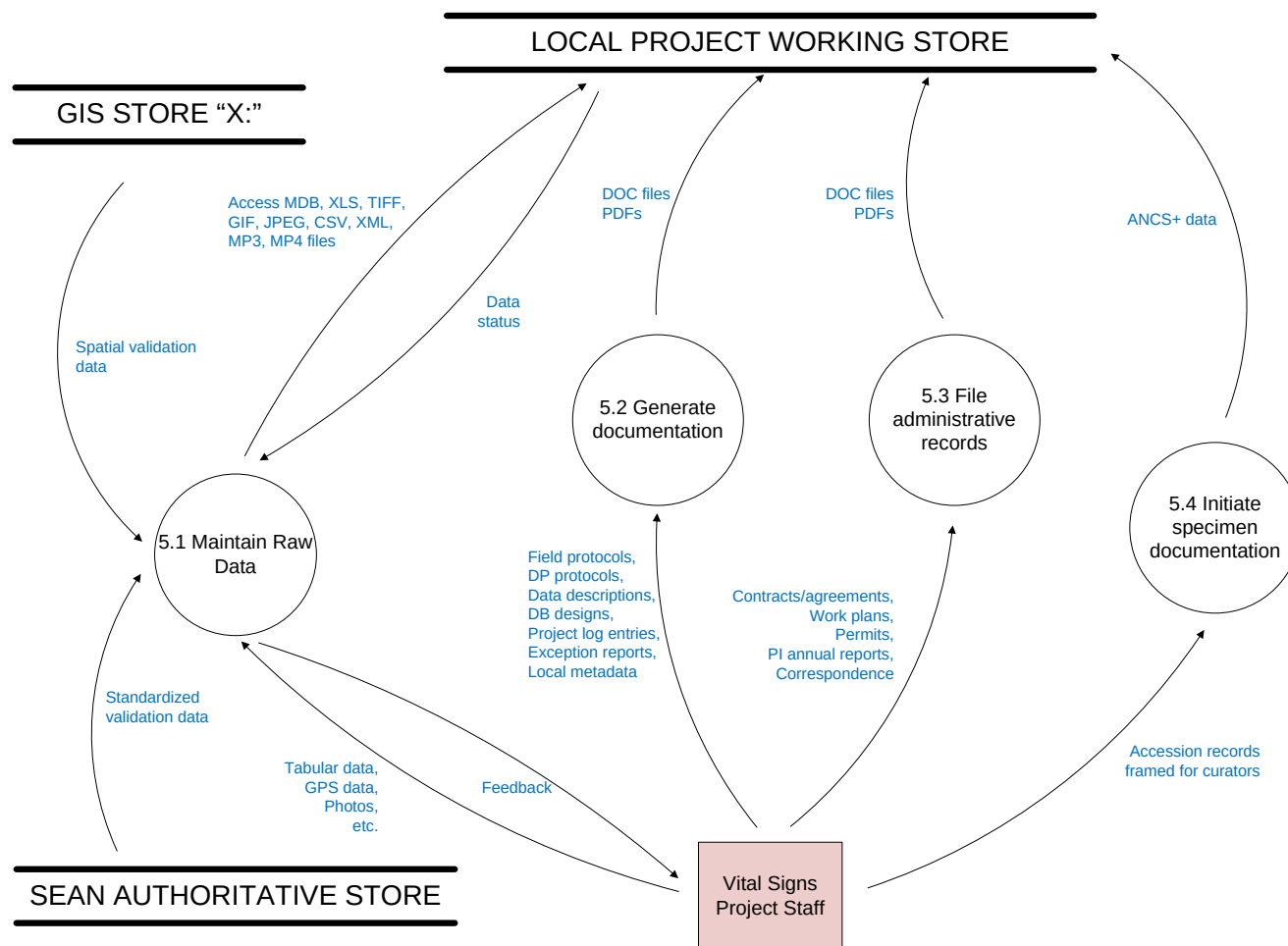
SEAN logical DFD

4.1

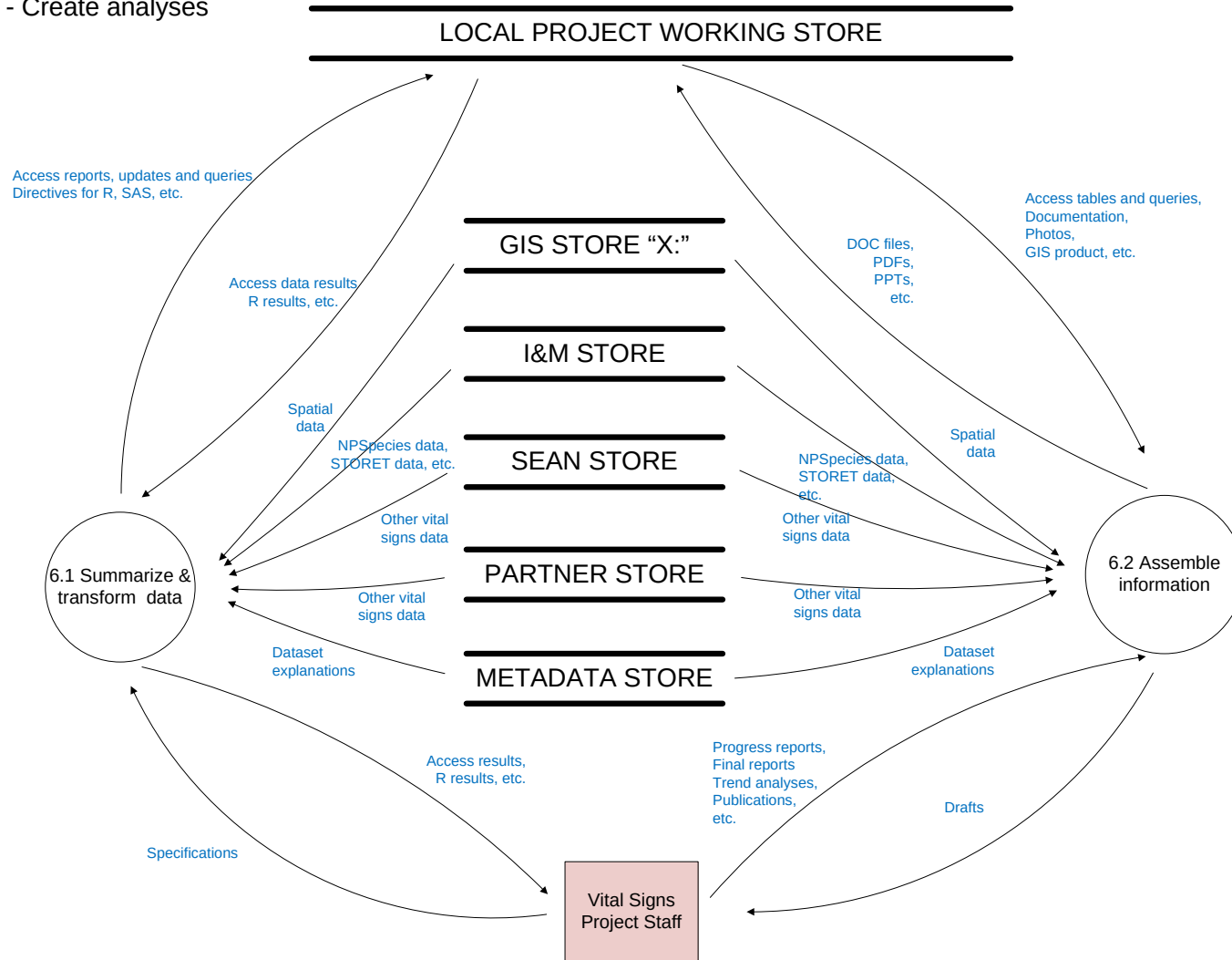
Manage repository data



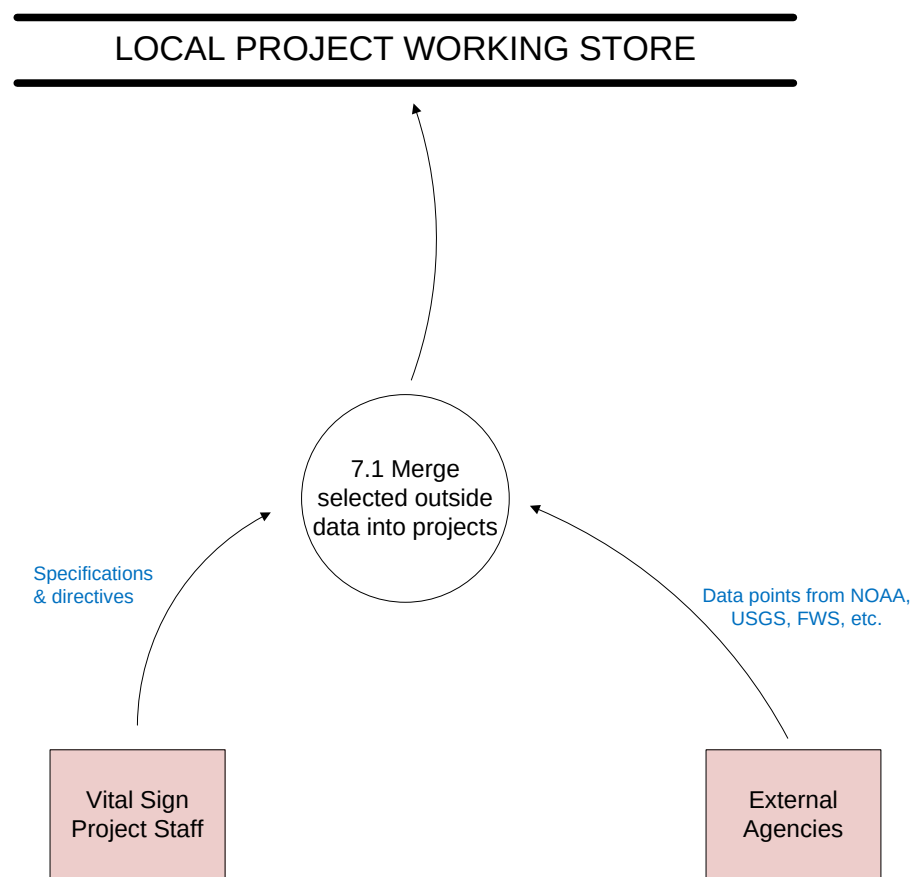
SEAN logical DFD
5 - Accurately enter & edit data



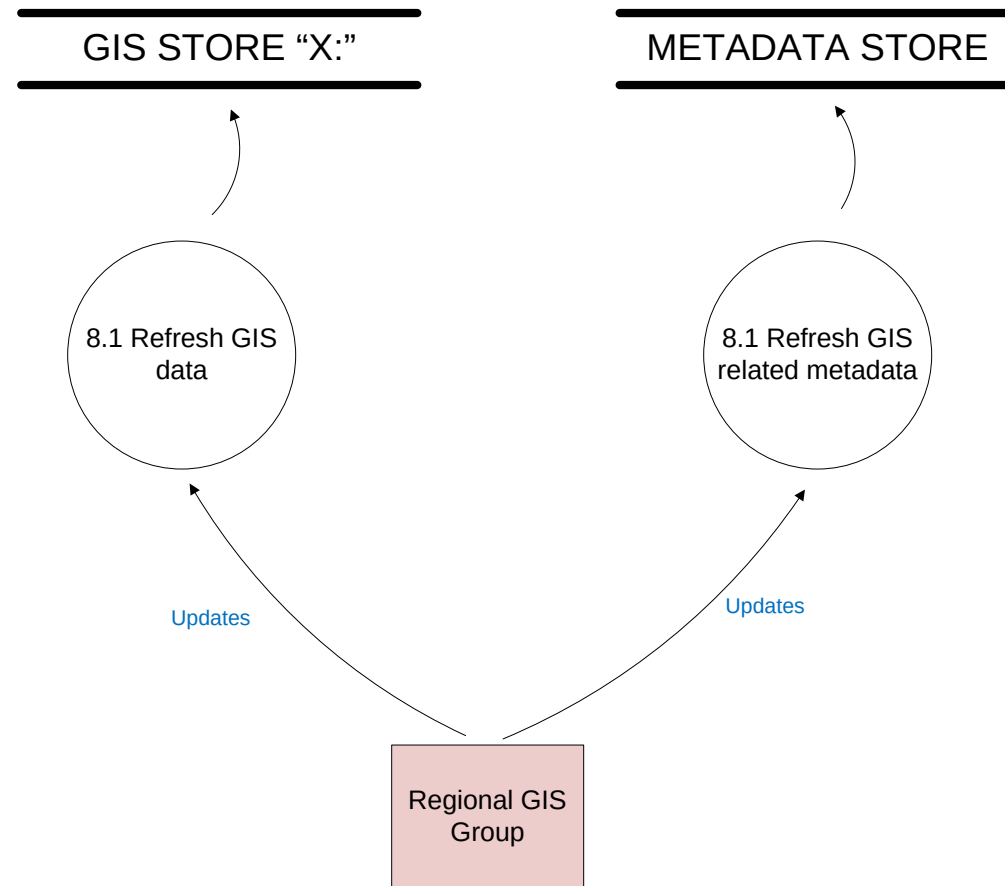
SEAN logical DFD
6 - Create analyses



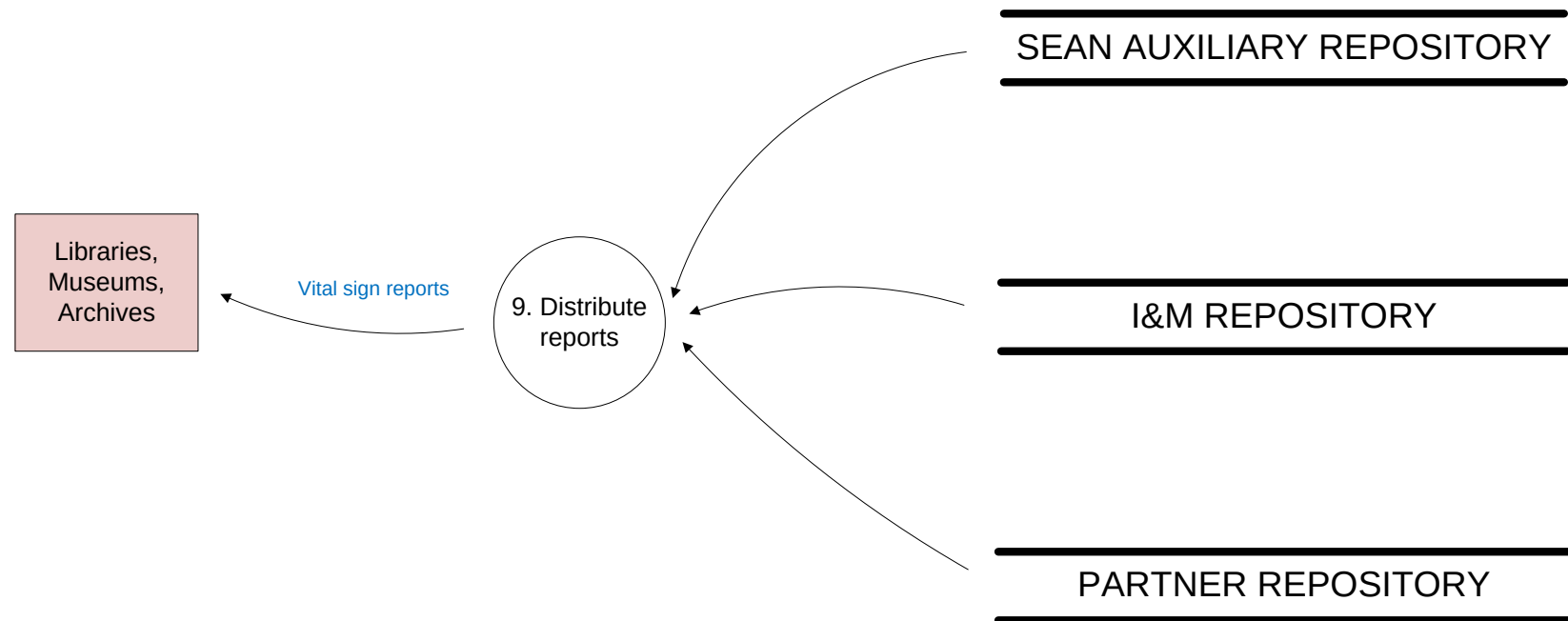
SEAN logical DFD
7 - Synchronize external data



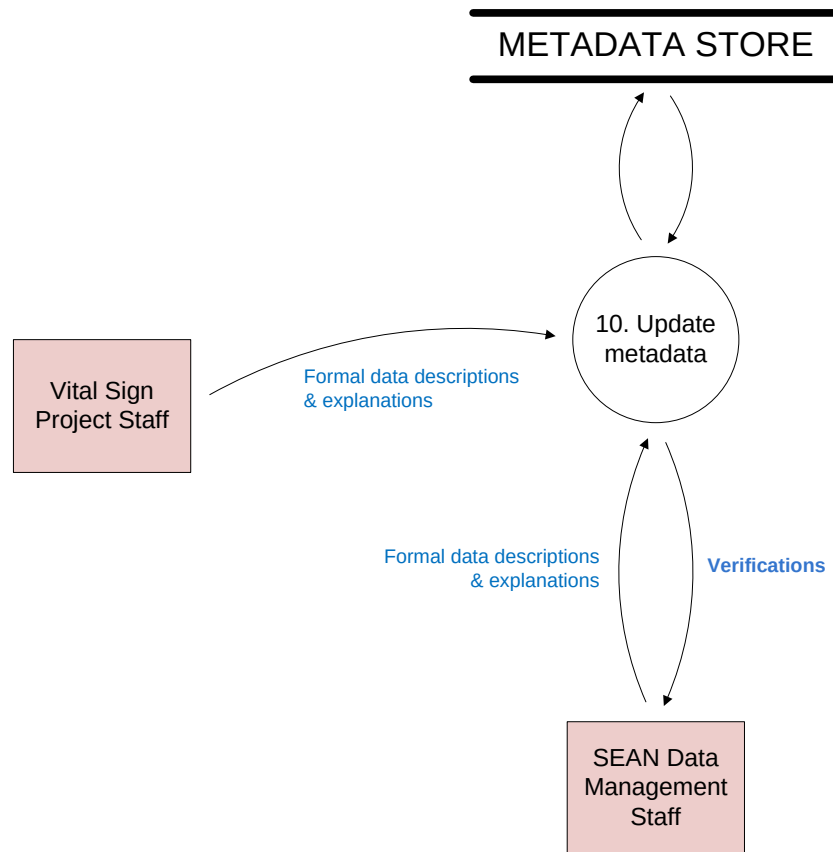
SEAN logical DFD
8 - Update GIS sources



SEAN logical DFD
9 - Distribute reports



SEAN logical DFD
10 - Update metadata



Appendix B: Standard Operating Procedures

The following SOPs are in place for the Southeast Alaska Network. Each SOP maintains its own revision information. SOPs may be revised as needed, without requiring a formal update to the Data Management Plan (DMP). SOPs are numbered with the first digit(s) reflecting the chapter number in the DMP for which the SOP is most relevant.

- SOP 101 – Review Process for Network Guidance Documents
- SOP 201 – Management of Network Internet and Intranet Sites
- SOP 202 – Infrastructure Organization and Maintenance
- SOP 203 – Project Directory Structure Recommendation
- SOP 204 – Backup and Restore Routines for I&M Network Data
- SOP 205 – SEAN Internal Directory and File Naming
- SOP 301 – Checklist for Initiating New Data Management Projects
- SOP 302 – Data Management Considerations in Protocol Development
- SOP 401 – Operational Roles and Responsibilities
- SOP 501 – Managing Spatial Data Products
- SOP 502 – Managing Photographic Data Products
- SOP 601 – Procedures for Certifying Project Data
- SOP 602 – Version Control
- SOP 603 – Documenting Product Provenance
- SOP 1101 – Network Archiving Procedures
- SOP 1201 – Project Tracking



Southeast Alaska Network Standard Operating Procedures

101 - Review Process for Network Guidance Documents

Abstract

The SEAN's data management operations are defined in a hierarchy of three guidance documents: the national data management plan, network data management plan, and appendix B of the network plan concerning standard operating procedures. When a new version of the national plan is published by I&M, the SEAN data manager must review the network data management plan and revise it so it is in conformity with national guidance.

The SOP appendix must be updated by the data manager whenever new operating process details are identified. The SOP appendix must be in conformity with the network data management plan at all times.

This version saved: September 24, 2008

Table of Contents

1	Network Guidance Documents.....	3
2	Development of SOPs	3
3	Updating the Network Data Management Plan.....	3

Revision history

Revision Date	Base Version	Author	Nature of Change
3/14/2008	-new-	Bill Johnson	Initial version
7/24/2008	3/14/2008	Bill Johnson	Peer review revisions

1 Network Guidance Documents

Overarching guidance for network data management is defined by the national data manager at I&M WASO. The single document providing this guidance is the NPS national data management plan: “Data Management Guidelines for Inventory and Monitoring Networks.” The first version of this was published in Ft. Collins in 2008. It consists of 13 well-defined chapters.

The SEAN has developed its network data management plan, initially published in Juneau in 2008. The network plan is subsidiary to and conforms to the national plan. Its structure matches the 13 chapters in the national plan.

Within the network plan is an appendix that lists the standard operating procedures to be used in running the network. The SOP appendix is in conformance with the network plan. It uses a numbering scheme where the most significant digits indicate the plan chapter being addressed.

2 Development of SOPs

Standard operating procedure documentation is created as the network is implemented. The network data manager is responsible for generating new, and updating existing, SOPs. The network program and supporting SOPs must be developed within the guidelines of the network data management plan. No formal review of the SOP appendix is required. However, whenever the network data management plan is revised, it is peer reviewed. If resources permit, the SOP appendix may also be scrutinized at that time.

3 Updating the Network Data Management Plan

Each time a new edition of the national data management plan is published, the network plan must be revised to make it conform to national guidance. The network plan is revised by the SEAN data manager in consultation with the SEAN network coordinator. Regardless of action at the national level, the network plan must be revised at least once every five years in order to keep it current with fast-changing accepted IT methods. Updates to the network plan require review of existing SOPs by the data manager. Updates to the network plan may also require additional SOP topics be addressed in the appendix.

A revised network plan must be approved by the national data manager before being published and used.



Southeast Alaska Network
Standard Operating Procedures

**201 - Management of Network Internet and Intranet
Web Sites**

Abstract

The SEAN relies on two classes of Internet web sites. The first is housed at I&M and disseminates the contents of the I&M and Auxiliary repositories. The second are web sites of individual partners with whom SEAN has agreements. Due to SEAN's limited staff, its intranet site is a stub pointing to the Internet site.

This version saved: September 24, 2008

Table of Contents

1	I&M Production Web Site.....	3
2	SEAN Development Web Site	3
3	Partner Production Web Sites.....	3

Revision history

Revision Date	Base Version	Author	Nature of Change
3/14/2008	-new-	Bill Johnson	Initial version
7/24/2008	3/14/2008	Bill Johnson	Peer review revisions

1 I&M Production Web Site

Customers pull most SEAN information from the single I&M web site whose root is at URL *HTTP://SCIENCE.NATURE.NPS.GOV/IM/UNITS/SEAN/*. The production web server is located at I&M in Fort Collins. The I&M repository is collocated with the web server. The Auxiliary repository is mapped to the Fort Collins web server as a virtual directory. The Auxiliary repository is physically located in Juneau. The web server runs Microsoft IIS and supports ASPX.

Web content is generated by SEAN staff. It is produced in a development environment. Updates are copied to the production site using MS Visual Studio tools over a Windows share.

Management of the production web site, including backing up the web programming and content housed on the web server, is the responsibility of I&M technical staff.

2 SEAN Development Web Site

Web application development is performed at the network office in Juneau. It is done using the MS Visual Studio IDE. Unit testing is done at the developer's workstation with a local web service. NPS security prevents local workstations from serving web pages to other nodes. System testing that spans multiple users is done using an IIS web server installed on the SEAN file server, INPGLBAFS03.

Source code is stored at \\INPGLBAFS03\SEAN_Data_Mgmt\Web_Sites\Internet. Standard Visual Studio project directory structures are employed. Because the web staff is one person, the effort and expense of a source code control system has not been undertaken.

3 Partner Production Web Sites

SEAN is attempting to partner with certain existing agency web sites to deliver SEAN-collected information. Until partnership agreements are executed, there is no concern regarding maintenance.



Southeast Alaska Network
Standard Operating Procedures

202 - Infrastructure Organization and Maintenance

Abstract

The SEAN's IT infrastructure is straightforward. In the Juneau office, each of two staff has a workstation connected to the local fileserver INPGLBAFS03. The LAN is connected to the NPS network with a leased T-1 line. Most support is provided by GLBA IT staff. Production web servers are part of the I&M Fort Collins infrastructure.

This version saved: September 25, 2008

Table of Contents

1	In Case of Emergency.....	3
2	SEAN Infrastructure Overview	3
3	Google Search Appliance	3
4	SEAN Building Infrastructure Schematic	4
5	SEAN Equipment Architecture Overlaid with Functional Components.....	5

Revision history

Revision Date	Base Version	Author	Nature of Change
3/18/2008	-new-	Bill Johnson	Initial version
7/25/2008	3/18/2008	Bill Johnson	Peer review revisions
9/22/2008	7/25/2008	Bill Johnson	Added equipment architecture figure

September 25, 2008

202 - Infrastructure Organization and Maintenance

Page 2 of 5

1 In Case of Emergency

For most network, server and workstation outages: GLBA IT support at 907.697.2663.

For failed T-1 line or router: Verizon NOSC, 866.376.6672 – specify IP address 165.83.57.225.

For failed production web server: Jay Rhoten, 970.267.2178 or Gina Glodowski, 970.267.2124.

2 SEAN Infrastructure Overview

SEAN is a two-person office in Juneau. Each staff member has a desktop workstation. Their files are kept on file server INPGLBAFS03. There is one shared laser printer. Network connectivity is over a T-1 leased line from Verizon. The line has a low capacity of 1.54 megabits/second. This infrastructure is shared with NPS Coastal Cluster program.

The office is shared with USGS. However, no network interconnections exist; the two agencies are completely separate.

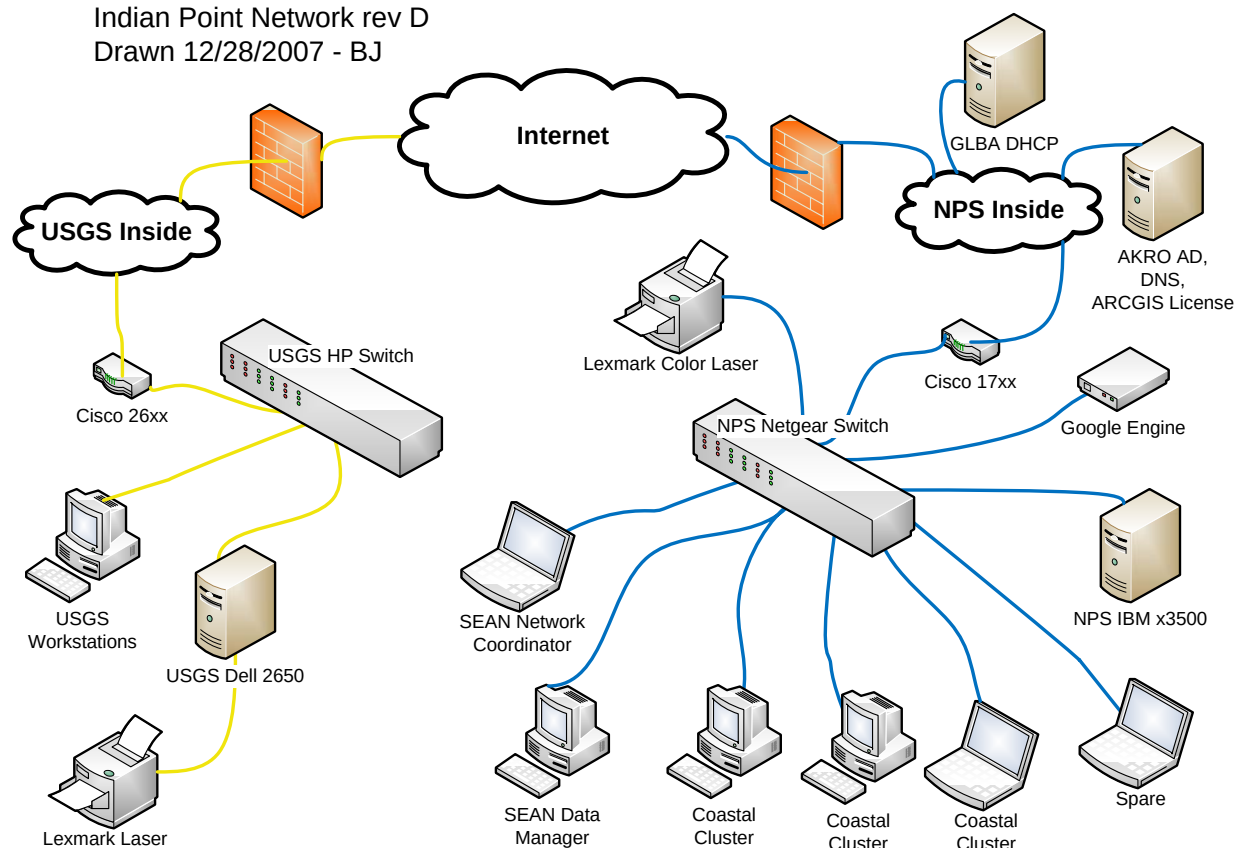
Most network, server and workstation support is done by the GLBA IT staff. They also provide the DHCP IP address leasing. Additional network services provided by AKRO in Anchorage include Active Directory and DNS. AKRO also houses the licensing server SEAN relies on for ESRI access. Production web serving is not part of SEAN's infrastructure. For that, we depend on I&M in Fort Collins. Web development, however, takes place on Juneau workstations using local web services.

3 Google Search Appliance

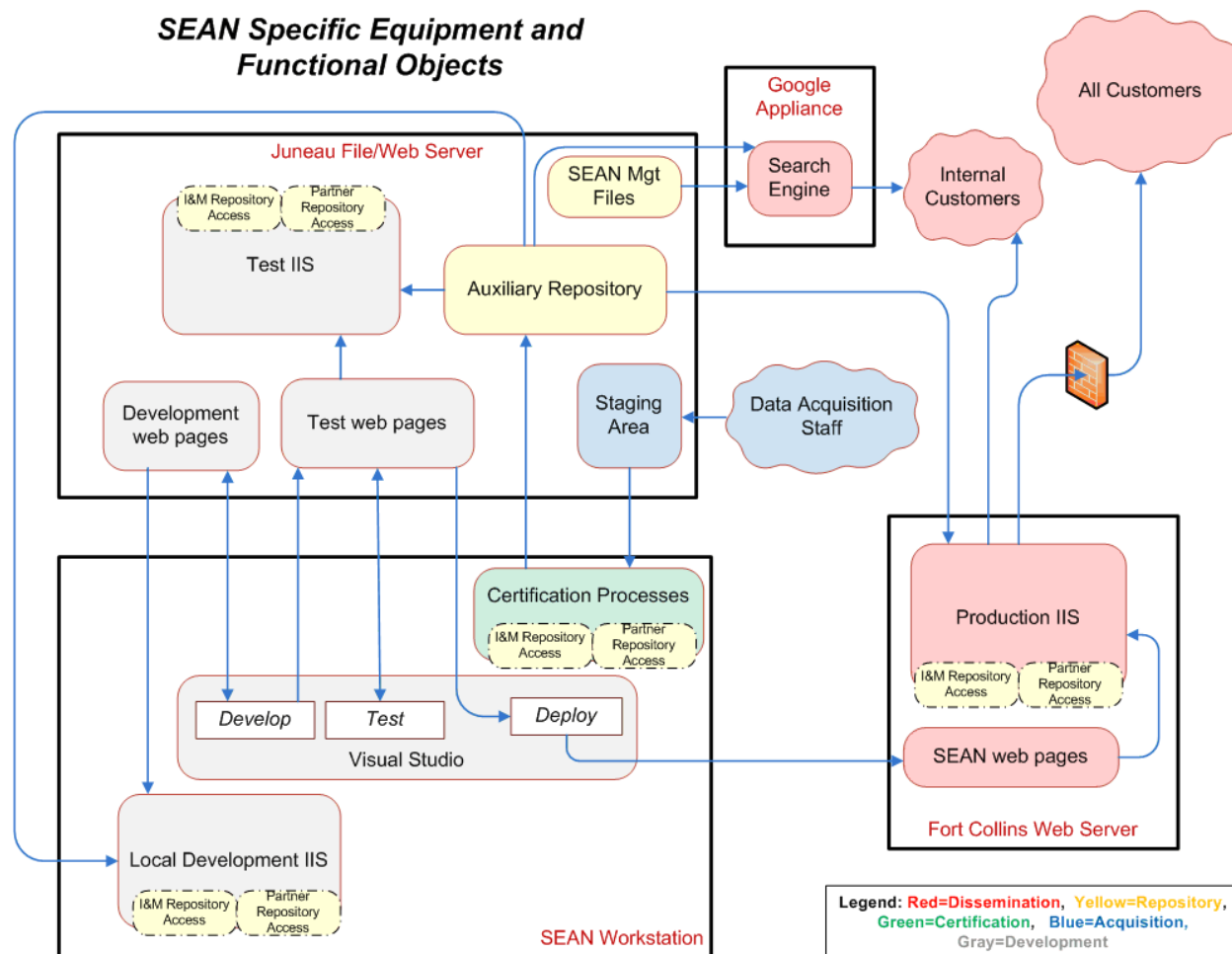
A local Google "Mini" appliance is maintained by SEAN staff. It is used to index and cache the contents of the INPGLBAFS03 file server, not of a web server. Google mini support is available at <https://support.google.com/enterprise>. Predictably, they do not publicize any phone contacts.

4 SEAN Building Infrastructure Schematic

Indian Point Network rev D
Drawn 12/28/2007 - BJ



5 SEAN Equipment Architecture Overlaid with Functional Components





Southeast Alaska Network
Standard Operating Procedures

203 - Project Directory Structure Recommendation

Abstract

In the course of a project, leaders are responsible for collecting, manipulating and submitting to SEAN data of many types. Such data files are stored in the project leader's locale, not at SEAN. In order for other project participants to readily locate objects among a project's set of computer files, leaders are requested to follow a standard directory structure. The SEAN does not maintain, regulate, or force specific usage of the many draft, temporary, and ancillary files managed locally by the project leader for a specific vital sign.

(Files maintained in the SEAN repositories, web sites, staging areas, and archives employ special directory layouts unrelated to those covered in this SOP.)

This version saved: September 24, 2008

Table of Contents

1	Project Organization Template	3
2	Directory Structure Layout.....	3
3	Specific Usage of Directory Entries	4
4	File Naming Standards	4

Revision history

Revision Date	Base Version	Author	Nature of Change
3/24/2008	-new-	Bill Johnson	Initial version
7/24/2008	3/24/2008	Bill Johnson	Peer review revisions

September 24, 2008

203 - Project Directory Structure Recommendation

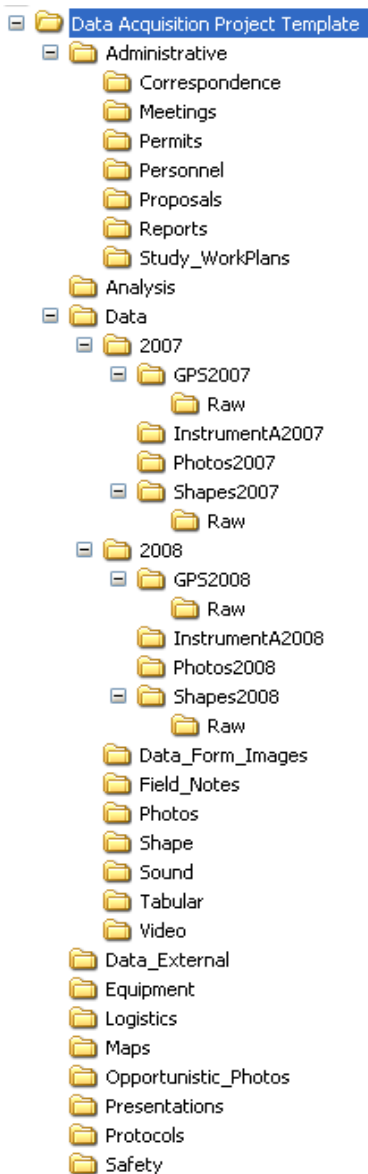
Page 2 of 5

1 Project Organization Template

A template has been developed by GLBA Resource Management staff to organize the files of their projects. Since this template is well-designed and is already used by many associates, SEAN project leaders are requested to observe this form when storing project files. Following this structure will make it easier for staff and management of SEAN projects to locate specific files. SEAN has slightly simplified the original structure in order to facilitate its use.

This SOP covers data acquisition projects only. Files maintained in the SEAN repositories, web sites, staging areas, and archives employ special directory layouts unrelated to those covered here.

2 Directory Structure Layout



3 Specific Usage of Directory Entries

All files in these structures are used exclusively by the project staff to develop deliverable work products. The SEAN does not maintain any intermediate or draft items in its repositories. SEAN repositories only contain completed deliverables as explicitly defined in a project's protocol. Many files in the project folders typically don't make it into SEAN's repositories.

Most of the directory names are self-explanatory. To avoid confusion, several clarifications are made below.

Administrative/Reports contain documents like progress reports and IARS used in administering the program. Reports showing analysis of scientific information belong under "analysis".

Analysis contains deliverable products in the form of Word documents or Adobe PDF files.

Data stores raw data obtained from devices. If multiple years of data are needed by the project leader to generate analyses, the data files should be segregated by year.

Data/Tabular is for MS Access MDB/ACCDB files.

Data_External holds data files obtained from outside the project. These are typically tabular data which are combined with project observations so inferences may be made.

Maps should hold the complete set of GIS files required to make project deliverables.

Opportunistic_photos are JPEGs for public communications use. Actual photos used for analysis should be saved in *Data/Photos*.

Presentations is the place to store PowerPoint PPT files and video MP4 files of recorded talks.

Protocols may contain a convenience copy of the project protocol. It may also contain working notes regarding protocol operation and recommended improvements. The actual protocol that a project leader is held accountable to, however, is the authoritative copy found on the project's main web page.

4 File Naming Standards

The project leader is responsible for, and is the primary (or exclusive) user of the local project directory. A particular vital sign may require hundreds of files of dozens of types to be generated during the data acquisition function. The SEAN will not micromanage scientists by insisting on specific name formats for every single file generated.

However, it is reasonable for SEAN to require that application-specific default file name extensions be employed wherever possible. That is, all Word2003 documents should be saved

with the extension of '.DOC'. All Access2007 database files should be filed as '.ACCDB', and so on. This will allay confusion regarding file content on the part of the project leader, project staff, and affiliated investigators needing to use project files.

File naming for products the SEAN, itself, accepts and maintains are a separate issue. This is covered in SOP 205.



Southeast Alaska Network
Standard Operating Procedures

204 - Backup and Restore Routines for I&M Network Data

Abstract

The SEAN develops data in a number of locations. Backup and restore is done by SEAN only for locally maintained data and programming on the Juneau file server. It is under the day to day operation of GLBA IT staff.

This version saved: July 30, 2008

Table of Contents

1	In Case of Emergency.....	3
2	SEAN Data Overview	3
3	Backup and Restore of I&M Repository	3
4	Backup and Restore of Partner Repositories	3
5	Backup and Restore of Auxiliary Repository.....	4

Revision history

Revision Date	Base Version	Author	Nature of Change
3/14/2008	-new-	Bill Johnson	Initial version
7/24/2008	3/14/2008	Bill Johnson	Peer review revisions

1 In Case of Emergency

To restore Auxiliary Repository items from backup: Patrick Chipley, GLBA, 907.697.2663.
To request I&M Repository restores: Margaret Beer, I&M, 970.267.2168.

2 SEAN Data Overview

SEAN is charged with developing data for its identified vital signs monitoring and inventory. Each data set is stored in one and only one repository of three possible types: I&M, Partner, or Auxiliary. The I&M repository holds various classes of NPS-wide data. In particular, its NPS Data Store facility is relied on to hold most SEAN tabular data as well as all of its metadata. The Partner repository holds specifically agreed on data elements for particular purposes. The Auxiliary repository is currently implemented as files on a local Windows file server in Juneau. The owner of each repository also always provides a web dissemination facility for distributing the data widely.

In keeping with its authoritative source policy, only one copy of each particular vital sign data is kept for use. However, since SEAN has no control over the backup and archiving operations of Partners, partner submissions are stored for safekeeping in Juneau. The safekeeping store is only for business continuity/disaster recovery purposes after a partner suffers catastrophic loss.

The following procedures cover checkpointing and restoring damaged repositories by technical experts. Customers never have direct access to repository files. The only way for customers to obtain data and information is through the web site. Only system administrators and the web service have permission to read the repositories.

3 Backup and Restore of I&M Repository

The I&M repository is professionally managed by fulltime NPS staff in Fort Collins, Colorado. Data entrusted to I&M is wholly contained at I&M, who are responsible for backup and restore. The SEAN does not maintain additional copies of these data outside the I&M repository. In the event recovery is ever required, contact the I&M WASO technical staff to request it.

4 Backup and Restore of Partner Repositories

It is the responsibility of partners to backup and restore SEAN data entrusted to it. But if partners fail to perform, data may still be recovered. The individual submissions made to partners have safekeeping copies stored by SEAN at `\\INPGLBAFS03\DATA\SEAN DATA\PARTNER SAFEKEEPING`. In the event of catastrophic loss by a partner, the stream of submissions used to populate the partner's data set, along with the order they must be applied in, may be sent to the partner. The safekeeping store is backed up by the processes in section 5, below. It is not available for use other than repopulating a failed partner.

5 Backup and Restore of Auxiliary Repository

The auxiliary repository is currently stored in Juneau at \\INPGLBAFS03\DATA\SEAN\DATA\AUXILIARY_REPOSITORY. It contains much of the material distributed by SEAN's only dissemination point: its Internet web site. These data are the authoritative source for many vital sign projects. It is essential backup and restore be treated professionally on this site. The contents are irreplaceable.

Routine daily backup of the entire volume \\INPGLBAFS03\DATA\SEAN_DATA is managed by GLBA IT staff. This captures the repository, the partner safekeeping copies, and the internal staging area. Backup medium is high-capacity USB disk. After a drive has received backups for a week, it is swapped with another drive. The original disk is taken off-site and stored at the data manager's home. One disk is designated F: and the other is G:. At any given time either F: or G: is attached to the server, but never both.

EMC "Retrospect" software is used to manage the content. Retrospect maintains a history of multiple file versions. The exact number of versions stored varies and cannot be guaranteed at a specific level. When a destination drive is filled, backups are saved on top of the oldest versions. Retrospect runs daily as a scheduled task, both on the F: drive and on the G: drive. Since only one drive is connected to the server at a given time, one of the scheduled jobs fails every day. This is by design and does not indicate a problem. If both jobs fail, however, data are at risk. The jobs are monitored over the network by GLBA IT staff in Bartlett Cove. Because the drives are used for no more than a week, it can only be assured SEAN has at least the last two weeks of data. To restore beyond the two week barrier, refer to the archiving SOP in the 1100 series.

Restore is accomplished by logging onto INPGLBAFS03 as an administrator. The RETROSPECT software is available from the START menu. It offers a rich, guided interface that leads an operator through the restore process. Unless one is comfortable with this tool from prior use, it may be safer to explain restore needs with GLBA IT staff and invite them to operate the facility. Errors in restore specifications could cause serious data loss.

A representative repository restore is tested every year during the last week of December, to be supervised by the data manager. Additional tests must be performed after every upgrade to the RESTROSPECT software.



Southeast Alaska Network
Standard Operating Procedures

205 - SEAN Internal Directory and File Naming

Abstract

SEAN products are stored in its auxiliary repository and in outside repositories. Certain ancillary items such as web pages, partially-staged candidate deliverables, and disaster recovery files are stored directly on SEAN's single file server.

Directory layouts and file naming standards are documented for structures that SEAN directly maintains.

SOP 203 should be referenced for directory structures to be used by projects.

This version saved: September 24, 2008

Table of Contents

1	Repository Structures	3
2	Other Structures.....	3
3	Structure Search Ability	4
4	Vital Signs Identification Prefixes.....	4
5	Directory Structure Layout.....	5
6	File Naming Standards	7

Revision history

Revision Date	Base Version	Author	Nature of Change
8/4/2008	-new-	Bill Johnson	Initial version
9/11/2008	8/4/2008	Bill Johnson	Marine Contaminants added

1 Repository Structures

SEAN data products are stored in either the I&M repository, partner repositories, or the SEAN's own auxiliary repository. The SEAN has neither the jurisdiction nor the need to control internal data management of I&M and partner repositories. SEAN only requires that these entities accurately and effectively serve SEAN's specific products to SEAN's customers.

The auxiliary repository, however, is totally a SEAN creation. The standard layout for organizing it is presented in this SOP.

2 Other Structures

The SEAN maintains files in addition to the final products it stewards and disseminates. These additional items are for internal use and are only accessible to SEAN staff: they are generally never passed on to customers. They exist on SEAN's local file server. Included are:

- Inventory area – used only to stage the original inventory of park species. These data do not directly correspond to a monitored vital sign, so they are not stored in the formal vital sign repositories. I&M's separate NPSpecies facility does both repository and dissemination of these data. This area is saved solely to support NPSpecies disaster recovery. It does not house data from inventories beyond species (such as weather and climate).
- Submission safekeeping area – for disaster recovery of SEAN products. Shadow copies of all certified submissions are archived here. In the event of catastrophic failure of any repository, the original data streams may be re-sent from this area in order to accurately restore the SEAN products. Submission files are never disseminated directly; they are only used to recover a repository failure.
- Staging area – for validating and reformatting candidate products from the projects. No products are directly submitted to a repository. They must first be scrutinized for validity. Both tabular and non-tabular objects are kept in staging until properly certified. Only the Data Manager has access to the staging area.
- Google engine – stores configurations and software to maintain the SEAN's search appliance.
- Vital signs heap – data manager's temporary work area for collecting, examining, revising, and reforming miscellaneous vital signs data files and legacy data. Heap objects are always deleted as soon as they have served their purpose. (To be saved, an item must have been identified as a formal deliverable in the protocol and installed in one of the three repositories.)
- Web sites – maintains the source code and other objects for SEAN's development, test, and production web sites. This also stores pages used by I&M and Park web sites as linking pages to the SEAN.
- SEAN Management area – stores correspondence, documentation, and management information of a sensitive nature. Sensitive item access is limited to authorized persons only.

- SEAN Management public area – stores correspondence, documentation, and management information developed by the Network Coordinator.

3 Structure Search Ability

Most of the above structures are indexed and served to web clients by our local web appliance. The search engine can display SEAN's normal web pages. A subset of other file types are also indexed and served to web clients by the appliance, even though they are internally not web page files. The types served are: .PDF, .DOC/.DOCX, .XLS/.XLSX, and .PPT/.PPTX.

The search engine serves all structures **except for** the SEAN management area, the staging area, and the Google engine area itself. In the future, when the search engine becomes accessible outside the firewall, search engine reachable structures must be audited to assure they never contain sensitive information.

4 Vital Signs Identification Prefixes

Some directories exist solely for holding vital signs objects. Such directories are subdivided into individual vital signs areas. The subdirectories are always named with two-character prefixes for the vital signs they contain. The current prefixes are:

Prefix	Vital Sign
AC	Airborne Contaminants
FC	Freshwater Contaminants
FM	Freshwater Macroinvertebrates and Algae
FQ	Freshwater Water Quality
GD	Glacial Dynamics
HU	Human Use / Access
HW	Humpback Whales
IC	Intertidal Communities
IP	Invasive and Exotic Plants
KM	Kittlitz's Murrelets
LL	Landform and Landcover
MC	Marine Contaminants
MP	Marine Predators
OC	Oceanography
SF	Streamflow
US	Underwater Sound
WC	Weather and Climate
WT	Western Toads

An additional subdirectory named 0_SEAN may be present under some folders. This contains dissemination objects specific to the SEAN program that are not vital signs: for example, the PDF of the Data Management Plan or the latest SEAN newsletter.

5 Directory Structure Layout

The root of SEAN's file server directory is <\\inpglbafs03\data> . This is a Windows UNC, not an Internet URL. Within that region, the following structure is to be observed:

- [-] SEAN_Data
 - [-] Auxiliary_Repository
 - [-] 0-SEAN
 - [-] AC
 - [-] FC
 - [-] FM
 - [-] FW
 - [-] GD
 - [-] HU
 - [-] HW
 - [-] IC
 - [-] IP
 - [-] KM
 - [-] LL
 - [-] MP
 - [-] OC
 - [-] SF
 - [-] US
 - [-] WC
 - [-] WT
 - [-] Inventory Area
 - [-] Partner Safekeeping
 - [-] Staging
 - [-] AC
 - [-] FC
 - [-] FM
 - [-] FW
 - [-] GD
 - [-] HU
 - [-] HW
 - [-] IC
 - [-] IP
 - [-] KM
 - [-] LL
 - [-] MP
 - [-] OC
 - [-] SF
 - [-] US
 - [-] WC
 - [-] WT
 - [-] SEAN_Data_Mgmt
 - [-] Google_engine
 - [-] Vital_signs_heap
 - [-] 0-SEAN
 - [-] AC
 - [-] FC
 - [-] FM
 - [-] FW
 - [-] GD
 - [-] HU
 - [-] HW
 - [-] IC
 - [-] IP
 - [-] KM
 - [-] LL
 - [-] MP
 - [-] OC
 - [-] SF
 - [-] US
 - [-] WC
 - [-] WT
 - [-] Web_Sites
 - [-] SEAN_Mgmt
 - [-] SEAN_Mgmt_Public

6 File Naming Standards

SEAN requires that application-specific default file name extensions be employed wherever possible. That is, all Adobe Portable Document Format files should be saved with the extension of '.PDF'. All Access2007 database files should be filed as '.ACCDB', and so on. This allays confusion regarding file content on the part of those needing to access them.

File name forms may be anything supported by the Windows NTFS file system. According to Microsoft Knowledge Base article #100108:

NTFS Naming Conventions

File and directory names can be up to 255 characters long, including any extensions. Names preserve case, but are not case sensitive. NTFS makes no distinction of filenames based on case. Names can contain any characters except for the following:

? " / \ < > * | :

Mixed case is recommended as a way to clarify file content to users. However, keep in mind it is impossible for Windows to store in one folder a file named "AvailableData.PDF" as well as a file named "availabledata.pdf". Also, the longer the name, the less convenient it is to reference; terse, but meaningful, file names should be used.

Auxiliary repository files must be named in a general non-version-specific manner. That is, the file name should not reflect any date updated or revision number. This is because every file in the auxiliary repository is referenced by name in the web site mechanics. Changing a name from, say, WC_RAW_DATA.XLS to WC_RAW_DATA_2.XLS would break links and require costly reprogramming of the web site. Versions of auxiliary repository contents shall be distinguished using each file's NTFS "modification date" entry.

Each file in the auxiliary repository should be stored in its appropriate vital sign subdirectory with a name beginning with the vital sign's prefix in upper case. Examples are "OC_1996Master.xls" and "WC_PDS.pdf".

Files submitted to the staging area for certification are assigned a unique "submission number" when received by the Data Manager. The data manager revises the name of the original submission file for internal purposes using the vital sign prefix and submission number. This readily describes the source and intended destination of each file. Related files generated during certification, as appropriate, may have their purpose appended to this name root. Examples include "WC_1024.MDB", "WC_1024_ErrorLog.doc", and "WC_1024_RejectionMemo.doc". By using this scheme, all files associated with a particular submission are readily identifiable. Also, multiple attempts at certifying a deliverable are stored each under a unique file name for auditing purposes. The submission number is associated with a single file, not just a class of data product. It may never be reused for a subsequent submission.



Southeast Alaska Network
Standard Operating Procedures

301 - Checklist for New Data Management Projects

Abstract

SEAN data management projects are developed and operated through four identified stages. Checklists of required tasks are presented for each stage.

This version saved: September 24, 2008

Table of Contents

1	The Four Stages of SEAN Projects	3
2	Plan-Approve.....	3
3	Design-Develop-Test.....	3
4	Implement-Execute.....	4
5	Evaluate-Close.....	4

Revision history

Revision Date	Base Version	Author	Nature of Change
3/21/2008	-new-	Bill Johnson	Initial version
9/11/2008	3/21/2008	Bill Johnson	Peer review updates

1 The Four Stages of SEAN Projects

As described in Chapter 3 of the DMP, tasks in SEAN projects are categorized as being in one of the four possible project stages: plan-approve, design-develop-test, implement-execute, and evaluate-close. Basic requirements of each stage are detailed here.

2 Plan-Approve

- ☐ Propose project to Network Coordinator
- ☐ Verify there is a legitimate problem
- ☐ Verify there are measurable objectives to meet
- ☐ Acquire approval of concept by technical committee
- ☐ Locate possible partners and cooperators
- ☐ Identify any required permits and determine they are available
- ☐ Identify any staff training needed to develop or operate the project
- ☐ Draft project proposal
- ☐ Have proposal reviewed by regional I&M coordinator
- ☐ Obtain funding and permission to spend the funding
- ☐ Write detailed project management plan
- ☐ Obtain a 2-character vital sign prefix from data manager

3 Design-Develop-Test

- ☐ Complete any staff training needed for development
- ☐ Draft protocol with specific data deliverables and validation criteria listed in detail
- ☐ Identify details of any NPS application interfaces needed
- ☐ Identify any partner repository interfaces needed; negotiate any agreements
- ☐ Identify outside data sources needed to complete products
- ☐ Draft data model supporting deliverables
- ☐ Identify and procure samples of any required specialized equipment
- ☐ Draft any data collection forms – regardless of whether paper or electronic will be used
- ☐ Design, build, and test MS Access databases for project leader to use in collecting and verifying tabular data products
- ☐ Build and test any MS Access queries or reports needed to produce deliverables on a recurring basis
- ☐ Build and test any tools required for importing and verifying external data needed by collection databases
- ☐ Build and test any tools needed to validate candidate deliverables
- ☐ Build and test any tools needed to extract-transform-load data to repositories
- ☐ Draft and revise development web page sets that provide access to deliverables
- ☐ Obtain any permits needed
- ☐ Conduct any pilots
- ☐ Revise and freeze protocol
- ☐ Revise and retest all supporting databases and tools

4 Implement-Execute

- ☐ Create new deliverable tracking sheets
- ☐ Initiate and maintain a “project log” for recording collection events and exceptions
- ☐ Renew any needed permits
- ☐ Complete any staff training needed for project operations
- ☐ Install a production main web page for the project
- ☐ Populate production web pages with basic items like protocol, contacts, FAQs, etc.
- ☐ Obtain, configure and distribute field equipment and supplies
- ☐ Print and distribute sampling forms
- ☐ Distribute any MS Access data collection databases
- ☐ Enter CSDGM metadata into NPS Data Store
- ☐ Obtain any GIS objects needed for final deliverables
- ☐ Collect field data
- ☐ Enter, verify, correct data into collection databases
- ☐ Validate deliverable submissions
- ☐ Correct validation issues
- ☐ Certify data products
- ☐ Redact sensitive data
- ☐ Install work products in appropriate repositories
- ☐ Verify web links give proper access to repository products
- ☐ Update deliverable tracking sheets
- ☐ Produce/validate/store additional deliverables based on latest certified data
- ☐ Prepare specimens and deliver to collections
- ☐ Scan/archive physical forms and field records

5 Evaluate-Close

- ☐ Verify all products are complete and web accessible, and tracking sheet is complete
- ☐ Verify all metadata has been loaded into NPS Data Store
- ☐ Review accuracy and responsiveness of web site and partner repositories
- ☐ Evaluate cooperator, contractor and partner performance
- ☐ Evaluate staff performance
- ☐ Review capacity of local repositories
- ☐ Verify products were/will be collected in annual data archive
- ☐ Identify opportunities to standardize objects like lookup tables and area hierarchies within the Alaska region
- ☐ If ongoing, determine what additional development should be done



Southeast Alaska Network
Standard Operating Procedures

302 - Data Management Considerations in Protocol Development

Abstract

The foundation of effective data management lies in clearly and precisely defining what data products are to be delivered by the project. The distinction is made between information and data, as well as between tabular and non-tabular objects. A collection of detailed planning forms are presented.

This version saved: September 24, 2008

Table of Contents

1	Importance of Addressing Data Management in Protocols.....	3
2	Specifying Deliverable Data Products.....	3
3	Basic Data Management Specifications for Monitoring Protocols	3
4	Data vs. Information / Tabular vs. Non-tabular.....	4
5	Defining Non-tabular Products.....	5
5.1	Non-tabular Data	5
5.2	Non-tabular Information	5
6	Defining Tabular Products.....	5
6.1	Tabular Data	5
6.2	Tabular Information	6
7	Deliverable Work Product Specification Sheets	6
7.1	Form A: Non-Tabular Information Deliverable.....	9
7.2	Form B: Tabular Information Deliverable	10
7.3	Form C: Non-Tabular Data Deliverable.....	11
7.4	Form D: Tabular Data Deliverable.....	12
7.5	Form X: Relation Definition	13
7.6	Form Y: Attribute Definition	14

Revision history

Revision Date	Base Version	Author	Nature of Change
4/11/2008	-new-	Bill Johnson	Initial version
7/24/2008	4/11/2008	Bill Johnson	Peer review revisions

1 Importance of Addressing Data Management in Protocols

To be of use, time-series data require consistency from year to year. To achieve consistency, an exact protocol must be followed. For a protocol to be exact, it must precisely specify what products are to be produced and fully detail the technical definition of each data element. Setting precise technical definitions requires the specialized knowledge of a data management person. Therefore, for time-series data to be useful, a data management specialist is needed at the outset to frame parts of the protocol.

Extensive workforce resources are required to develop dissemination methods, reliable repositories, repository loading methods, comprehensive validation processes, and data acquisition tools. Every time a technical data definition changes, those five items have to be re-engineered. Frequent changes to definitions cause resources to be exhausted before a project can be fully implemented. Therefore, once detailed data descriptions are set and a data management system is built around them, they should not be changed under normal circumstances.

2 Specifying Deliverable Data Products

The purpose of every SEAN monitoring project is to produce sound data and information to be used for future decision making and scientific advancement. To create these products, a data management system must be built that is specific to each particular project. Because the system is engineered to generate specific products, the products should be defined in advance of building the sophisticated data management infrastructure. However, there are times when new valuable products are identified only after a protocol is implemented. In such cases, the protocol and data management tools must be revised, sometimes causing delays and depleting project support resources. It is to a project's advantage to identify as many solid deliverables as possible in the original protocol development process. A further explanation of this concept may be found in the SEAN data management plan, Section 1.8.

The deliverable products must be defined in detail, because the content of the deliverables determines exactly what data need to be collected.

3 Basic Data Management Specifications for Monitoring Protocols

When a new protocol is begun, it will be assigned a "version number". If and when the protocol is revised and republished, the version number must be changed. A suggested numbering scheme is of the form VV-YYYY.n; i.e., the vital sign code, dash, revision year, a dot, and a sequence number beginning with one each time a new year becomes current. An example protocol version is "MC-2009.1", the Marine Contaminant first revision of 2009. A version number is required for subsequent data management processes, so it must be declared in the protocol.

Deliverables that are to be merged into time-series data sets must have a “submission unit” defined in the protocol. This indicates the granularity used for applying partial updates to the time series. For example, a particular deliverable may use CALENDAR_YEAR+PARK_ID as the unit, indicating that all data for a particular deliverable collected in a single year at a single park will be submitted to the data manager in one complete unit. When a deliverable is applied to the time series, it always replaces the previous unit. If a project leader mistakenly certifies a deliverable that has only 25% of the data for a submission unit, then the time series data set will be updated to contain only 25% of the observations. The other 75% will effectively be erased. This is further explained in SOP 601 – Procedures for Certifying Project Data.

4 Data vs. Information / Tabular vs. Non-tabular

For purposes of data management a distinction is drawn between data and information. Data are values of individual points. Discrete sampling observations, for example, result in data. An automated oceanographic probe that measures water salinity every 2 minutes records data.

Information is created by manipulating data using such processes as statistical analysis, synthesis, and modeling. An analysis of surface water temperature and its probable impacts on biota provides information.

Tabular items are capable of being represented in a grid, such as an Excel worksheet or an MS Access table. They are frequently of simple structure but high volume. Tabular data lend themselves particularly well to generation of information, especially large-scale automated generation.

Non-tabular items are represented in forms other than grids of discrete values. They are frequently of complex structure but low volume. Audio recording of whale songs may be non-tabular data that record the presence of particular members of a population in a certain location.

	<i>Tabular</i>	<i>Non-Tabular</i>
<i>Data</i>	Access MDB table of high and low temperatures at specific locations recorded every 10 minutes	Individual orca specimen identification photos as JPEG files
<i>Information</i>	XLS file of total monthly precipitation by location, obtained by summing daily data	MP4 Video file of conference address presenting humpback whale migration routes

Table 302-1. Example of how work products fit in the four-cell classification scheme.

5 Defining Non-tabular Products

5.1 Non-tabular Data

Non-tabular data are required to be in one of the acceptable formats specified in Table 5.3 of the SEAN data management plan. These are all Windows files. If for some reason it is impossible to initially gather data in one of the standard file types, then a protocol process must be specified for the project staff to translate the originally captured data into an acceptable file; non-standard files will not be accepted by the SEAN.

The protocol must define the type of file each product will be delivered in. It should also specify whether data are to be provided as individual files, or as wrapped composites (e.g., JPEG photo sets stored in a single PDF file). Since certain file types are subject to major format revisions, the required software version must be specified for such files. It should be the current version, or no more than one version prior. For example, one may specify the product has to be delivered as a PDF file of version 1.7, the standard file type used by Adobe Reader 8.

5.2 Non-tabular Information

Non-tabular information is plentiful: reports, analyses, photos, PowerPoints, etc. The most important characteristic to define for these is a version of the file format that is not too old for SEAN's archiving/re-versioning processes to maintain. The recommended specification to put in the protocol is the format used by the latest released version of primary software intended to manipulate such files, or one version prior. For example, in year 2008, the default file type used by either Word 2007 (.docx) or Word 2003 (.doc) is acceptable for word processing documents.

6 Defining Tabular Products

6.1 Tabular Data

Tabular data are frequently used for generating other complex deliverables. For example, a straightforward table of rainfall in locations throughout the country over many years could be the basis for an elaborate report on climate trends. Therefore, it is particularly important that tabular data be thoroughly and unambiguously defined; and that they are held to explicitly stated quality standards.

The following parameters for tabular data products must be specified in each protocol:

- Parent relation/table/file name
- Attribute/column/field standard name
- Attribute/column/field default column heading for reports
- One-line description of purpose
- Data type: integer, real, character, date, currency, binary, Boolean
- Units: degrees Celsius/Fahrenheit, grams/kilograms, meters/fathoms, etc.
- Maximum character length of data values

- Mandatory or optional
- Case: upper/lower/sentence/proper /any

Validation criteria should be selected for each attribute. Each individual criterion also needs to be designated as mandatory or optional. An observation that fails a mandatory criterion must either be revised by the project staff to be in compliance or, if that can't be done, be discarded. The following categories of criteria are among those often used:

- Value fits the data type
- Value is within a specific numeric or string range
- Value resolves properly as a foreign key to a lookup table
- Value, or value combined with other values present, is unique to the dataset
- Value is not null
- Value does not exceed maximum attribute size
- Value is one of a particular subset of possible values

The above mostly list individual attribute criteria. To obtain adequate data quality, cross-attribute validations are often necessary. In such cases, several attributes must jointly meet combinations of criteria. For example, if the attribute LENGTH contains a valid value, then LENGTH_TYPE_CODE must not be null. Or, if MONTH is earlier than April, then TEMPERATURE must be below 21C.

6.2 Tabular Information

Tabular information fits in grid form. It is frequently created by straightforward manipulation of tabular data, for example using means such as stratified summation “rollups.” However, complex information may also be represented, as by having grid cells hold multi-value structures.

7 Deliverable Work Product Specification Sheets

During protocol development, it is imperative to precisely define the data deliverables that will be created by the project. Because information products ultimately rely on data products as a basis, the information products need to be defined first so that they may drive specification of data that are ultimately collected. The following forms should be filled out by skilled data designers during protocol development. The forms serve as the foundation for the IT systems that must be built to support the project. Not every product will have every deliverable type. Most projects, however, will have multiple deliverables of a single type. Forms that depend on other deliverables should be filled out first; their requirements will help define the lower level products.

Completing the deliverable specification has to be a joint effort of the project leader and data manager. The project leader provides the content. The data manager provides the concise technical structure definitions. Six forms are employed:

<i>Form</i>	<i>Defines</i>	<i>Primary Person</i>
Form A	Non-Tabular Information	Project Leader
Form B	Tabular Information	Project Leader
Form C	Non-Tabular Data	Project Leader
Form D	Tabular Data	Project Leader
Form X	Relation Definition	Data Manager
Form Y	Attribute Definition	Data Manager

Table 302-2. Purpose of forms and the primary source of form contents.

Forms A, B, C, and D define deliverable products. Form X documents underlying table structure for tabular deliverables (Form B and Form D items). If a deliverable has multiple tables, then multiple Form Xs are needed. Form Y defines the column structure. Every Form X has one or more Form Ys that provide fine detail regarding each individual data element. Figure 302-1 shows a hypothetical arrangement of specification forms that are needed to define the deliverable data products of a vital sign project. Notice deliverables are named using the Vital Sign Identification Prefix defined from SOP 205. Following that are examples illustrating proper use of each form.

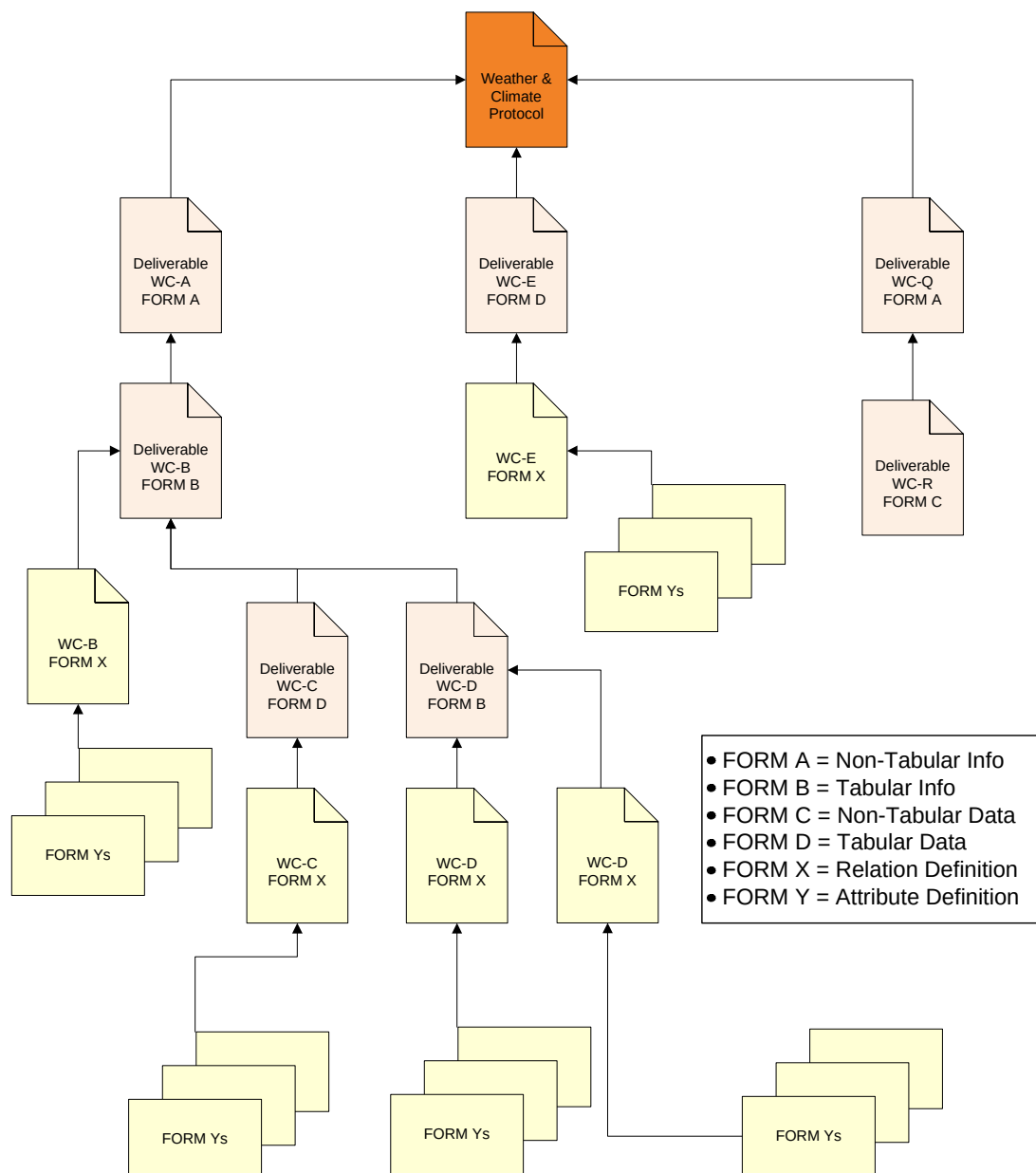


Figure 302-1. Example Deliverable Specification Form Structure for a Vital Sign Protocol. This vital sign has 7 deliverables; which is determined by counting the total occurrences of forms A, B, C, and D. Deliverable WC-A is non-tabular information based solely on WC-B. WC-B is derived from WC-C and WC-D. Deliverable WC-E stands on its own, which is typical of raw data. WC-Q is built on top of WC-R; these could be, for example, an interpretive slide show (Form A) extracted from a library of raw project data images (Form C). Each tabular deliverable needs one or more definitions of relations (Form X) and attributes (Form Y). Orange items may be done by project leader; yellow are best done by data manager.

7.1 Form A: Non-Tabular Information Deliverable

<i>Vital Sign:</i> 1-Weather and Climate	<i>Deliverable ID:</i> WC-A	<i>Deliverable Title:</i> Annual climate analysis
<i>File format:</i> .PPT	<i>Associated software and version:</i> MS PowerPoint 2007	<i>Revision Date / protocol version:</i> 04-09-2008 / 2008.1
<i>Expected frequency:</i> 1/year	<i>Likely dissemination partners</i> WRCC/NOAA	<i>Submission unit:</i> None: does not get merged into a time series
<i>What purpose does this deliverable serve?</i> Explains to the customer summarized historic weather indicators, recent trends, and upcoming climate forecasts.		
<i>Summary of content:</i> This is a PowerPoint presentation aimed at a general audience. It summarizes the weather for a specific year. Analysis of the most recent climate trends is addressed. Brief predictions are made for changes likely in climate variables over the next one and five years. Information presented includes mean weekly high and low temperature Celsius, mean precipitation in centimeters, peak wind direction in degrees, peak wind speed in km/hour, high and low relative humidity in percent. Trends are shown by fitting a line to each of the above attributes over a five year period. Forecasts are general statements of likely changes suggested by projecting the trends forward for two years.		
<i>Mandatory validation criteria:</i> 1. Must be able to be opened with full functionality using Microsoft PowerPoint 2003 or more recent version. 2. All photos and diagrams must be visible in the single deliverable file. No links to outside objects may be included.		
<i>Optional validation criteria:</i> 1. Should be able to be opened with full functionality using Microsoft PowerPoint 2007.		
<i>Deliverable ID of any other SEAN data products prerequisite to this product:</i> VS-1B: Summarized weather table.		
<i>Description and source of any outside data required to create this product:</i> - None -		

7.2 Form B: Tabular Information Deliverable

<i>Vital Sign:</i> 1-Weather and Climate	<i>Deliverable ID:</i> WC-B	<i>Deliverable Title:</i> Summarized weather table
<i>File format:</i> MDB	<i>Associated software and version:</i> MS Access 2003	<i>Revision Date / protocol version:</i> 04-09-2008 / 2008.1
<i>Expected frequency:</i> 1/year	<i>Likely dissemination partners:</i> None – served by SEAN	<i>Submission unit:</i> Year+Station_ID
<i>What purpose does this deliverable serve?</i> Provides detailed weather observations over a calendar year in all three parks, summarized at the weekly level for each station.		
<i>Identifiers of relations that compose the tabular deliverable (“Relations” are tables or files that provide information which may be represented in a grid format. Each relation listed must be fully defined in its own accompanying Form X.):</i> Tbl_weekly_weather		
<i>Mandatory validation criteria for overall deliverable:</i> 1. Must be able to be opened with full functionality using MS Access 2003 or later.		
<i>Optional validation criteria for overall deliverable:</i> - None -		
<i>Deliverable ID of any other SEAN data products prerequisite to this product:</i> WC-C: Weather station calibration details. WC-D: Raw NOAA weather data logger file.		
<i>Description and source of any outside data required to create this product:</i> - None -		

7.3 Form C: Non-Tabular Data Deliverable

<i>Vital Sign:</i> 1-Weather and Climate	<i>Deliverable ID:</i> WC-C	<i>Deliverable Title:</i> Station maintenance records
<i>File format:</i> PDF	<i>Associated software and version:</i> Adobe Reader 8.1	<i>Revision Date / protocol version:</i> 04-09-2008 / 2008.1
<i>Expected frequency:</i> 1/station/year	<i>Likely dissemination partners:</i> None – served by SEAN	<i>Submission unit:</i> Year+Station_ID
<i>What purpose does this deliverable serve?</i> Demonstrates that all sensors have been maintained according to standards and that all calibrations are accurate.		
<i>Summary of content:</i> This is a collection of the equipment service/calibration documentation for each weather station followed by SEAN. The data may be used in analyses that must account for the accuracy of individual station data.		
<i>Mandatory validation criteria:</i> 1. Must be able to be opened with full functionality using Adobe Reader 8.0 or later.		
<i>Optional validation criteria:</i> - None -		
<i>Deliverable ID of any other SEAN data products required to create this product:</i> - None -		
<i>Description and source of any outside data required to create this product:</i> Annual maintenance of all stations is performed by the BLM RAWS Depot in Boise, ID. They supply a written record of work performed, which must be converted to PDF by SEAN.		

7.4 Form D: Tabular Data Deliverable

<i>Vital Sign:</i> 1-Weather and Climate	<i>Deliverable ID:</i> WC-D	<i>Deliverable Title:</i> Raw NOAA weather data logger file
<i>File format:</i> CSV	<i>Associated software and version:</i> n/a	<i>Revision Date / protocol version:</i> 04-09-2008 / 2008.1
<i>Expected frequency:</i> 1/year	<i>Likely dissemination partners:</i> None – served by SEAN	<i>Submission unit:</i> YEAR
<i>What purpose does this deliverable serve?</i> Provides automated weather station observations taken at 10-minute intervals. Defined in: National Wildfire Coordinating Group, 2005. National Fire Danger Rating System Weather Station Standards, PMS 426-3.		
<i>Identifiers of relations that compose the tabular deliverable (“Relations” are tables or files that provide information which may be represented in a grid format. Each relation listed must be fully defined in its own accompanying Form X.):</i> WC-D_yyyy.CSV, where “yyyy” is the calendar year submission unit.		
<i>Deliverable ID of any other SEAN data products required to create this product:</i> - None -		
<i>Description and source of any outside data required to create this product:</i> Raw ASCII data for SEAN stations may be ftp'd from NOAA's NCDC at http://www4.ncdc.noaa.gov/cgi-win/wwcgi.dll?wwDI~SelectStation~USA~AK~DIV~01		

7.5 Form X: Relation Definition

Vital Sign: 1-Weather and Climate	Relation identifier: tbl_weekly_weather	Used by deliverable ID: WC-B
Revision date / protocol version: 04-09-2008 / 2008.1	Type of relation ☒ Database table ☐ Data file	Estimated rows: 156/year
Primary key for this relation: Year+Station_ID+Week_number		
Purpose: Network weather attributes for each station summarized by week.		
Identifiers of attributes defined over this relation (“Attributes” are columns of the grid. Each attribute must be defined in an accompanying Form Y.): Year Station_ID Week_number Temperature_high Temperature_low Precip Wind_direction Wind_speed Humidity_high Humidity_low Protocol_version Update_user Update_timestamp		
Mandatory validation criteria involving multiple attributes: - If Temperature_high is present and Temperature_low is present, then Temperature_high must be $\geq$ Temperature_low. - If Humidity_high is present and Humidity_low is present, then Humidity_high must be $\geq$ Humidity_low.		
Optional validation criteria involving multiple attributes: If Wind_speed is present, then Wind_direction should be present.		

7.6 Form Y: Attribute Definition

<i>Vital Sign:</i> 1-Weather and Climate	<i>Attribute identifier:</i> year	<i>Relation (from Form X):</i> tbl_weekly_weather
<i>Revision date / protocol version:</i> 04-09-2008 / 2008.1	<i>Default report heading:</i> CALENDAR YEAR	
<i>Purpose:</i> Calendar year in which an observation was made.		
<i>Data type:</i>	varchar(4)	
<i>Maximum length</i>	4	
<i>Required:</i>	yes	
<i>Measurement units:</i>	n/a	
<i>Format:</i>	nnnn	
<i>Foreign key to (relation+attribute):</i>	n/a	
<i>Case:</i>	n/a	
<i>Mandatory validation rules for this attribute:</i>	1. Must be 4 digits in the range of 2000 through the current year. 2. Must not be null.	
<i>Optional validation rules for this attribute:</i>	1. Should represent a year no more than 2 years prior to the current year.	



Southeast Alaska Network
Standard Operating Procedures

401 - Operational Roles and Responsibilities

Abstract

The data management plan lists a number of roles and tasks required for developing and operating SEAN. Responsibility for day to day operation is concentrated in the Project Manager and Data Manager. In order to be effective, each person should understand the specific duties required. In order to be accountable, responsibilities should not be shared.

This version saved: September 24, 2008

Table of Contents

1	Operations versus Development.....	3
2	Project Leader Operations Tasks.....	3
3	Network Data Manager Operations Tasks.....	4

Revision history

Revision Date	Base Version	Author	Nature of Change
3/21/2008	-new-	Bill Johnson	Initial version

1 Operations versus Development

The data management plan consistently distinguishes between development phase items and operations phase items. This procedure covers only operations of already developed projects.

Chapter 4 of the DMP details 16 service providers used by the network. Only two of them, Project Leader and Data Manager, have duties that are repeatedly performed in the operations phase on a routine basis. Each one has certain responsibilities which, of course, they also have the authority to perform. In order for a project to operate effectively, staff must be accountable for accomplishing specific tasks. For an individual to be accountable, no responsibility should be shared. Cooperation is essential on many levels in order to complete tasks. But an individual should ultimately be responsible for each one.

This procedure lists the tasks each of those two providers is responsible for. It is general guidance; specific projects may have extraordinary needs that require shifted responsibility to achieve best results. In the case task responsibilities are adjusted, it must be made absolutely clear exactly who is accountable for the job.

2 Project Leader Operations Tasks

- A. Oversee creation of all deliverables, both tabular and non-tabular.
- B. Procure data collection equipment.
- C. Develop and distribute sampling forms.
- D. Train staff.
- E. Implement the standardized local directory structures on acquisition computers.
- F. Initiate and maintain a “project log” for recording collection events and exceptions.
- G. Oversee field logistics.
- H. Manage field data collection, entry, and editing.
- I. Merge external data from other agencies, laboratories, and contractors into collection databases.
- J. Build and test any MS Access queries or reports needed for one-time use.
- K. Develop GIS products with assistance from park and regional experts.
- L. Maintain log of data exceptions and corrective actions.
- M. Supervise data entry and verification.
- N. Submit deliverables to Data Manager for validation.
- O. Correct validation violations in data products.
- P. Enter CSDGM metadata sections 0 and 2 into NPS Data Store.
- Q. Certify valid data products.

- R. Send specimens to collections.
- S. Maintain and archive physical forms and records.
- T. Review deliverable schedule progress with Data Manager.
- U. Maintain backups of all project work files.

3 Network Data Manager Operations Tasks

- A. Lay out and create deliverable tracking sheets for each new collection cycle.
- B. Design, build, test, and deliver basic MS Access databases for project leader to use in collecting and verifying tabular data products.
- C. Build, test, and distribute any basic MS Access queries or reports needed to produce deliverables on a recurring basis.
- D. Build and test any tools required for importing and verifying external data for collection databases.
- E. Build and test any tools needed to validate candidate deliverables.
- F. Build and test any tools needed to extract-transfer-load data to repositories.
- G. Assist Project Leader by explaining technical IT processes that assist acquisition.
- H. Record progress toward completing deliverables on the tracking sheets.
- I. Remind project leader of upcoming deadlines.
- J. Enter CSDGM metadata sections 1, 4, 6, 7, and 8 into NPS Data Store.
- K. Audit metadata contributions of others.
- L. Attempt validation of candidate products supplied by project leader and report results to project leader.
- M. Give advice to project leader regarding resolution of data errors.
- N. Record project leader's certification of validated products.
- O. Segregate any sensitive items, as determined by park superintendent's staff, out of the standard dissemination facility.
- P. Configure any repository containers needed to hold new data.
- Q. Arrange installation of certified products in proper repositories.
- R. Create web site links to delivered products.
- S. Periodically audit the accuracy of repositories.
- T. Monitor the performance and functionality of dissemination web sites.
- U. Create and manage archives.
- V. Perform annual version upgrades to repository files.



Southeast Alaska Network Standard Operating Procedures

501 - Managing Spatial Data Products

Abstract

Geospatial data are integral to many monitoring programs. Managing them requires specialized staff and is often resource-intensive. SEAN accepts national, regional, and park standards and guidance. Web-based geospatial infrastructure at NPS does not appear to be a mature function yet. Since SEAN dissemination is totally web-based, SEAN's role must be limited to certifying spatial data, installing it in repositories, and disseminating it over the web in the form of file transfers.

The datum for all geospatial files submitted to SEAN must be NAD83. That datum must be explicitly declared in the file's metadata. If a file represents a projection, the exact projection specification must also be declared in the metadata.

This version saved: September 24, 2008

Table of Contents

1	Guidance and Standards	3
2	Datum and Projection	3
3	Metadata	3
4	SEAN's Role	3
5	Data Packaging	3
6	Designing Geodatabases.....	4

Revision history

Revision Date	Base Version	Author	Nature of Change
9/18/2008	- New -	Bill Johnson	Initial version

1 Guidance and Standards

Detailed national standards regarding spatial data may be found at:

<http://science.nature.nps.gov/nrgis/standards.aspx>

Of particular relevance is Specifications for Geospatial and Other Data Deliverables of GIS and Resource Mapping, Inventories, and Studies at:

http://science.nature.nps.gov/nrgis/pdfs/NPS_GISSpecs_20070302.pdf

The Alaska region offers tools, datasets, training, and guidance at:

<http://165.83.62.205/rgr/akgis/index.cfm?action=dsp&topic=home&item=home>

2 Datum and Projection

SEAN's required datum for geospatial data is NAD83. The preferred projection for most of Southeast Alaska is UTM Zone 8. UTM Zone 7 is to be used for Dry Bay and westward.

3 Metadata

Certain metadata are essential in order to use geospatial data. ESRI files must always include datum and projection metadata. Additional metadata to meet the FGDC standard, as explained in Section 8.1 of SEAN's Data Management Plan, must also be provided.

4 SEAN's Role

The complexity and significant resource requirements of GIS limit SEAN's ability to directly work with it. Also, SEAN delivers all its products over the web. While facilities are marketed for interactively working with GIS object on the web, these technologies are not yet available to the SEAN. Therefore, SEAN must limit its role in supporting GIS data to only certification, repository management, and dissemination in the form of downloading GIS files from its web pages. SEAN does not now support online manipulation or querying of GIS data. For users to do this, they must retrieve files from SEAN to their local GIS-equipped computers, and work with the data there.

5 Data Packaging

In order to keep up with the ongoing need to upgrade repository files so they work with new software releases, SEAN must limit the number of file types it accepts.

GIS data destined for SEAN are to be saved by originators as File Geodatabases stored in a folder whose name ends in ".GDB". The folder is then compressed using LVZ zip technology and submitted to SEAN. Once certified, such files will be made available for download by web browsers.

File Geodatabases were chosen because:

- They are an allowed type under current NPS national standards
- They support very large data sets
- They are highly recommended by ESRI, the primary GIS software developer
- They are more robust than personal geodatabases
- They are readily transported among various classes of computers and operating systems
- They do not require additional software, such as particular versions of MS-Access

6 Designing Geodatabases

It is beyond the scope of SEAN to advise among the myriad methods and approaches to designing GIS structures to meet particular purposes, given the availability of particular data and tools. But SEAN strongly recommends these structures be spelled out in detail well in advance of releasing the protocol. GIS experts are available to assist with this in a number of places, including the parks, the region, and some cooperators.



Southeast Alaska Network Standard Operating Procedures

502 - Managing Photographic Data Products

Abstract

Photos that serve as data, as opposed to strictly illustrative photos, have special requirements. They must be submitted as either JPEG or TIFF files, depending on whether compression loss is acceptable. Projects must convert any “raw” camera files to one of the two allowed forms before being submitted to SEAN. Hardcopy photos must be scanned to files for submission. JPEG/TIFF files may be packaged in a zip file or be submitted individually.

Eight embedded metadata elements are mandatory for all photo submissions. Metadata are to be stored only within the actual photo files and according to “IPTC Core” schema for XMP. Protocols will explicitly state these requirements.

This version saved: September 24, 2008

Table of Contents

1	Data Photos v. Incidental Photos.....	3
2	Acceptable Form.....	3
3	Photo Packaging and Submission.....	3
4	Metadata Requirements	4
5	Protocol Considerations.....	6

Revision history

Revision Date	Base Version	Author	Nature of Change
9/17/2008	- New -	Bill Johnson	Initial version

1 Data Photos v. Incidental Photos

This procedure applies to photos that serve as data which are identified as specific deliverables in the protocol. These could include collections of photos used, for example, in land-cover classification, glacial dynamics, and individual fauna identification. Photos used to document experiment details may also be treated as data so long as they are intended to quantify, qualify, and/or justify results.

Incidental photos, such as those used to illustrate reports and presentations, are to be directly and wholly embedded in those products and do not require the additional management processes defined here.

2 Acceptable Form

SEAN accepts photos in digital file format only. These are typically recorded by digital camera. If hardcopy is the only medium available for certain photos, they must be scanned into electronic files which are then delivered to SEAN. The SEAN has neither the physical facilities nor the curatorial expertise to adequately manage collections of physical objects like paper photos, negatives, and slides.

In order for it to be practical to refresh file versions in SEAN's database, only a limited number of file types are supported. These are listed in Table 5.3 of the Data Management Plan. Individual photos are to be submitted as either TIFF files, when it is essential to preserve the maximum detail, or JPEG files, when compact size and versatility are important. If cameras are used to record in raw or proprietary formats, they must have their files converted to either TIFF or JPEG before submission to SEAN. It is noted that once a file is converted to JPEG, some loss of information is incurred. It is not possible to perfectly recreate a raw or TIFF high resolution file by transforming its JPEG. By design, JPEG compression is usually not noticeable to human eyes. However, if JPEG photos are subjected to advanced image processing techniques, compression will result in image processing artifacts.

Resolution requirements are to be determined by the protocol developers on a case by case basis. SEAN is not able to set a single standard resolution range that will optimally serve the intended purpose of every data photo series.

3 Photo Packaging and Submission

Photo submission packaging specifications will be written in each protocol. Submissions may be made as a group of individually named and labeled files. If there is a natural directory structure appropriate to a set of photo files, in order to retain that structure they may be packaged as a single zip file for submission to SEAN certification. "Zipping" may also be used on individual

files to save file space, though photo data typically do not shrink significantly. Zip file compression does not cause any loss of photo information. Again, consult the protocol to determine exactly how to package the data photo product.

Regardless of whether or not photo files are packaged in zips, they must be submitted with clear labeling as to what the deliverable ID, protocol version, and submission unit are. The data manager will assign a submission number to uniquely identify them thereafter. This mechanism is detailed in SOP 601 – Procedures for Certifying Project Data.

4 Metadata Requirements

NPS recommended standards for photo metadata may be found at:

http://focus.inside.nps.gov/docs/DigitalPhotoMetadata/ImageMetadataStandard_FINAL-Sept8-2006.pdf

While the recommended standards are comprehensive, only seven items are mandatory for every photo. One other element, indicating the 2-character monitoring project prefix, is additionally required by SEAN. Every electronic photo file must have those eight values embedded following the “IPTC Core” metadata schema for XMP. (XMP is a structure that carries metadata inside both TIFF and JPEG files.) The structure may be populated using a number of tools, such as the “File Info” dialog in Adobe Photoshop or the ThumbsPlus editor. Every protocol that includes a data photo product must require the presence of these eight elements as a mandatory validation criterion.

NPS Metadata Element	Content	IPTC Core XMP Entry
Title	Who or what is in the image	IPTC Status: Title
Image_Content_Place	Where the image was taken	IPTC Image: Location
Image_or_Set_Create_Date	When the image was taken	IPTC Image: Date Created
NPS_Unit_Alpha_Code	Park Code (e.g., KLGO) or Network Code (e.g., SEAN)	IPTC Status: Source
Metadata_Access_Constraints	Who may view the image (e.g., park only, NPS only, public)	IPTC Status: Copyright Notice
Constraints_Information	Explanation of restrictions, if any, on access or use of the image or metadata – or “NA”	IPTC Status: Rights Usage
Contact_Organization	Who to contact for further information	IPTC Status: Provider
Controlled_Terms=>Controlled Term	Vital signs project prefix (e.g., “OC” or “WC”)	IPTC Status: Job Identifier

It is essential these eight elements be embedded within the photo file using XMP. Alternatives, such as external “side-car” files or a database, make the photos worthless in the event the external metadata source is misplaced.

The requirement for embedding the eight mandatory elements within each file is a minimum. Project leaders are encouraged to more fully document their work by supplying additional IPTC Core XMP entries for each file. Also note other metadata are often inserted into files by hardware like cameras and software like ThumbsPlus. An example of embedded XMP metadata entry, using PhotoShop in this case, follows:

The screenshot shows a software interface for managing photo metadata. At the top, a blue header bar displays the filename "100_1736.JPG". Below this, on the left, is a vertical sidebar with a list of menu items: Description, Camera Data 1, Camera Data 2, Categories, History, IPTC Contact, IPTC Content, IPTC Image, IPTC Status (highlighted in blue), Adobe Stock Photos, Origin, and Advanced. The main area of the interface is titled "IPTC Status" and contains the instruction "Use this panel to record workflow and copyright information." Below this instruction are several input fields: "Title" with the value "Location of cast c20051/c02", "Job Identifier" with "OC", "Instructions" (an empty text area), "Provider" with "NPS Southeast Alaska Network", "Source" with "SEAN", "Copyright Notice" with "Public domain", and "Rights Usage Terms" with "NA". Each of these fields has a small upward and downward arrow icon on its right side. At the bottom left of the interface, there is a logo that says "Powered By xmp".

100_1736.JPG

- Description
- Camera Data 1
- Camera Data 2
- Categories
- History
- IPTC Contact
- IPTC Content
- IPTC Image**
- IPTC Status
- Adobe Stock Photos
- Origin
- Advanced

IPTC Image

Use this panel for formal descriptive information about the image.

Date Created ▼

Intellectual Genre

IPTC Scene*

Scene values are defined at <http://www.newscodes.org>

Location

City

State/Province

Country

ISO Country Code ▼

Country codes may be either 2- or 3-letter codes as defined by the ISO 3166 standard.

* Multiple values may be separated by a comma or semicolon.

Powered By **xmp**

5 Protocol Considerations

Protocols must be written to conform to the strictures of this SOP. That is, specifications have to set file format (either TIFF or JPEG), any submission packaging structure, the submission unit, and the embedded eight-element metadata requirement.

Beyond that, considerable latitude has been left to the scientist to determine how best to collect and use data photos to meet their goals. However, once the best parameters have been selected, it is essential they be cast in the protocol and followed consistently in every monitoring cycle.



Southeast Alaska Network Standard Operating Procedures

601 - Procedures for Certifying Project Data

Abstract

Data are not made available to customers until their accuracy and quality are assured. Criteria for validity are always expressed explicitly in the project protocol. The project leader creates tabular and/or non-tabular products, which are considered *non-validated*. Each deliverable must be of a pre-defined scope called the *submission unit*. The project leader submits a candidate product to the data manager for formal validation. Products are cycled through project leader and data manager until all mandatory validation criteria are met, at which point they are called *valid*. The project leader performs any final review and, when satisfied with the quality, notifies the data manager that the product is *certified*. Once certified, the data manager immediately arranges to have the product installed in the appropriate repository and made accessible over the web.

This version saved: September 24, 2008

Table of Contents

1	Network Data Management Plan Reference	3
2	Possible Data States.....	3
3	Creating Non-validated Work Products	3
4	Validating Products	3
5	The Submission Unit	5
6	Certifying Products.....	5
7	Non-tabular Products.....	6
8	Schematic Illustrating Steps to Certification	7

Revision history

Revision Date	Base Version	Author	Nature of Change
4/14/2008	-new-	Bill Johnson	Initial version
9/11/2008	4/14/2008	Bill Johnson	Peer review revisions

1 Network Data Management Plan Reference

The method for validating and certifying work products is explained in Chapter 6 of the SEAN DMP. Validation criteria are addressed in Chapter 7.

2 Possible Data States

At a given time, SEAN data and information products occupy one of three possible states. *Non-validated* products are raw and intermediate components in the custody of the project leader. *Validated* products have been rigorously examined by the data manager against the validation criteria specified in their project's protocol, and have met all mandatory criteria. *Certified* products are validated products that the project leader attests are of adequate quality to be distributed, interpreted, and used as the basis for other analysis and decision making.

3 Creating Non-validated Work Products

Non-validated products are assembled by the project leader. These may be tabular data, photo collections, trend reports, PowerPoints or any other deliverable. The exact products are explicitly defined in each project's protocol. If a product is not defined in the protocol, it is never possible to validate or certify it.

For a monitoring project that spans multiple parks, it is the responsibility of the single project leader to submit from all parks.

The data manager has two isolated roles in the process of building work products. First, the data manager contributes details to the protocol to assure that every required data and information deliverable is fully defined in precise IT terms. Second, for tabular data only, the data manager creates and delivers to the project leader an MS Access database to be used for data collection. All other activity in preparing candidates for validation is the responsibility of the project leader.

4 Validating Products

The project leader submits non-validated deliverables to the data manager for validation. The standard method for this is sending clearly labeled email attachments to the data manager, as illustrated in figure 601-1. The email message must contain the deliverable ID (e.g., AC-F), the protocol version (e.g., 2007.1), and the submission unit (e.g., 2008 KLGO). All three of those attributes are defined in advance in the protocol.

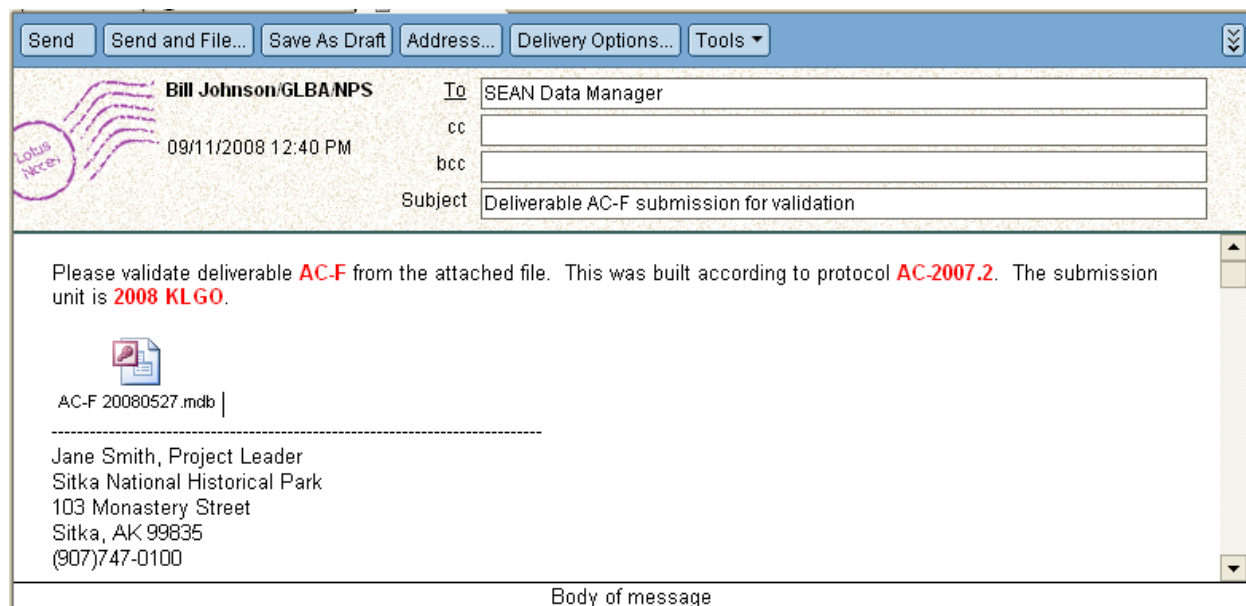


Figure 601-1. Example submission from project leader to data manager.

If technical issues prevent email submission, a mutually agreed upon alternative will be employed. Each submission is given a unique serial number by the Data Manager to identify its reports and audit trails. (This submission number is also used to permanently document the data's provenance in the submission log, discussed in SOP 603.) Validation is a formal process defined in each project protocol. Validation tests the product against explicit criteria. A project's protocol may address two types of validation criteria: mandatory and optional. Mandatory criteria must be met completely for a product to be considered valid. An example mandatory criterion is WATER_TEMPERATURE must be an integer between -5 and 35. Optional criteria are used to identify outliers. An optional criterion might be WATER_TEMPERATURE must be an integer between 0 and 15. A submission is valid if it passes all mandatory criteria, regardless of whether or not optional criteria are satisfied. A required mandatory criterion for every project is that adequate metadata have been provided and recorded in NPS Data Store.

Detailed results of each validation attempt are always reported immediately to the project leader by the data manager in an email, conveying information depicted below:

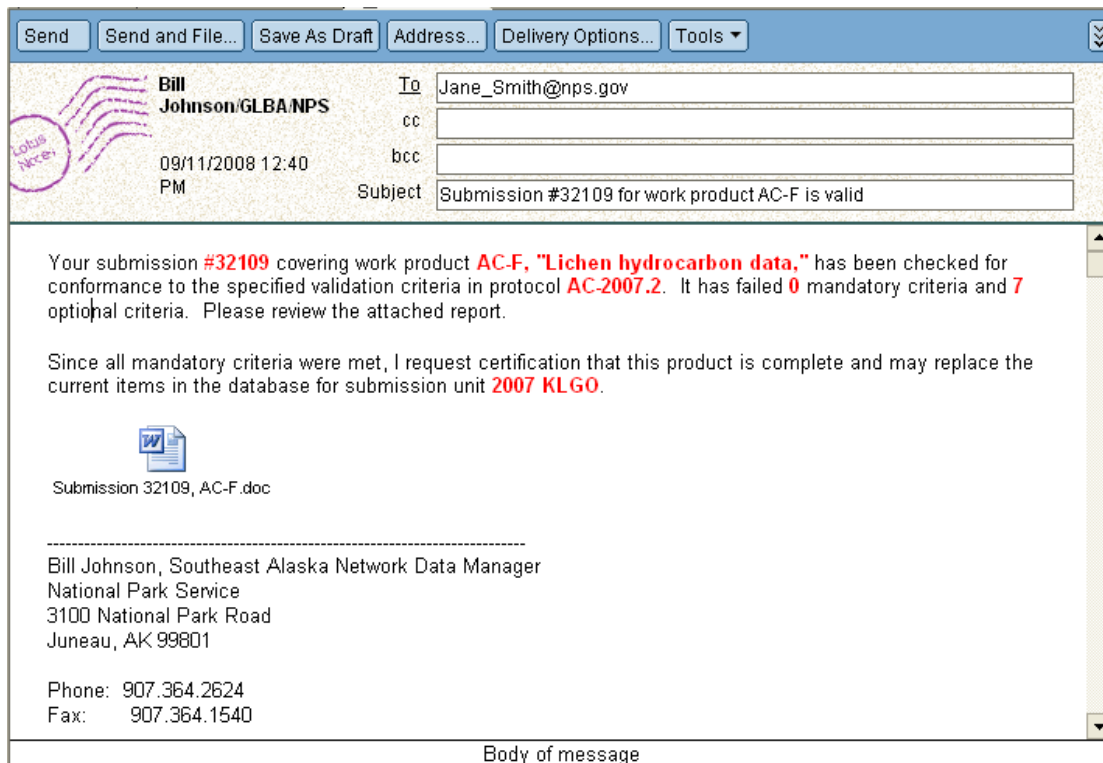


Figure 601-2. Example notification of validation results from data manager to project leader.

5 The Submission Unit

Validation candidates must be of a defined scope, referred to here as a *submission unit*. Since work products for vital signs are often added to a time-series dataset, a method is required to assure items are added only once to the time-series, deleted items are reliably removed from the time-series, and corrections to existing time-series values are applied to the proper items. The submission unit is key to accomplishing this. A submission unit specifies the complete scope of a product. A possible submission unit could be a specific park for a specific year, e.g. GLBA 2011. Submission units must be defined in advance in the protocol, be consistently applied on a specific work product, and use attributes that can never be unknown. The certified data will be used to **replace** all existing data for that submission unit. Under no circumstances are partial submissions permitted. That is, if the submission unit is park and year, then one is not allowed to submit “SITK 2009 spring” for validation and then separately submit “SITK 2009 autumn”. In such a case, after the update process has finished, all spring data would be replaced with autumn, effectively deleting spring.

6 Certifying Products

When a project leader receives a validation report that indicates a submission is valid, the leader must consider whether to certify that submission. First, violation of optional criteria should be

reviewed to see if additional edits are required to make the data of adequate quality for use. Then, the data scope should be reviewed to ensure no observations, or blocks of observations, have been left out. Certifying a submission unit that is missing data will result in any previous record of the omitted data being totally removed from the repository. It is the responsibility of the project leader to ensure every submission is complete; the data manager does not have enough knowledge to determine completeness.

To grant certification, the project leader notifies the data manager by email that the submission is certified. The email must reference the submission's sequence number. It should be of the form illustrated below. The submission sequence number and work product name are the only items that must be originated by the sender. The remaining contents are standard for all certifications.

If a validated submission is not certified by the project leader within a reasonable time, the data manager may unilaterally install the submission in a repository if the data manager judges this to be in the best interest of the NPS.

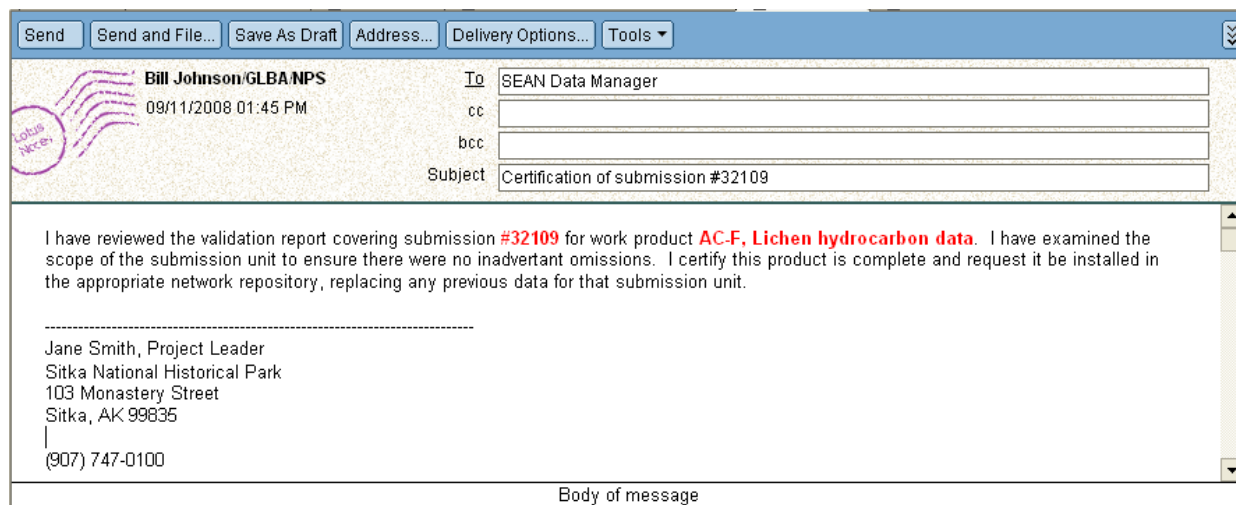


Figure 601-3. Example notification of certification by project leader to data manager.

The data manager will arrange for the submission to be installed in the repository as soon as the certification is received. Certifications will typically be a benchmark to log in the project tracking sheet. When installation is complete, the project leader will be notified in the form of a reply to the certification email.

7 Non-tabular Products

Tabular products lend themselves to automated rigorous validation, and typically have very detailed validation criteria set in the protocols. It is harder to establish complex rules for other data products such as Word documents, MP4 videos, PowerPoints, etc. One mandatory validation that must be specified in the protocol is that such files are capable of being successfully opened by target software, of a version used by SEAN at the time of submission.

Regardless of product format, every item submitted for validation is assigned a sequence number, has QC results reported to the submitter, and needs to be certified by the project leader.

8 Schematic Illustrating Steps to Certification

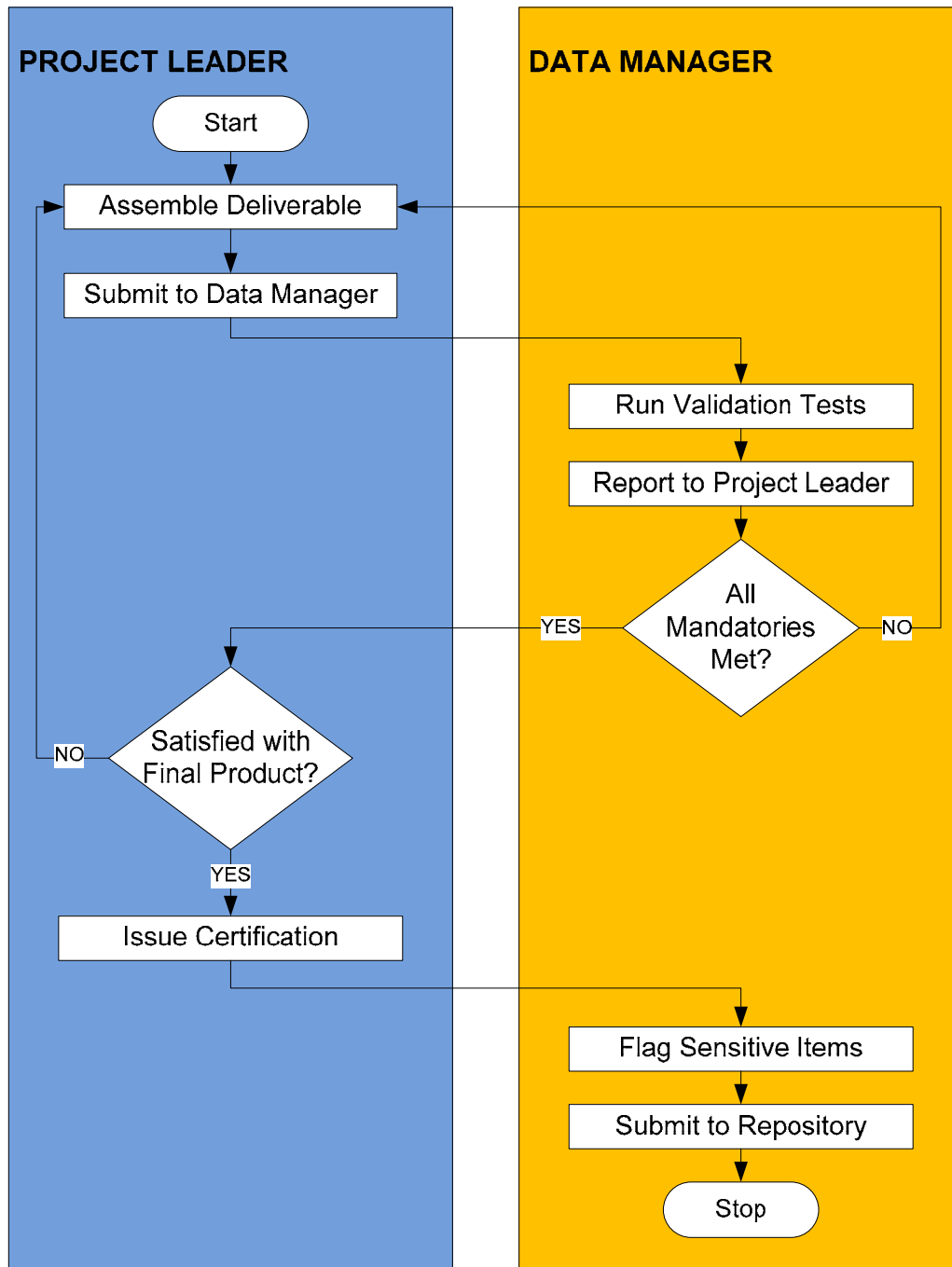


Figure 601-4. Sequence of tasks in certification process.



Southeast Alaska Network
Standard Operating Procedures

602 - Version Control

Abstract

An essential aspect of data management involves assuring data and information are the most current. This requires tracking the version of both data products and the rules used to build them. While specific methods are described in several SOPs, they are recapped here in one location as an aid to implementing them.

This version saved: September 24, 2008

Table of Contents

1	Summary of Version Control Structures Defined by SEAN.....	3
2	Related SOPs	4

Revision history

Revision Date	Base Version	Author	Nature of Change
9/11/2008	-new-	Bill Johnson	Initial version

1 Summary of Version Control Structures Defined by SEAN

Subject	Example	Method
Protocol	OC-2009.2 - Two-character vital sign prefix - Dash - Year accepted - Period - Sequence number within the year	When a protocol for a particular vital sign is formally accepted, it is labeled with the year of acceptance. The first protocol version of a particular year is given sequence number 1.
Deliverable	AC-H - Two-character vital sign prefix - Dash - Designator letter(s)	Each deliverable is formally defined in the protocol. It is designated by the vital sign prefix and a deliverable designator. The first deliverable defined in the protocol is assigned designator A, the second B, and so on. If a protocol revision results in discontinuing a particular product, its designator is not reused. If more than 26 deliverables need to be defined, the designator may be extended with an additional character: X, Y, Z, AA, BB, etc. In the case where a deliverable is “tweaked” but is substantially the same, the deliverable ID remains the same; however, the protocol revision number should be incremented.
Submission files	WC_67823.MDB - Two-character vital sign prefix - Underscore - Submission number - Period - Accurate file extension reflecting type	Submission number is assigned by data manager on receipt of file. File is saved in the staging area using a revised standard name as illustrated in the example. Subsequent log and report files associated with the submission retain the 8-character prefix and append further descriptive information about the type of file it is; e.g., MP_00137_ErrorListing.txt. A submission number is used to version a single submitted file and resulting validation files. It is never reused.
Database rows	2009-05-23 13:07:22 - Timestamp WFJOHNSON - Userid SF-2011.3 - Protocol identifier (when appropriate) 36783 - Submission number (when appropriate)	For tabular data stored in a repository as DBMS objects, whenever a row is inserted it is stamped with two attributes: the current date-time as a DATE object, and the inserting userid as a VARCHAR object. If an inserted row is the result of a certified submission process, then the protocol version and submission number are also recorded, both as type VARCHAR. (Some tables, such as lookup tables, get updated by means other than the submission process.)

2 Related SOPs

Additional aspects of versioning are detailed in:

- SOP 205 – SEAN Internal Directory and File Naming
- SOP 302 – Data Management Considerations in Protocol Development
- SOP 601 – Procedures for Certifying Project Data



Southeast Alaska Network Standard Operating Procedures

603 - Documenting Product Provenance

Abstract

Data and information products reach repositories only after rigorous validation. They may be stored as individual items or may be merged into existing data objects, such as multi-year tables. In order to document the origin of particular items in the repositories, a tracking table records all attempts at submitting new information (both successful and unsuccessful) to the SEAN broad database. The table may be used to audit the origin and content of particular data objects.

This version saved: September 24, 2008

Table of Contents

1	The Need to Document Provenance	3
2	Submission Number as Key	3
3	The Submission Log Table.....	3

Revision history

Revision Date	Base Version	Author	Nature of Change
9/12/2008	- New -	Bill Johnson	Initial version

1 The Need to Document Provenance

Vital sign monitoring data products are intended to consistently cover a long period of time. It is desired the data outlast supporting staff over multiple turnovers. Furthermore, individual deliverables will be merged into existing monitoring sets on a regular basis. This includes adding everything from large PowerPoint files down to individual rows in DBMS tables. In order to be able to audit the origin and basis for particular items in the broadly-defined SEAN database, a tracking system that follows each repository submission must be adhered to.

2 Submission Number as Key

The process for getting a deliverable into a repository is explained in SOP 601 – Procedures for Certifying Project Data. Often there will be an iteration of submission attempts, the final one being certified. For each attempt the data manager assigns a serial *submission number*. The submission number of certified deliverables may have many attributes of interest associated with it. Rather than attempting to include such values as Project Leader, Certification Date, and Protocol Version in every single record of every deliverable, it is much more manageable to just record submission number. This may then be used to look up answers regarding where particular items originated and under what circumstances they were recorded.

3 The Submission Log Table

A table called the *submission log*, having one row for each submission attempt, is maintained by the data manager. Its scope is the entire SEAN database, not individual vital signs or individual deliverables. Currently the table is an Excel spreadsheet. (It is expected to be a DBMS table in the future when such facilities become available.) The submission log table is stored in the root of the auxiliary repository.

SEAN deliverable product submission log															
Submis- sion Number	Protocol ID	Deliverable ID	Submission Unit	Project Leader	Submis- sion Date	Validation Attempt Date	Mandatory Errors	Optional Errors	Certifi- cation Date	Safe- keeping Date	Installer Name	Installed Date	Deleted Rows	Inserted Rows	
00001															
00002															
00003															
00004															
00005															
00006															
00007															
00008															
00009															
00010															

The following attributes are recorded for every submission number:

- Submission number: a unique integer assigned sequentially beginning at one.
- Protocol ID: the formal version identifier for the protocol used to generate the deliverable (see SOP 602)

- Deliverable ID: the formal identifier for the deliverable (see SOP 602)
- Submission unit: a string explaining what set within the time-series is to be replaced
- Project leader: name of the person responsible for the deliverable being submitted
- Submission date: the day the data manager assigns the submission number
- Validation attempt date: the day the data manager ran a validation suite and sent result reports back to the project leader
- Mandatory errors: the number of mandatory validation errors encountered for this attempt
- Optional errors: the number of non-fatal warnings reported

The following attributes are additionally recorded for each submission that successfully passes validation:

- Certification date: the day the project leader notifies the data manager the product is complete and may be installed in a repository
- Safekeeping date: the day the original valid submission file is archived in the safekeeping directory
- Installer name: the person who actually merges the product into the repository
- Installed date: the day the product is verified to be properly installed in the repository
- Deleted rows: in the case of tabular data, the number of rows deleted from the primary table because they were in the scope of the submission unit.
- Inserted rows: in the case of tabular data, the number of rows inserted into the primary table from the submission.



Southeast Alaska Network
Standard Operating Procedures

1101 - Network Archiving Procedures

Abstract

The SEAN archives objects stored in the Auxiliary Repository and the partner safe-keeping area. I&M WASO is responsible for archiving the I&M repository. SEAN exerts no control over partner archiving.

Maintaining archives requires work in two distinct areas: physical archiving and version updating. Physical archiving involves copying repository contents to secure long-life media which have extremely wide acceptance and, hence, longevity. Version updating is the process of identifying files using old proprietary formats, and updating them to newer formats. For example, Excel 97 is an old format that needs to be updated to Excel 2007 soon, before it becomes so old that software can no longer open it.

This version saved: September 24, 2008

Table of Contents

1	Scope of Archiving.....	3
2	Two Parts to Archiving.....	3
3	Physical Archiving	4
4	Version Updating.....	4

Revision history

Revision Date	Base Version	Author	Nature of Change
3/20/2008	-new-	Bill Johnson	Initial version
7/30/2008	3/20/2008	Bill Johnson	Peer review updates

1 Scope of Archiving

SEAN has strong interest in keeping its information products available for the long term, so it must maintain very reliable archives. Archives differ from backups in that they cover long periods and are retained permanently. Also, archives do not provide the finer temporal resolution of backups. SEAN archives cover the Auxiliary repository and the safe-keeping area that holds submissions made to partners and retained for business continuity purposes. SEAN does not archive the I&M repository, which is the responsibility of WASO. SEAN does not archive draft, incomplete or non-validated project work products. That is in the realm of backups, and is the responsibility of project leaders. It is inherent that SEAN archives continually grow in size.

2 Two Parts to Archiving

Archiving procedures address two different aspects of retention: physical archiving and version updating. Physical archiving is copying repository files to high-quality media of a type having very high market acceptance. There is safety in numbers. The more popular a technology, the longer it will be kept available and supported. SEAN employs single-layer DVD+R technology using the longest-life media available, which typically has a gold reflective layer that is particularly resistant to oxidation.

Version updating procedures involve determining files that are using older than current proprietary file formats, and converting them to new files in the latest formats. In order to make this manageable, SEAN explicitly restricts work products to fourteen possible formats:

File Extension	Contents Supported
.DOC/.DOCX	Microsoft Word document
.PDF	Adobe Portable Document Format document
.MDB/.ACCDB	Microsoft Access database
.XLS/.XLSX	Microsoft Excel worksheet
.MP3	Audio
.MP4 ¹	Video
.JPG	Image
.GIF	Image
.TIF ²	Lossless image
.PPT/.PPTX	Microsoft PowerPoint presentation
.CSV	Comma-separated ASCII text data
.XML	Extensible markup language data
.VSD	Microsoft Visio drawing
.ZIP	Zipped folder (e.g., ESRI file geodatabase objects)

¹ MP4 refers to the MPEG-4/H.264 group of specifications. MP4 content may also be “wrapped” in AVI or WMV containers.

² TIFF version 6.0 or greater should be used; earlier versions may face software compatibility issues.

3 Physical Archiving

The SEAN archive is created once a year during the last week of December. Two complete copies of the repositories are made. One is permanently retained in the fireproof cabinet at the SEAN offices. The second is mailed to the GLBA IT staff, who keep a duplicate archive. Each year a complete, cumulative pair of DVD+R sets is made and retained permanently. SEAN archive DVDs are never discarded under any circumstances.

The DVDs are burned using standard Windows software on the INPGLBAFS03 server. A DVD+R burner has been installed in the server box for this explicit purpose. They are simply direct file copies. No compression or reorganization is done, since that could impede archive recovery in later decades. All files located at [\\INPGLBAFS03\DATA\SEAN_DATA](#) are copied. Since each DVD+R has a capacity of 4.7GB, multiple DVDs will be required. The person responsible for generating the two archive sets will ensure the file distribution across multiple disks is logical and intuitive to future users.

Once the disks are created, they must be verified. This is done by inserting each disk in a DVD reader on a machine other than the server used to create it and copying them all to scratch space on a hard drive. If the DVD reads completely and without error, it is considered viable. It should be labeled with date and sequence number using a permanent marker having low volatile organic content; some solvents can damage media.

4 Version Updating

Version updating procedures should be performed immediately before the annual permanent archive is completed. This function should only be performed on files that are identified as having dangerously-old formats. It is prohibitively expensive to prove software works exactly according to design, so there is always a possibility that by using version converting software errors get introduced into the files. The slight likelihood is definitely overwhelmed, however, by the value of keeping data in a form that can be used by current-issue software. When a format is two versions past the current level, it is essential to update the file.

Be aware that version updating is required only by older files. In effect, such files are guaranteed to be archived since two formats never are superseded within one year. Knowing there is an original permanently available for restoring means that updating files is protected by a good safety net. The conversions need not be weighed down with extraordinarily careful and complex procedures.

Certain file formats are not so proprietary as to be locked into limited software tools. These include MP3, MP4, JPG, GIF, TIF, CSV, and XML. These formats do not need to be annually screened for refresh. Because they are used in so many ways by so many products, the formats are cast in concrete. The only time a conversion is needed is when the particular format has been superseded by a completely new technology. On the rare instances this occurs, the DMP has to

be revised to redefine the acceptable file types SEAN will accept. Once a file type has been replaced with something new, all the old data must be converted to the new format. This typically requires a special project.

Other file formats have their internal definitions altered more frequently. Among these are DOC, PDF, MDB, XLS, PPT, VSD, and the ESRI GIS files. An annual review of version is needed for these. The method involves opening every file in the repository using its associated software and, if the software alerts that the file is of an obsolete version, resaving the file in the current version format. If a file is resaved with a slightly different name, the original obsolete file left remaining must be deleted immediately. Such files violate the authoritative source principle and are totally unacceptable.

Those files that are updated must be noted. After all files are updated, the noted files are opened, one at a time, to demonstrate they are not defective. This step is essential. If an error is not noticed until years later and the original format has become totally unsupported, then restoration would not be possible.

The work is tedious. This is mitigated by the procedure only having to be performed once each year. Automation is possible, but not cost-justifiable until the number of files in SEAN's Auxiliary repository is considerably larger.



Southeast Alaska Network
Standard Operating Procedures

1201 - Project Tracking

Abstract

Each specific project operational task must be clearly defined at the outset as the project management responsibility of either the project leader or the data manager. The SEAN data management staff will only manage tasks for which it has authority and direct responsibility.

The SEAN uses a standard Excel spreadsheet for project tracking. Data deliverables, as defined in the protocol, always have their completion recorded when each product emerges from the Operations phase. Tasks associated with the Development phase of a Vital Sign project may also be tracked with the same spreadsheet tool, though these will only occur in the early stages of a monitoring project.

This version saved: September 30, 2008

Table of Contents

1	Scope of Project Management Responsibility within the Realm of Data Management	3
2	Example Tracking Sheet.....	4
3	Identification of Deliverable Products.....	5
4	Identification of Deliverable Milestones	5
5	Recording Progress.....	5
6	Specific Steps to Take to Implement Tracking	5

Revision history

Revision Date	Base Version	Author	Nature of Change
3/20/2008	-new-	Bill Johnson	Initial version
9/30/2008	3/20/2008	Bill Johnson	Peer review updates

1 Scope of Project Management Responsibility within the Realm of Data Management

Under the SEAN system, the Project Leader and Data Manager are assigned domains of different and non-overlapping tasks.

Responsibility for task completion can only be taken by persons having the authority to commit to action. Furthermore, in order to maintain accountability, ultimate responsibility for a particular task must rest with a single person. The SEAN Data Manager typically does not serve as an overall program manager, but is instead responsible for managing the Certification, Repository, and Dissemination functions. The Data Manager does complete several tasks related to Data Acquisition, primarily during protocol development:

1. Assist in identifying and technically defining deliverable work products.
2. Assist in defining mandatory and optional validation criteria.
3. Assist in defining submission units.
4. Design and build basic MS Access databases for project leader to use in collecting and verifying tabular data products.

The remaining Data Acquisition tasks are the responsibility of, and are managed by, the Project Leader in charge. It is important the Project Leader and Data Manager understand what they are each responsible for.

SOP 401 lists the normal delegation of tasks in detail.

2 Example Tracking Sheet

The SEAN uses Excel spreadsheets, laid out as follows, for tracking monitoring progress.

A	B	C	D	E	F	G	H	I	J
Vital sign 1: Weather and climate									
1	Year: 2008								
2	Deliverable								
3	Milestone	Responsible Person	Standard data log feed from all stations	Periodic sampling data taken at designated critical locations	Mid-season sampling progress report	Photo set documenting weather station placement	End-season sampling progress report	Trend analysis of precipitation and temperature by site	GIS theme: annual high and low temperatures by site
4	Data model complete	SEAN data manager	na	1/15/2008	na	na	na	na	na
5	Protocol frozen	SEAN network coordinator	2/12/2008	2/12/2008	2/12/2008	2/12/2008	2/12/2008	2/12/2008	2/12/2008
6	Sampling database designed	SEAN data manager	na	3/21/2008	na	na	na	na	na
7	Sampling forms created	Project Leader	na	4/12/2008	na	na	na	na	na
8	Sampling database initially created	SEAN data manager	na	4/22/2008	na	na	na	na	na
9	Sampling database testing completed	SEAN data manager	na	4/22/2008	na	na	na	na	na
10	Sampling database delivered to project leader	SEAN data manager	na	4/23/2008	na	na	na	na	na
11	Acquisition commenced	Project Leader	3/15/2008	5/1/2008	na	4/23/2008	na	na	na
12	Data entry / edit begun	Project Leader	na	na	na	na	na	na	na
13	Verification completed	Project Leader	11/20/2008	na	na	na	na	na	na
14	Deliverable sent to SEAN	Project Leader	12/1/2008	7/15/2008	7/15/2008	5/1/2008	na	na	na
15	Deliverable validated successfully	SEAN data manager	12/12/2008	na	7/15/2008	5/1/2008	na	na	na
16	Deliverable certified	Project Leader	12/12/2008	7/17/2008	7/17/2008	5/12/2008	na	na	na
17	Deliverable packaged	SEAN data manager	12/19/2008	7/19/2008	7/19/2008	5/12/2008	na	na	na
18	Deliverable accepted by repository	SEAN data manager	1/25/2009	7/19/2008	7/19/2008	5/12/2008	na	na	na
19	Deliverable posted to repository	Repository staff	3/15/2009	7/20/2008	7/20/2008	na	na	na	na
20	Web site linked to deliverable	SEAN data manager	3/15/2009	7/21/2008	7/21/2008	na	na	na	na
21	Web access tested	SEAN data manager	3/15/2009	7/22/2008	7/22/2008	na	na	na	na
22	Sampling forms archived	Project Leader	na	na	na	na	na	na	na
23	Specimens sent to curator	Project Leader	na	na	na	na	na	na	na
24	Specimen access on recorded	SEAN data manager	na	na	na	na	na	na	na

3 Identification of Deliverable Products

SEAN tracks its progress in delivering certified work products to customers. Specific work products (that are subject to tracking) are identified by completed protocols, so SEAN's data deliverable tracking is activated only after the protocol is frozen, during the real Operations phase.

Once the protocol has been approved, a list of deliverable data products is available. These are recorded as column headings in an Excel spreadsheet which is updated only by the Data Manager. The sheet is accessible at all times through a web site link on the project's main page and is stored in the Auxiliary repository under the corresponding project directory.

4 Identification of Deliverable Milestones

Each row in the tracking sheet is assigned a milestone name. This results in a grid of milestones (rows) by deliverables (columns). The milestones are extracted from the protocol by the Data Manager and Network Coordinator. Milestones are listed in expected operational order, and grouped by the four core functions.

There are always cases where a grid cell is not applicable: some milestones have nothing to do with certain deliverables. For example, a milestone for completing design of a collection form is not applicable to the deliverable of a set of photographs documenting site selection. When the grid of milestones by deliverable is initially composed, the data manager marks inapplicable cells with an 'NA' designation.

5 Recording Progress

When a project's main page is created on the web site, the project tracking sheet should be immediately defined and linked. The tracking sheet identifies for customers the products that will be delivered, and shows the latest progress toward completion. It is the responsibility of the Data Manager to maintain effective communications and keep the tracking sheets current. Because monitoring projects run long-term, a separate tracking page is maintained for each collection cycle.

6 Specific Steps to Take to Implement Tracking

- Wait for protocol to be frozen
- Copy a blank spreadsheet template to the vital sign's area of the auxiliary repository
- Establish one spreadsheet tab for the first collection year
- Define column headings using the titles of the explicit deliverables identified in the protocol

- Define row headings with milestones relevant to the deliverables, in expected chronological order grouped by Core Function
- Assign position responsible for each milestone – only one position should be defined as the responsible person for a milestone
- Assign ‘NA’ values to those cells for which a milestone is Not Applicable for a particular deliverable
- Create a web link to the tracking sheet from the vital sign’s main page
- Whenever a milestone is reached for a deliverable, insert its date into the appropriate cell
- For subsequent cycles, copy the prior sheet into a new tab and repopulate it as the latest monitoring effort proceeds.

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.

National Park Service
U.S. Department of the Interior



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

www.nature.nps.gov

EXPERIENCE YOUR AMERICA™