Data Management Plan
for the
Pacific Island Network

Prepared by
Gordon H. Dicus (NPS)

Pacific Island Network (PACN)

Territory of Guam
War in the Pacific National Historical Park (WAPA)

Commonwealth of the Northern Mariana Islands
American Memorial Park, Saipan (AMME)

Territory of American Samoa
National Park of American Samoa (NPSA)

State of Hawaii
USS Arizona Memorial, Oahu (USAR)
Kalaupapa National Historical Park, Molokai (KALA)
Haleakala National Park, Maui (HALE)
Ala Kahakai National Historic Trail, Hawaii (ALKA)
Puukohola Heiau National Historic Site, Hawaii (PUHE)
Kaloko-Honokohau National Historical Park, Hawaii (KAHO)
Puuhonua o Honaunau National Historical Park, Hawaii (PUHO)
Hawaii Volcanoes National Park, Hawaii (HAVO)

File Name: data-management-plan_pacn_20060811

Recommended Citation:
Dicus, G. H. 2006. Data Management Plan for the Pacific Island Network. National Park Service, Pacific Island Network, Hawaii National Park, HI.

Last revision: 10 August 2006

Status: DRAFT

Organization Contact Information:
National Park Service, Inventory and Monitoring Program, Pacific Island Network, PO Box 52, Hawaii National Park, HI 96718, phone: 808-985-6180, fax: 808-985-6111, http://www.nature.nps.gov/im/units/pacn/index.htm

Acronyms:

NPS	National Park Service
PACN	Pacific Island Network
I&M	Inventory & Monitoring Program
FGDC	Federal Geographic Data Committee
NBII	National Biological Information Infrastructure
USGS	United States Geological Survey
PIERC	Pacific Island Ecosystems Research Center
ALKA	Ala Kahakai National Historic Trail
AMME	American Memorial Park
HALE	Haleakala National Park
HAVO	Hawaii Volcanoes National Park
KAHO	Kaloko-Honokohau National Historical Park
KALA	Kalaupapa National Historical Park
NPSA	National Park of American Samoa
PUHE	Puukohola Heiau National Historical Park
PUHO	Puuhonua o Honaunau National Historical Park
USAR	USS Arizona Memorial
WAPA	War In The Pacific National Historical Park

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

EXECUTIVE SUMMARY

Information is the common currency among the activities and staff involved in the stewardship of natural resources for the National Park Service (NPS). This Executive Summary chapter summarizes the data management strategy for the Pacific Island Network (PACN) of the Inventory and Monitoring (I&M) Program. The Data Management Plan is a guide for current and future project leaders and PACN staff to ensure the continuity and documentation of data management methods and procedures over time. The Data Management Plan, in turn, refers to other guidance documents and standard operating procedures which convey the specific standards and steps for achieving the network's data management goals. The Data Management Plan is a companion document to the PACN Monitoring Plan (see References Cited section).

The Data Management Plan focuses on the processes used to:

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

Data Management Goals

The overall goal of the NPS I&M Program is to provide scientifically and statistically sound data to support management decisions for the protection of park resources. The Program's data management mission is to ensure the quality, interpretability, security, longevity and availability of program natural resource data.

The data management goals of the PACN are:

- *Goal 1* - Ensure the high quality and long-term availability of the ecological data and related analyses produced from the network's inventory and monitoring work.
- *Goal 2* - Integrate data management activities with all aspects and at all stages of network business.
- *Goal 3* - Specify data stewardship responsibilities for all network personnel.
- *Goal 4* – Work within the network and outside the network, as appropriate, to improve the quality and availability of legacy NPS datasets and data from outside sources.

Data Management Priorities

The priorities for network data management efforts are:

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

Data Stewardship Roles and Responsibilities

Every individual involved in the I&M Program is required to understand and perform data stewardship responsibilities in the production, analysis, management, and end use of the data as

described in the Data Management Plan and the specific vital sign monitoring protocols. PACN vital sign monitoring protocols will describe specific roles and responsibilities in detail. Senior network staff (described in Chapter 8 of the PACN Monitoring Plan) share responsibility in ensuring that data management procedures are followed.

Data and Information Workflow

Understanding the life cycle of data throughout a project will help to manage the staffing resources necessary to complete and support quality data. PACN projects include short-term data collection, analysis, and reporting efforts, such as inventories, and long-term efforts such as vital sign monitoring, as well as efforts external to the I&M Program that generate data of interest to PACN. For data management to be effective, it must occur throughout all stages of the project.

A project is divided into the following stages:

1. Planning and Approval
2. Design and Testing
3. Implementation
4. Product Integration
5. Evaluation and Closure

PACN uses a project tracking database (see Appendix B) to document and support the progress of information collected for PACN projects. Most notably, this database tracks project status, changes to protocols, and archiving and distribution of deliverables.

Infrastructure and System Architecture

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

The infrastructure supports these required functions:

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

Database Design Strategies

For PACN inventory projects and vital sign monitoring projects, the project leader and the data manager will work together to develop conceptual and logical data models to:

1. Understand the data life cycle flow of the data collection process; identify the starting point of data collection (e.g., a visit to a site) and the steps involved in data processing.
2. Determine the data relationships for database development (e.g., one site visited on multiple dates with numerous data elements measured on each visit).
3. Determine how the information will be organized for efficient retrieval and presentation.

The Data Management Plan specifies the standards by which data will be handled. Data management elements or principles common to more than one vital sign protocol will, to the greatest extent practical, be standardized so as to enhance overall data integrity and the comparability of data across the network.

Acquiring and Processing Data

The types of data handled by the I&M Program fall into three general categories:

- Program data – produced by projects that are either initiated (funded) by the I&M Program or involve the I&M Program in another manner (e.g., natural resource inventories and vital sign monitoring projects)
- Non-program legacy/existing data – produced by NPS entities without the involvement of the I&M Program (e.g., park or regional projects)
- Non-program external data – produced by agencies or institutions other than the National Park Service (e.g., weather and water quality data)

Most data acquired by the network will be collected as field data (inventories and long-term monitoring) or discovered through data mining initiatives (legacy/existing data). Methods of field data collection, such as paper field data forms, field computers, automated data loggers, and GPS units, will be specified in individual monitoring protocols and study plans. Field crew members must closely follow the established standard operating procedures (SOPs) in the project protocol. Data acquired by non-program sources, such as data downloaded from other agencies, will also be specified in individual monitoring protocols.

Ensuring Data Quality

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

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

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

Quality assurance methods must be in place at the inception of any project and continue through all project stages to final archiving of the dataset. It is critical that each member of the team work to ensure data quality. The final step in project quality assurance is the preparation of summary documentation that assesses the overall data quality. A statement of data quality will be composed by the project leader and incorporated into formal metadata. Metadata for each dataset

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

Data Documentation

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

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

Data documentation will be available and searchable in conjunction with related data and reports via the PACN website as well as the national I&M Program's NR-GIS Metadata and Data Store, a searchable online application for managing and sharing natural resource and GIS metadata and data generated by the National Park Service.

Data Analysis and Reporting

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

Data Dissemination

The PACN data dissemination strategy aims to ensure that:

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

Access to PACN data products will be facilitated by a variety of means that allow users to browse, search and acquire network data and supporting documents. These means include, but are not limited to:

- Links to public data products will be maintained on the PACN public website
- NR-GIS Metadata and Data Store, an online application for managing and sharing natural resource and GIS metadata and data (distribution instructions for each dataset will be provided in the respective metadata)
- Service-wide databases, such as NPSTORET for water quality data, NPSpecies for species biodiversity data, and NatureBib for bibliographic data
- Regional, Network, or Park data servers for providing datasets in a read-only format

- External repositories such as the University of Hawaii, US Geological Survey, US Forest Service, Bishop Museum, Western Regional Climatic Center, and many others
- FTP sites, CDs, DVDs, or hard drives, as appropriate

Ownership, FOIA, and Sensitive Data

PACN products are considered property of the NPS. However the Freedom of Information Act (FOIA) establishes access by any person to federal agency records that are not protected from disclosure by exemption or by special law enforcement record exclusions. The NPS is directed to protect information about the nature and location of sensitive park resources under one Executive Order and four resource confidentiality laws:

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

When any of these regulations are applicable, public access to data can be restricted. If disclosure could result in harm to natural resources, the records may be classified as 'protected' or 'sensitive'. The NPS recognizes the following resources as sensitive:

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

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

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

Classification of sensitive data will be the responsibility of network staff, park superintendents, and project leaders. Network staff will classify sensitive data on a case-by-case, project-by-project basis and will work closely with project leaders and park staff to ensure that potentially sensitive park resources are identified, that information about these resources is tracked throughout the project, and that potentially sensitive information is removed from documents and products that will be released outside the network.

Digital Data Maintenance, Storage, and Archiving

PACN data maintenance, storage and archiving procedures aim to ensure that digital data and related metadata documentation are:

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

Primary data maintenance occurs on the PACN file server and on service-wide servers maintained by NPS staff and cooperators at the Washington Area Support Office in Fort Collins, Colorado. PACN staff are responsible for keeping data and information current on PACN and service-wide servers, and depend on national and regional IT staff for assistance with regular data backups. PACN staff will ensure that the latest versions of primary data are available in conventional formats reflecting common data usages in the resource management community.

Project data are electronically archived as stand-alone products that include:

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

Non-Digital Data Archiving and Records Management

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

The PACN data management strategy includes assisting project leaders in complying with archival directives. Whenever necessary, physical items considered project products such as reports, maps, photographs, or notebooks will be cataloged and archived by the park(s) involved with the project. When this is not possible, an alternative storage strategy and location will be found and fully described in the project documentation. Physical specimens, such as plants and animals, will be accessioned and housed at the appropriate archival institution (typically a park archival facility, but may be a partner institution such as the Bishop Museum).

Water Quality Data

Water quality data are managed according to guidelines from the NPS Water Resources Division (WRD). This includes using the NPSTORET database application to help manage data entry, documentation and transfer. The PACN oversees the use of NPSTORET according to the network's integrated and regulatory water quality monitoring protocols and ensures the content is transferred at least annually to WRD for upload to the STORET database.

Implementation

The Data Management Plan (DMP) contains practices that may be new to staff and principal investigators. With a few exceptions, however, the DMP does not include any requirements that are new. Almost every requirement comes from law, Director's Orders, or the I&M Program. The DMP helps to put these requirements into context and in sequence, provides operational guidance for achieving these requirements, and outlines short- and long-term goals. In order to remain current with Information Technology advances and with PACN needs, the DMP will be reviewed and revised on a five year cycle.

1. INTRODUCTION

1.1. The Inventory and Monitoring Program

The Inventory and Monitoring (I&M) Program represents a long-term commitment by the National Park Service (NPS) to assess and document the status and trends of park resources. In 1998, the National Parks Omnibus Management Act established a framework for the I&M Program, which integrates natural resource monitoring and other scientific activities into the management processes of the National Park system. As part of the strategy to institutionalize inventory and monitoring on a programmatic basis throughout the agency, all national park units with significant natural resources have been grouped into 32 networks linked by similarities in geography, natural resources, and resource protection challenges. The I&M network approach facilitates collaboration, information sharing, and economies of scale in natural resource monitoring. The approach also provides parks with a "minimum infrastructure" to initiate natural resource monitoring.

The Inventory and Monitoring Program's long-term goals are to:

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

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

1.2. Scope of the Data Management Plan

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

The DMP focuses on the processes used to:

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

The I&M Program anticipates 30% of the program's efforts will involve data management. Providing the overarching guidance for this effort, the DMP aims to ensure that program data are documented, secure, and remain accessible and useful for future generations. The DMP, in turn, refers to other guidance documents and standard operating procedures which convey the specific standards and steps for achieving the network's data management goals. The DMP is a companion document to the Pacific Island Network Monitoring Plan, and is intended for use by the parks or other programs.

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

1.3. Data Management Goals

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

The data management goals and associated objectives of the PACN are as follows:

- *Goal 1* - Ensure the high quality and long-term availability of the ecological data and related analyses produced from the network's inventory and monitoring work.
  - Objective - Outline the procedures and work practices that support effective data management.
  - Objective - Establish an organizational schema for PACN program data and information so that they are retrievable by staff, cooperators, and the public.
  - Objective - Establish standards for data, data distribution, and data archiving to ensure the long-term integrity of data, associated metadata, and any supporting information.

- *Goal 2* - Integrate data management activities with all aspects and at all stages of network business.
  - Objective - Encourage effective data management practices as an integral part of project management so all data are available and usable for park management decisions, research, and education, now and into the future.
  - Objective - Establish quality control and quality assurance standards.

- *Goal 3* - Specify data stewardship responsibilities for all personnel.
  - Objective - Guide current and future staff of the PACN to ensure that sound data management practices are followed.
  - Objective - Establish data management roles and responsibilities of PACN staff.

- *Goal 4* - Work within the network and outside the network, as appropriate, to improve the quality and availability of legacy NPS datasets and data from outside sources
  - Objective – As time and resources permit, migrate high-priority legacy datasets into modern formats and improve the quality and documentation of these datasets or other data originating from outside the Inventory and Monitoring Program.
  - Objective – Work with partner agencies and institutions to promote the sharing and development of data, software applications, and analyses.

1.4. Audience

The intended audiences of the Data Management Plan are:

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

1.5. Revisions to the Data Management Plan

The Data Management Plan is scheduled for a full review and revision every five years, coinciding with the PACN Inventory and Monitoring Program review schedule. Until then, minor revisions may be made as needed and will be documented in the corresponding chapters.

1.6. Types of Data Covered by this Plan

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

This plan encompasses a range of products that are coordinated or managed by PACN. These products fall into four general categories: data, documentation, reports, and administrative records (Table 1.1). Documentation, in the form of protocols, data dictionaries, database user guides, and standard operating procedures, provides the long-term value of data by setting the context of how and why the data were collected, analyzed, and reported.

Table 1.1. *Categories and examples of data products covered by the Data Management Plan.*

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

1.7. Sources and Priorities of Natural Resource Data

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

- I&M Inventories
- I&M Vital Signs Monitoring
- Natural resource projects, monitoring, and research done by park staff and cooperators

- Monitoring or research done by other agencies or on adjacent lands

The priorities for network data management efforts are:

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

1.8. Pacific Island Network Organization and Structure

The Pacific Island Network consists of 11 park units located in Hawaii, the west Pacific (Guam and Saipan), and the south Pacific (American Samoa) (Figure 1.1).

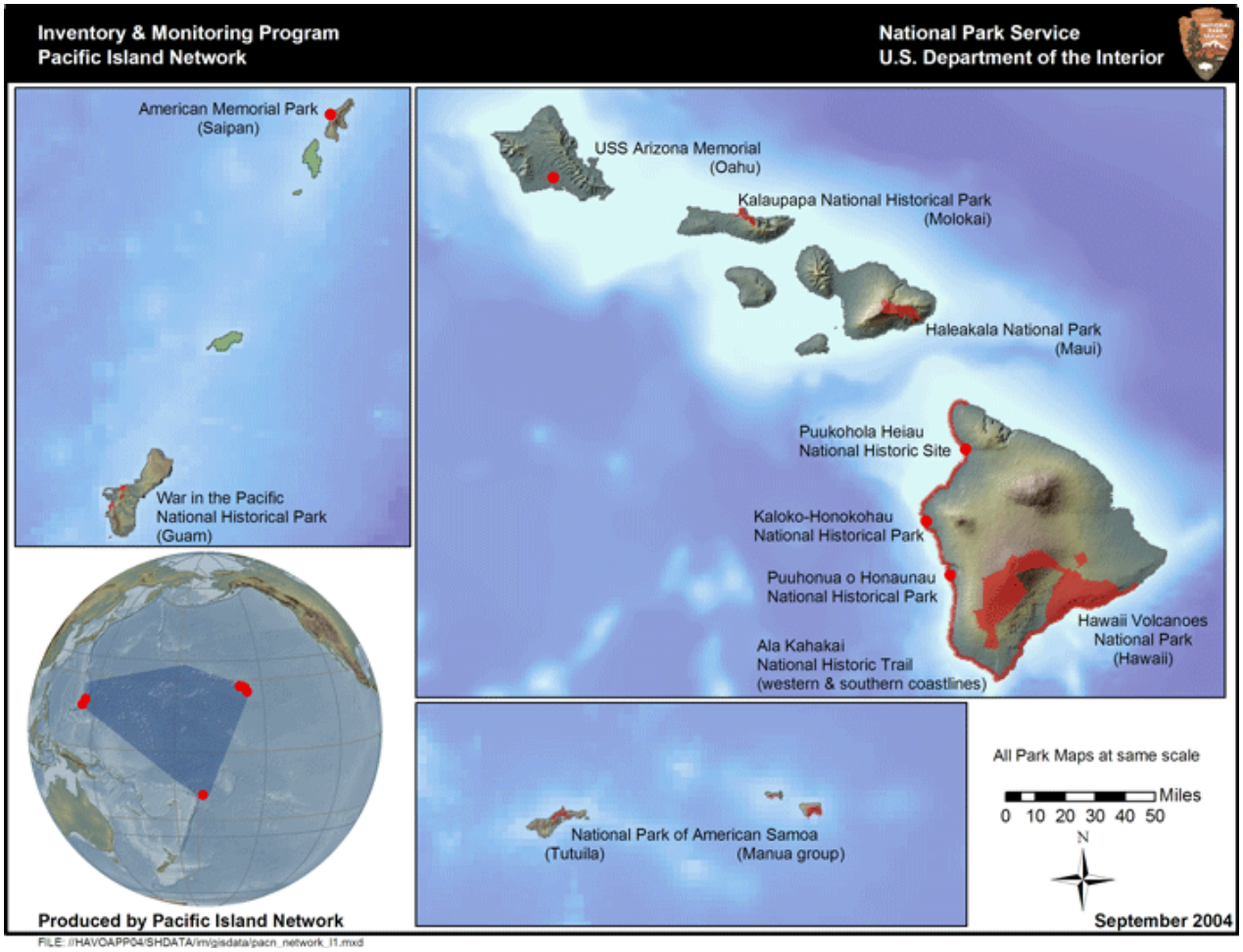


Figure 1.1. Map of the Pacific Island Network.

Below is a list of these 11 parks, their four-character park codes, and the island or territory in which they are located. This list is followed by a table showing, for each of these 11 parks, the

year the park was established, the length of the park coastline, the size of the park (terrestrial and marine), and the park's elevation range (Table 1.2).

War in the Pacific National Historical Park (WAPA) – Territory of Guam
American Memorial Park (AMME) – Saipan (Commonwealth of the Northern Mariana Islands)
National Park of American Samoa (NPSA) – Territory of American Samoa
USS Arizona Memorial (USAR) – Oahu, State of Hawaii
Kalaupapa National Historical Park (KALA) – Molokai, State of Hawaii
Haleakala National Park (HALE) – Maui, State of Hawaii
Ala Kahakai National Historic Trail (ALKA) – Hawaii, State of Hawaii
Puukohola Heiau National Historic Site (PUHE) -- Hawaii, State of Hawaii
Kaloko-Honokohau National Historical Park (KAHO) -- Hawaii, State of Hawaii
Puuhonua o Honaunau National Historical Park (PUHO) -- Hawaii, State of Hawaii
Hawaii Volcanoes National Park (HAVO) -- Hawaii, State of Hawaii

Table 1.2. *Summary of land and water spatial characteristics for each park in the PACN*.*

Park	Authorized (year)	Coastline length (km)	Authorized marine size (ha)[a]	Authorized terrestrial size (ha)	Authorized total size (ha)[b]	Elevation Range (m)[c]
WAPA	1978	9	401	400	801	-50 to 318
AMME	1976	5	0[e]	58	58	0 to 3
NPSA	1988	70	2129[e]	3786	5915	-50 to 966
USAR[d]	1978	1	2	4	7	-12 to 23
KALA	1980	21	834	3519	4353	-61 to 1287
HALE	1916	4	0	11849	11849	0 to 3055
ALKA[f]	2000	282	TBD	TBD	TBD	0 to 122
PUHE	1972	0	3	32	35	-15 to 52
KAHO	1978	5	217	264	481	-46 to 24
PUHO	1961	3	0	231	231	0 to 274
HAVO	1916	53	0	135091	135091	0 to 4169

* Data Sources (except as noted): NPS Lands Division Authorized Park Boundary Maps, USGS DLG Political Boundaries, TMK Parcels (from GDSI), and 2001 IKONOS Imagery (used for NPSA calculations). Converted to metric units when necessary. Compiled by S. Margriter.

a. Identified as 0 where authorized park boundary ends at high tide line. Includes the shoreline and offshore waters of inholdings. At NPSA the offshore lease area, through an agreement with the American Samoa Government, begins at the mean high tide line and extends seaward to 1/4 mile. In Hawaii, neither management nor ownership of marine areas has been transferred to the NPS (as is also the case in several terrestrial areas throughout the network).

b. Determined by adding authorized marine size and authorized terrestrial size.

c. Determined using figures provided by the NPS Lands division (otherwise using USGS digital elevation models for land surfaces and, when available, bathymetry data for marine areas). Refers to authorized lands only.

d. USAR does not have formal congressional authorization, but operates under a memorandum of agreement with the U.S. Navy. Coastline length and marine size are estimates. Terrestrial and total size are from park webpage.

e. NPSA marine area is newly authorized, but not yet funded.

f. Coastline length and elevation range are estimates only.

Regional oversight is provided to the PACN, and to seven other Pacific West Region networks, by the regional I&M Program coordinator located in Seattle, Washington. In addition, staff at Pacific West Regional Office - Honolulu participate substantially in PACN planning and operations.

The PACN Board of Directors is composed of superintendents, the regional I&M coordinator, and the PACN coordinator. And the PACN Technical Committee consists of park resource staff, USGS-Biological Resources Division staff, selected regional resource staff, and a representative from the USGS-Hawaiian Volcano Observatory. The PACN Charter provides detailed information about the Board of Directors and the Technical Committee (see PACN Monitoring Plan).

PACN staff composition will evolve as the monitoring phase of the program progresses. Staff composition is addressed in detail in Chapter 8 of the PACN Monitoring Plan.

1.9. Organization of this Data Management Plan

Chapters 1 through 4 of this DMP are overview chapters describing roles, processes and framework. Chapters 5 through 11 discuss data management operations. When applicable, legal mandates are presented at the beginning of the chapter. Acknowledgements and References are included at the end of this document. The content and presentation of this DMP benefited immensely from discussions with other network data managers, and this plan draws extensively from material developed by several of those data managers, as reflected in the Acknowledgements section. The References section provides a selected list of documents, guidelines or policies that may be helpful in further discussion and development of data management operations. Appendices present additional materials that augment the data management strategies and procedures presented in the body of this plan.

2. DATA MANAGEMENT ROLES AND RESPONSIBILITIES

Data management is about people and organizations as much as it is about information technology and database theory and application. Nearly every person in an organization manages data and information at some level, and good data stewardship is truly a collaborative endeavor. Every person involved in PACN network business must understand and perform their individual data stewardship responsibilities in the production, analysis, management, and/or end use of data and information (Table 2.1).

This chapter clarifies roles and outlines the principal and ancillary data management responsibilities of network staff, park staff, and cooperators conducting network projects. Project leaders and staff (i.e., individuals carrying out monitoring protocols or inventory study plans) are responsible for reading and understanding guidance at two levels – the general, overarching level of this Data Management Plan and the specific, task-oriented level of the respective protocol or study plan. The network data manager is responsible for creating and maintaining a network information management infrastructure and standards, and for helping project staff to understand and adhere to these network standards.

Table 2.1. *Categories of data stewardship involving all Network personnel and cooperators.*

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

**NOTE – Each position is listed in only one category according to primary responsibilities; however, most positions contribute to activities in each category.*

2.1. Data Stewardship Roles and Responsibilities

A role is a function or position (e.g., Project Leader)
A responsibility is a duty or obligation (e.g., review data records)

An increasing demand for more detailed, higher quality data and information about natural resources and ecosystem functions requires a group of people working together to steward data and information assets. Knowledgeable individuals from scientific, administrative, and technological disciplines must work in concert to ensure that data are collected using appropriate methods, and that resulting datasets, reports, maps, models, and other derived products are well managed. Datasets and the presentations of these data must be credible, representative, and available for current and future needs. Stewardship responsibilities apply to all personnel who handle, view, or manage data (Table 2.2). Vital sign monitoring protocols will describe more detailed, project-specific data stewardship roles and responsibilities.

Table 2.2. *Programmatic roles and associated data stewardship responsibilities.*

Role	Data Stewardship Responsibilities
Project Crew Member	Record and verify measurements and observations based on project objectives and protocols. Document methods and procedures.
Project Leader/Principal Investigator	Direct project operations. Communicate data management requirements and protocols to project staff, network data manager, and resource specialist(s). Supervise crew members to ensure adherence to data collection and data processing protocols. Ensure data verification, validation, and documentation. Responsible for final submission and review of all products and deliverables.
Database Manager (Project)	Apply particular knowledge and abilities related to database software and associated application(s).
Data/GIS Specialist or Technician (Project or Network)	Support the network data manager and project leaders in implementing data management procedures. Assignments may include GIS activities, cataloging, data conversion and integration, documentation, and reporting.
Statistician/Biometrician (Project or Network)	Develop project sampling design. Analyze data, develop models, and documentation procedures. Review methods for statistical soundness.
Park Research Coordinator	Facilitate data acquisition by external researchers; communicate NPS requirements to permit holders.
Park Resource Specialist	Understand project objectives, data, and management relevance. Make decisions about validity, sensitivity, and availability of data. Describe, publish, release, and discuss data and information products.
Curator (Park or Region)	Manage collection, documentation, and preservation of specimens.
Network Data Manager	Oversee development, implementation, and maintenance of data infrastructure and standards. Facilitate and integrate data and metadata. Oversee long-term data storage and maintenance. Design and develop databases and applications.
Network Ecologist	Ensure useful data are collected and managed by integrating natural resource science in network activities and products, including specifying objectives, sample design, data analysis, synthesis, and reporting.
Network Coordinator	Ensure programmatic data and information management requirements are met as part of overall network business.
GIS Manager (Region)	Provide support for long-term storage of GIS data. Update and maintain GIS software and tools. Provide technical assistance.
Information Technology Specialist (Region)	Maintain local area network, establish and maintain system security, update software and hardware, and implement secure file server backup system.
I&M Data Manager (National)	Provide service-wide database design, support, and services, including processing to convert, store, and archive data in service-wide databases.
End Users (managers, scientists, interpreters, public)	Provide feedback on scientific information, presentation needs, and interpretation.

Among all of the data management responsibilities shared by project and network personnel, the careful documentation of datasets, data source(s), and data collection methodology is paramount. This careful and thorough documentation establishes the basis for the appropriate use of the data

in resulting analysis and products, both in the short-term and long-term. PACN monitoring protocols contain key elements of data documentation. Network data records collected according to these protocols will include the name, date, and version of the associated protocol. (Chapters 7 and 8 present important guidance and reference for documentation and metadata.)

2.2. The Hub of Data Stewardship

Network coordinators, project leaders, data managers, and GIS specialists comprise the central data management team for inventory and monitoring projects. Each is responsible for certain aspects of project data and all share responsibility for some overlapping tasks (Figure 2.1). Because of the collaborative nature of project data management, good communication among these personnel is essential to meeting program goals.

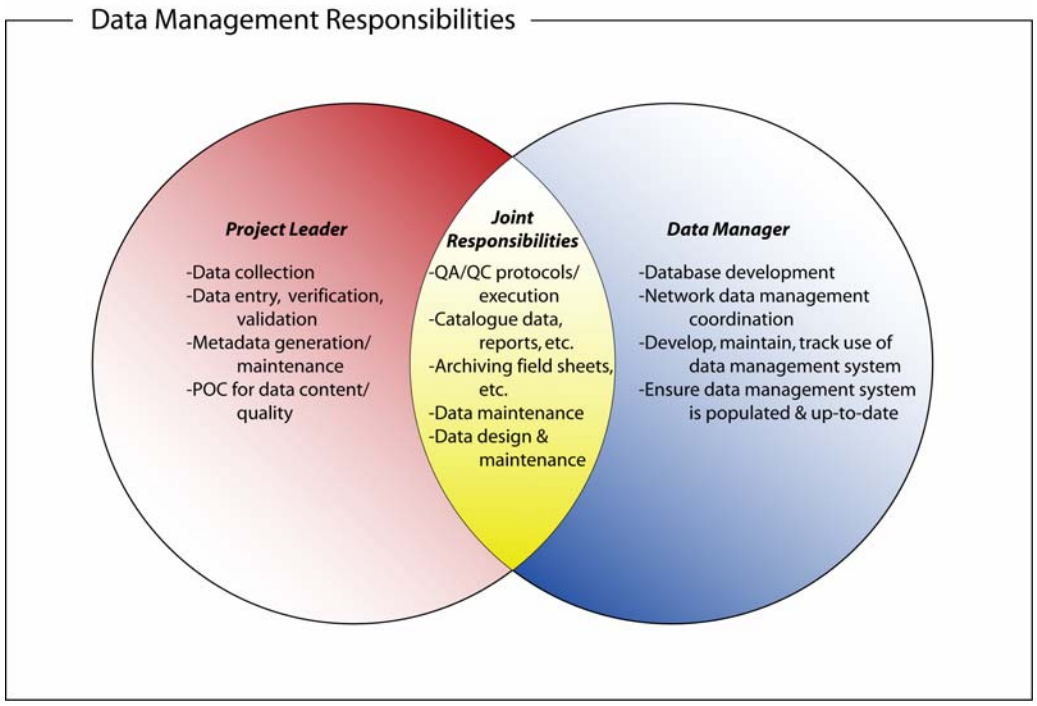


Figure 2.1. Core data management responsibilities for a project leader and the network data manager.

Communication is promoted by providing:

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

Network Coordinator

The Network Coordinator supervises the project leaders and has the ultimate responsibility for data entry, validation, verification, summarization/analysis and reporting. The Board of Directors

and Technical Committee give final approval for products to be integrated, distributed to the public or parks, or protected, in the case of sensitive information.

Project Leader

Project leaders oversee and supervise all phases of an inventory and monitoring project from initiation to product delivery. Assuming "ownership" of the data, the project leader's active involvement in data management determines the quality and usefulness of the project data. The project leader provides project oversight, directs on-the-ground data collection efforts, and provides continuity and cohesion among data collection, synthesis, interpretation, and reporting. Project leaders must act as a steward for all project data, and must work with project and PACN personnel to ensure the proper handling of project data. Project leaders are responsible for designating an alternate leader who is capable of maintaining project operations in his or her absence, in order to ensure project continuity and data integrity.

Specifically, a project leader is responsible for:

- project documentation that describes the 'who,' 'what,' 'where,' 'when,' 'why,' and 'how' of a project
- documentation and implementation of standard procedures for field data collection and data handling
- quality assurance and quality control measures, which include the supervision and certification of all field operations, staff training, equipment calibration, species identification, data collection, data entry, verification, and validation
- maintenance of concise explanatory documentation of all deviations from standard procedures
- detailed post field-collection documentation
- maintenance of hard copies of data forms and archiving of original data forms
- scheduling of regular project milestones such as data collection periods, data-processing target dates, and reporting deadlines
- regular summary reports, periodic trend analysis of data, resulting reports, and their public availability
- identifying sensitive information that requires special consideration prior to distribution
- acting as the main point of contact concerning data content/quality

The project leader may also work closely with the data manager and/or a data specialist (such as a biometrician) to:

- develop quality assurance and quality control procedures specific to project operations
- identify training needs for staff related to data management and quality control procedures
- coordinate changes to the field data forms and the user interface for the project database (coordinate data entry procedures – data design maintenance)
- document and maintain master data (including metadata generation and maintenance)
- manage the archival process to ensure regular archival of project documentation, original field data, databases, reports and summaries, and other products from the project
- define the process of how project data will be transformed from raw data into meaningful information
- create data summary procedures to automate and standardize this transformation process

- identify and prioritize legacy data for conversion and convert priority datasets to a modern format
- increase the interpretability and accessibility of existing natural resource information
- catalog project data and reports in nationally maintained I&M Program databases such as NPSpecies, NatureBib and NR-GIS (see Chapter 10)

Network Data Manager

The data manager oversees the development, implementation, and maintenance of data infrastructure and standards for the PACN. The data manager is responsible for ensuring the compatibility of project data with program standards and the long-term integrity and availability of project data. Working with project leaders and other network staff to design databases, applications, and products, the data manager facilitates dissemination of project datasets and information products. The data manager exercises data stewardship to ensure data are archived, documented, and compatible with other program data. The data manager ensures the maintenance of mechanisms for making data discoverable and available. Both the data manager and the project leader ensure that the information conveyed via these mechanisms is up-to-date and accurate.

General data management duties for the data manager are:

- overall coordination of data management activities in the network
- develop and maintain data and information storage and dissemination system for all program data and information
- work to improve the acquisition, accessibility and transparency of digital data
- ensure that the data and information system is populated and kept up-to-date with all relevant network output
- develop and maintain logs recording changes/enhancements to data products, data management processes, and/or the data and information system
- act as point of contact for access to network data and information
- ensure data security (archiving operations, etc.)

Data managers will also work closely with project leaders to:

- develop and maintain the infrastructure for metadata creation, project documentation, and project data management
- create and maintain project databases in accordance with best practices and current standards
- provide training in the theory and practice of data management tailored to the needs of project personnel
- establish and implement procedures to protect sensitive data according to project needs
- collaborate with GIS specialists to integrate tabular data with geospatial data in a GIS system in a manner that meets project objectives
- define the scope of the project data and create a data structure that meets project needs
- become familiar with how the data are collected, handled, and used
- review quality control and quality assurance aspects of project protocols and standard operating procedures

- identify elements that can be built into the database structure to facilitate quality control, such as required fields, range limits, pick-lists and conditional validation rules
- create a user interface that streamlines the process of data entry, review, validation, and summarization that is consistent with the capabilities of the project staff
- develop automated database procedures to improve the efficiency of the data summarization and reporting process
- make sure that project documentation is complete, complies with metadata requirements, and supports the interpretability and longevity of the project data
- ensure regular archiving of project materials
- inform project staff of changes and advances in data management practices

Regional and Network GIS Specialists

The regional GIS Specialist and network GIS cooperators play a crucial supporting role for PACN projects. GIS support includes project planning to determine the GIS data and analysis needs for a project. GIS specialists will provide consultation to project leaders on field collection of spatial data including the use of GPS and other spatial data collection techniques. They will also coordinate importing of spatial data into the GIS, work with project leaders to analyze spatial data, and provide the results in map or tabular form. The GIS specialists will work with project leaders to properly document data in compliance with spatial metadata standards and will be responsible for stewardship of GIS data and products.

GIS specialists will also work directly with data managers to design databases and applications, to create relationships between GIS and non-spatial data, and to create appropriate database and GIS applications to facilitate the analysis of both spatial and non-spatial data. Maintaining standards for geographic data, GIS specialists are responsible for sharing and disseminating GIS data throughout the network.

2.3. Data Management Coordination

The Natural Resource Challenge states that collaboration among the National Park Service, other public agencies, universities, and non-governmental organizations is necessary to effectively acquire, apply, and promulgate the scientific knowledge gained in national parks.

The network data managers work with I&M Program data management staff and regional resource information management personnel to maintain a high level of involvement in service-wide and regional databases and data management policy. The data managers also work locally with network personnel, park staff, and cooperators to promote and develop workable standards and procedures that result in the compatibility and availability of datasets.

Key contacts include regional and park GIS specialists and the project leaders for each monitoring or inventory project. Involvement and input from park scientists and resource management staff is also essential. Each of these individuals plays an integral role in the successful development of planning materials, inventory study plans, and monitoring protocols. Consistent and productive communication among these individuals leads to common understanding and better synchronization of data management activities. This communication

may take the form of personal visits, phone calls, email, joint meetings and training sessions, and participation in meetings of the PACN Technical Committee and Board of Directors.

Data managers throughout the NPS regularly coordinate with each other and national program staff via annual meetings, conference calls, workgroups, a listserv, web sites, and informal communication. Data managers have already demonstrated effective cooperation in the ways that they share ideas and technology, and collaborate to develop data management standards and documentation. This communication and cooperation promotes practical levels of compatibility among protocols and datasets among networks and organizations.

3. PROJECT AND DATA MANAGEMENT OVERVIEW

This chapter describes the general work flow of projects and, in light of the data stewardship responsibilities presented in the previous chapter, the general process of generating, documenting, and integrating project products within the broad I&M data management framework.

A project is a short-term (e.g., inventory) or long-term (e.g., monitoring protocol) effort to collect, analyze, and report natural resource data, and includes efforts external to the I&M Program that generate data of interest to PACN. The effectiveness and future usefulness of a project depends upon data management considerations being pervasive in project planning, execution, and documentation. Making a distinction between project management and I&M data management helps to illustrate the coordination required to make effective use of the I&M Program's database standards, data documentation strategies, and data dissemination repositories. In most instances, PACN projects are carried out through contractual or cooperative agreements. Project leaders, or principal investigators, direct and coordinate all project data collection, data verification and validation, data analysis, and project reporting. PACN staff, and the data manager in particular, provides guidance and assistance in establishing standard operating procedures, review of project products to ensure compliance with all I&M Program standards and requirements, and integration into service-wide databases and repositories.

A generalized work flow model provides a framework for how data management elements fit into the five primary stages of project management: planning and approval; design and testing; implementation; product integration; and evaluation and closure (Figure 3.1). To further illuminate project management details, each of these stages can be broken into a set of tasks (Table 3.1). A successful project and a productive project evaluation and closure require the participation of the project leader, project staff, and all PACN staff. The resources and coordination necessary to produce, maintain, and deliver quality data and information are highlighted in this chapter. More details about data acquisition, quality assurance, data documentation, data maintenance, and information dissemination can be found in later chapters of this plan.

3.1. Project Management versus Data Management

If project management can be viewed as the logistics of carrying a project from initial planning to completion of final reports, publications, and presentations, then data management can be seen as the fine details concerning data collected and information generated along the way. In shepherding a project from conception to completion, the project leader holds primary responsibility for the first three project management stages – planning and approval, design and testing, and implementation. The bulk of data management considerations fall into the design and testing stage and into the fourth stage, product integration. Holding primary responsibility for data management considerations, the PACN data manager is closely involved in the database design component of the design and testing stage, and plays a central role in ensuring integration of project products into the broad I&M data management framework. Evaluation and closure stage responsibility rests equally with the project leader, project staff, the PACN data manager, and the rest of the PACN staff. While the entire PACN staff serves a review and approval role

for all project activities and delivered products, the following project workflow discussion is presented in light of this distinction between project management and data management, or between project leader responsibilities and the responsibilities of PACN data management staff.

3.2. Project Workflow

Projects may be either short-term or long-term. Short-term projects will include research, inventory, and pilot projects to prepare for long-term monitoring. Long-term projects will include the PACN vital sign monitoring protocols and other park-based long-term monitoring. Long-term projects generally have stricter requirements for adhering to standards, protocols, and peer review, including documentation of changes to protocols and standard operating procedures. Short- and long-term projects share workflow characteristics and both generate products that must be managed and made available. The workflow of project management can be divided into five primary stages (Figure 3.1), each of which can be further described as a set of project management and data management tasks (Table 3.1).

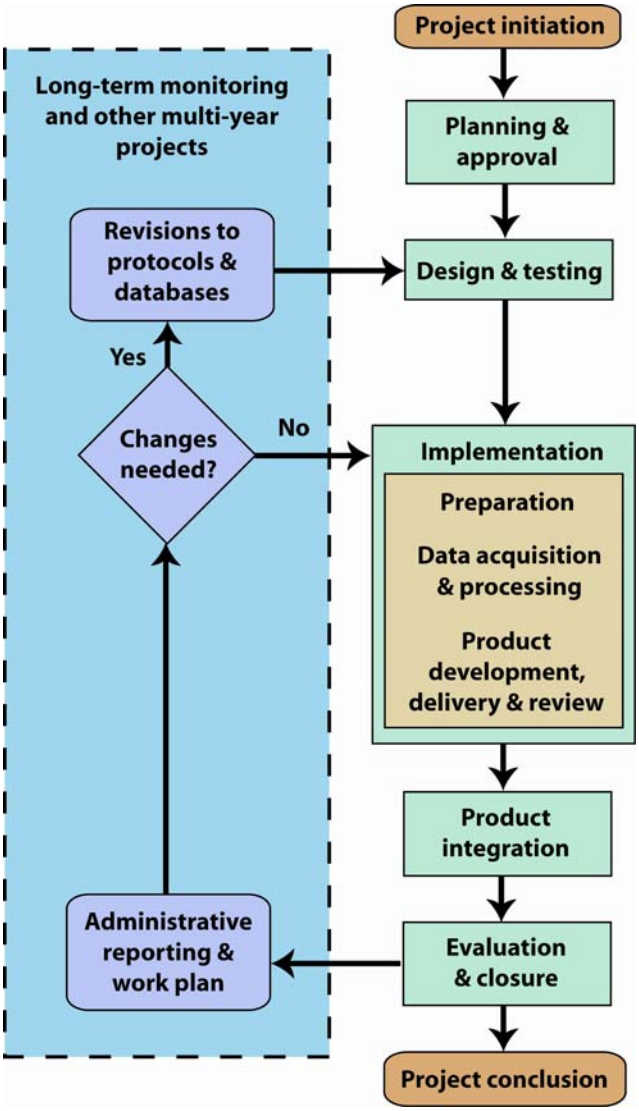


Figure 3.1. Project and information workflow model depicting the five primary project management stages.

3.2.1. Planning and Approval

Project Management

This is the stage during which the project leader coordinates preliminary decisions regarding project scope and objectives, seeks and secures funding sources, and addresses permits and compliance. Completion of these tasks is the responsibility of the project leader and program administrators. A detailed study plan will form the basis of any applicable contracts or cooperative agreements, and of applicable permits. Addressing compliance issues in detail may be delayed until project methods are thoroughly developed.

Data Management

Although this stage lacks specific data management activities, it is important that the PACN data manager remains informed so that project tracking can be initiated and input provided on timelines for receipt and review of deliverables. Involvement of PACN staff will help ensure that all contracts, agreements and permits include standard language that describes the formats, specifications, and timelines for project deliverables. A guidance document on project organization and project deliverables is provided to all project leaders at the time of project initiation (Appendix A).

3.2.2. Design and Testing

Project Management

Details about how data will be acquired, processed, analyzed, reported and made available to others are worked out during this project stage. The project leader is responsible for developing and testing project methodology, or for modifying existing methods to meet project objectives. Collaboration between the project leader and the PACN data manager throughout this stage will help to build and reinforce good data management throughout the project, which is especially critical for data acquisition, processing, and retrieval. By beginning collaborative development as soon after project approval as possible, data integrity and quality can most easily be assured. An important part of this collaboration is defining the data parameters to be collected and establishing the database design structure and associated data dictionary. Devoting adequate attention to this aspect of project development is possibly the single most important part of assuring the quality, integrity and usability of the resulting data. Development of field forms, databases, and standard operating procedures (for data collection, data entry and validation, and analysis and reporting) is the responsibility of the project leader. Because data products delivered by a project must be approved by the PACN data manger, close communication with the PACN data manager is strongly encouraged.

Data Management

The PACN data manager will communicate established I&M standards regarding development of database applications, data dictionaries, and metadata records, as well as make available all pertinent guidance documents and training materials. In addition to reviewing proposed database design structures and making recommendations for design modifications, the PACN data manager will approve final databases. In some cases, PACN data management staff may take on primary responsibility for database development and documentation. As time permits, PACN

Table 3.1. List of tasks involved in each project management stage and the associated staff responsibilities. PL = Project Leader; S/E = Statistician or Ecologist; DM = Data Manager; GIS = GIS Specialist. * = Primary role; - = Secondary role.

Project Stage	Task	Task Description	PL	S/E	DM	GIS
Planning & Approval	1.1	notify data manager of new project in planning stage	*			
	1.2	develop study plan and budget	*	-	-	
	1.3	initiate project tracking	-		*	
	1.4	identify project deliverables	*		-	-
	1.5	determine data sensitivity and ownership	*		-	-
	1.6	develop language for contracts, agreements, and permits	*	-	-	-
	1.7	apply for permits and update project tracking	*		-	
Design & Testing	2.1	notify data manager of changing status	*			
	2.2	update project tracking	-		*	
	2.3	clarify roles and responsibilities	*	*	*	*
	2.4	scoping for information systems design	*	*	*	*
	2.5	initiate metadata record	*		-	-
	2.6	create data dictionary and data forms	*		*	*
	2.7	define quality assurance measures	*	-	-	
	2.8	database and application development	*	-	*	*
	2.9	sample site selection and documentation	*	*	-	*
	2.10	provide training as needed and time permits			*	*
Implementation	3.1	notify data manager of changing status	*			
	3.2	training	*			
	3.3	update project tracking	-		*	
	3.4	data acquisition, data entry, upload data logger files	*			
	3.5	upload GPS and other data logger data into database	*			-
	3.6	quality assurance: verification, validation and processing, data certification	*	*		
	3.7	summarization, analysis, map production	*	*	-	-
	3.8	interpret results, reporting, peer review	*	*		
	3.9	check for completeness, proper format, metadata, and certification	*		-	-
	3.10	assemble and deliver all data products and metadata to DM/GIS (see Task 5.2)	*			
Product Integration	4.1	verify completeness, proper format, metadata, and certification	-		*	*
	4.2	update project tracking (deliverables complete and published)			*	-
	4.3	upload certified data into master project database(s)			*	-
	4.4	finalize and parse metadata record	-		*	*
	4.5	publish products and integrate with national applications (NatureBib, NPSpecies, NPSTORET, NR-GIS Data Store)			*	*
	4.6	electronic permanent storage			*	*
	4.7	hardcopy archiving and distribution	*		-	
Evaluation & Closure	5.1	verify that all products are accounted for	*		*	*
	5.2	administrative evaluation and reporting, decisions on project continuation and needed methodology changes	*	-		
	5.3	communicate administrative decisions and information management needs to data manager and GIS specialist	*			
	5.4	update project tracking	-		*	

data management staff may provide training to project staff on subjects such as GPS data collection and downloading, and GIS data processing.

3.2.3. Implementation

Project Management

During the implementation stage, data are acquired, processed, error-checked and documented. Activities may include data entry, quality control/quality assurance procedures, preliminary analysis, labeling and storage of materials, and documentation. Methods for data entry and verification must adhere to standard operating procedures (SOPs); any modifications, if needed, must be documented in those SOPs. The implementation stage is when products such as reports, databases, maps, and GIS themes are finalized. The project leader oversees all aspects of implementation – from logistics planning, contracting, training, and equipment procurement to data acquisition, data entry and verification, metadata creation, report preparation, and final product delivery. Project budgets should cover the hiring of staff with appropriate technical abilities, and the roles of PACN data management staff should be established in advance of project implementation. Toward the end of this stage, all product deliverables identified in project planning documents (i.e., protocol, study plan, contract, agreement, and/or permit) are submitted. In general, all raw and derived data products, metadata, reports and other documentation should be delivered to the PACN staff and to all appropriate park or regional staff (e.g., program administrators, collections managers, etc.). All project deliverables should be developed and delivered according to product specifications, which should be stipulated in all project planning documents. Products that do not meet program requirements must be revised.

Data Management

Throughout this stage, PACN data management staff function primarily as facilitators – providing support for database applications, GPS data collection, GIS, and other data processing applications; facilitating data summarization, validation, and analysis; and assisting with the technical aspects of product development and documentation. The specific roles of PACN data management staff will depend primarily on the technical capabilities of the project staff. All data-related project products are reviewed by PACN data management staff, and the PACN data manager is responsible for maintenance of final digital data products within the PACN infrastructure (see Chapter 11).

3.2.4. Product Integration

Project Management

The project leader must ensure the timely delivery of products that meet I&M program needs and technical standards specified during the planning and approval stage and the design and testing stage. If the project has additional integration needs, such as collaboration with other agencies or institutions using a common database, then successfully meeting these needs is the responsibility of the project leader.

Data Management

Primary responsibility for integration of project products into the broad I&M data management framework lies with the PACN data manager. This integration includes entering reports into appropriate bibliographic databases, integrating datasets and metadata into service-wide databases and repositories, archiving reports and data products, and distributing or otherwise making products available to their intended audiences. Archiving of digital data products is described in more detail in Chapter 11, while archiving and cataloging of non-digital project products by individual park curatorial programs is discussed in Chapter 12.

3.2.5. Evaluation and Closure

Project Management

For long-term monitoring and other cyclic projects, the evaluation and closure stage occurs when annual products (e.g., progress report and analysis summaries, updates to database and GIS themes, and metadata) are completed. For non-cyclic projects, this stage represents the completion of the project. Evaluation and closure is a collaborative effort to assess how well the project met its objectives, and to determine what might be done to improve various aspects of the methodology, implementation, and formats of the resulting information. The project leader and project staff bring a project perspective to this evaluation and discussion on project improvement. The project leader is responsible for recording all modifications to protocols and SOPs in the edit tracking table of each respective document. Major revisions may require additional peer review.

Data Management

The entire PACN staff participates in this annual evaluation of the project, and in making recommendations on how to improve the project. PACN data management staff is responsible for updating records to reflect project status and deliverables in network project tracking applications. If the PACN data management staff designed and maintains the project database, then they are responsible for documenting any modifications to database design structure and/or the data dictionary and database documentation.

3.3. Data Life Cycle

During a project's various stages, project data take different forms and are maintained in different places as they are acquired, processed, documented and archived. This data life cycle can be modeled as a sequence of events and tasks (Figure 3.2), which involve interaction with the following objects:

- *Raw data* – Analog data recorded by hand on hard-copy forms and digital files from handheld computers, GPS receivers, telemetry data loggers, etc.
- *Working database* – A project-specific database for entering and processing data for the current season (or other logical time period). This may be the only database for short-term projects with no need to distinguish current season data from the full set of validated data
- *Certified data and metadata* – Completed data and documentation for short-term projects, or one season of completed data for long-term monitoring projects. Certification is a confirmation by the project leader that the data have passed all quality assurance requirements and are complete and ready for distribution. Metadata records include the

detailed information about project data needed for its proper use and interpretation (see Chapter 8)

- *Master database* – Project-specific database for storing the full set of validated project data, used for viewing, summarizing, and analysis. Current season data from working database must pass all quality assurance steps prior to upload into this master project database
- *Reports and data products* – Information that is derived from certified project data
- *Edit log* – A means of tracking changes to certified data
- *National databases and repositories* – Applications and repositories maintained at the national level, primarily for the purpose of integration among NPS units and for sharing information with cooperators and the public (see Chapter 10)
- *Digital libraries and archiving* – All digital files associated with a project are stored on file servers at the network level, with data backups occurring through network-attached digital storage and off-site storage of backup tapes. Archiving of all project hard-copy items is accomplished at the park level by cultural resources staff, with coordination from both the project leader and the PACN staff

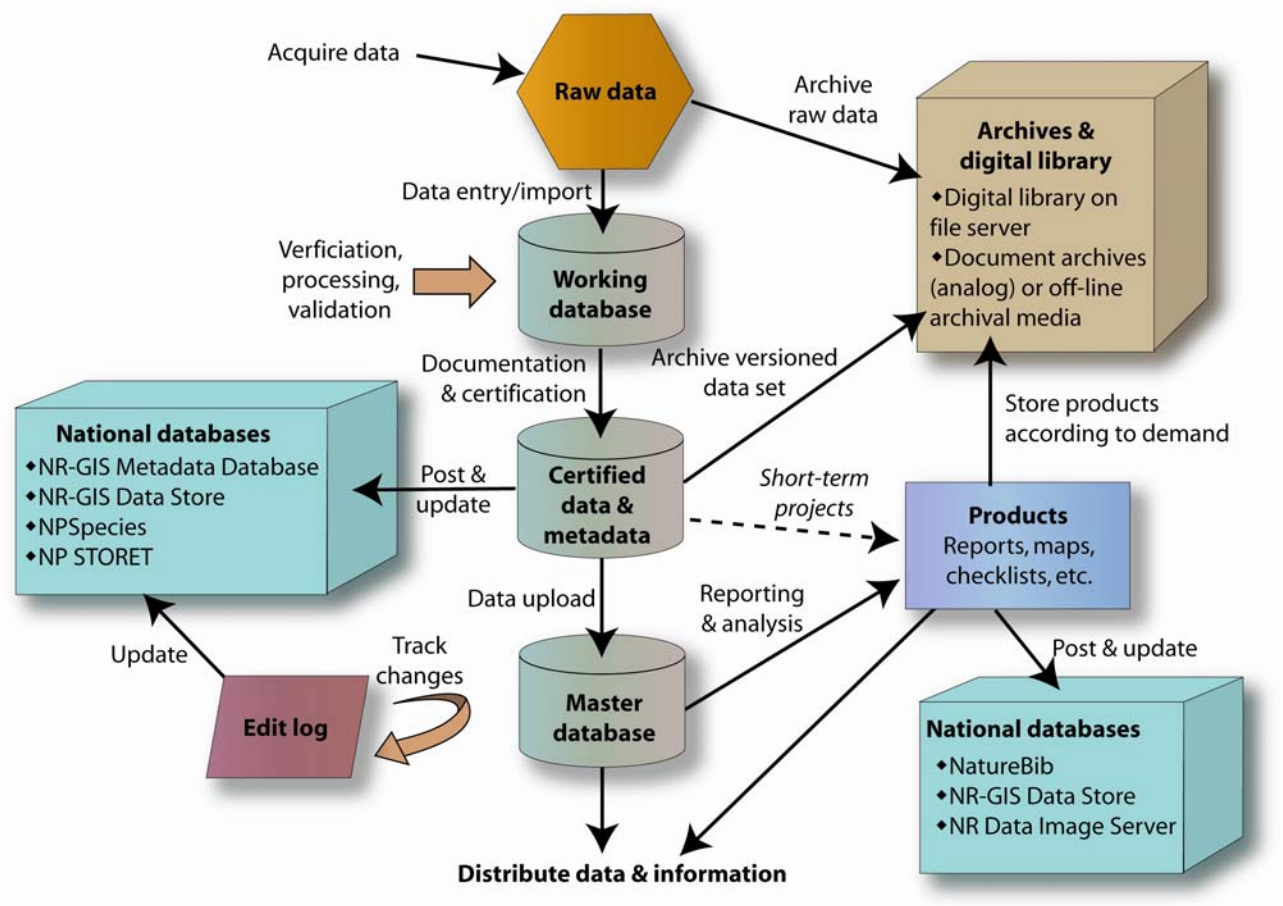


Figure 3.2. Data life cycle workflow model depicting sequence of associated events and tasks.

Although the data life cycle may vary depending on specific project needs and objectives, in general, the third and fourth project management stages (implementation and product integration) will accomplish the following data life cycle events and tasks (from Figure 3.2):

1. *Acquire data* – For data recorded by hand in the field, data forms should be reviewed regularly (at least daily) for completeness and validity in order to capture errors as close to their origin as possible
2. *Archive raw data* – Copies of all raw data files are archived intact. Digital files are copied to the digital library section for the project; hard copy forms are either scanned and placed in the digital library, or are copied and placed in the archives. Archiving or scanning of hard copy data forms may occur at the end of a season as a means of retaining all marks and edits made during the verification and validation steps
3. *Data entry / import* – Analog data are entered manually, and digital data files are uploaded to the working database
4. *Verification, processing and validation* – Verify accurate transcription of raw data; process data to remove missing values and other data flaws; and validate data using database queries to capture missing data, out-of-range values, and logical errors
5. *Documentation and certification* – Develop or update project metadata and certify the data set. Certification is a confirmation by the project leader that the data have passed all quality assurance requirements and are complete and documented. It also means that data and metadata are ready to be posted and delivered
6. *Archive versioned data set* – Copies of the certified data and metadata are placed in the digital library. This can be accomplished by storing a compressed copy of the working database, or by exporting data to a more software-independent format (e.g., ASCII text; see Chapter 10)
7. *Extract data, post data, and update national databases* – To make data available to others, certified data and metadata are posted to national repositories such as NR-GIS Data Store. In addition, national databases such as NPSpecies, NPSTORET, and NR-GIS Metadata Database are updated with data extracted from the certified data set. Note: Data and data products may not be posted if they contain protected information about the nature or location of rare, threatened or endangered species, or other natural resources of management concern (see Chapter 9)
8. *Upload data* – Certified data are uploaded from the working database to the master project database. This step might be skipped for short-term projects where there is no need to distinguish working data for the current season from the full set of certified project data
9. *Reporting and analysis* – Certified data are used to generate data products, analysis, and reports, including semi-automated annual summary reports for monitoring projects. Depending on project needs, data might be exported for analysis or summarized within the database
10. *Store products* – Reports and other data products are stored according to format and likely demand – either in the digital library, on off-line media, or in the document archives
11. *Post products and update national databases* – To make data available to others, reports and other products are posted to national repositories such as NR-GIS Data Store or the NR Data Image Server. In addition, products are catalogued in NatureBib. Data products may not be posted if they contain protected information about the nature or location of rare, threatened or endangered species, or other natural resources of management concern (see Chapter 9)
12. *Distribute data and information* – Data, metadata, reports and products can be shared and distributed in a variety of ways – especially via the web-based national databases and repositories, by FTP or mailing in response to specific requests, or by providing direct

access to project records to cooperators. In all cases, distribution will follow legal requirements under the Freedom of Information Act, and limitations established to protect information about sensitive resources (see Chapter 9)

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

For long-term projects, this sequence of events occurs in an iterative fashion, repeating at the end of each field season or other logical data collection and reporting period. Conversely, this sequence is followed only once for short-term projects.

4. INFRASTRUCTURE AND SYSTEM ARCHITECTURE

The ultimate measure of success for the data management strategy of the I&M Program is to supply quality information to the park and scientific communities for research and analysis. The most common mechanism for delivering information to parks and scientists is by the use of a data clearinghouse. The term 'clearinghouse' loosely refers to one-stop-shopping on a website. Clearinghouses provide a mechanism for searching, browsing and downloading information at the convenience of the user.

This chapter discusses the components of the PACN computer resources infrastructure and the I&M Program's overall information system architecture that allow data and information to be served to a broad audience. The PACN infrastructure is composed of computers and servers that are functionally or directly linked through computer networking services, and represents the foundation upon which the network information system is built. System architecture refers to the applications, database systems, repositories, and software tools that make up the framework of the I&M Program's data management enterprise.

4.1. Computer Resources Infrastructure

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

The PACN infrastructure consists of park and network servers available to staff workstations via the wide-area network (WAN), and servers maintained at the service-wide level that host web-based applications and clearinghouses (Figure 4.1). This infrastructure is maintained by park, regional, and national IT specialists who administer all aspects of system security and backups.

This infrastructure supports these critical functions:

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

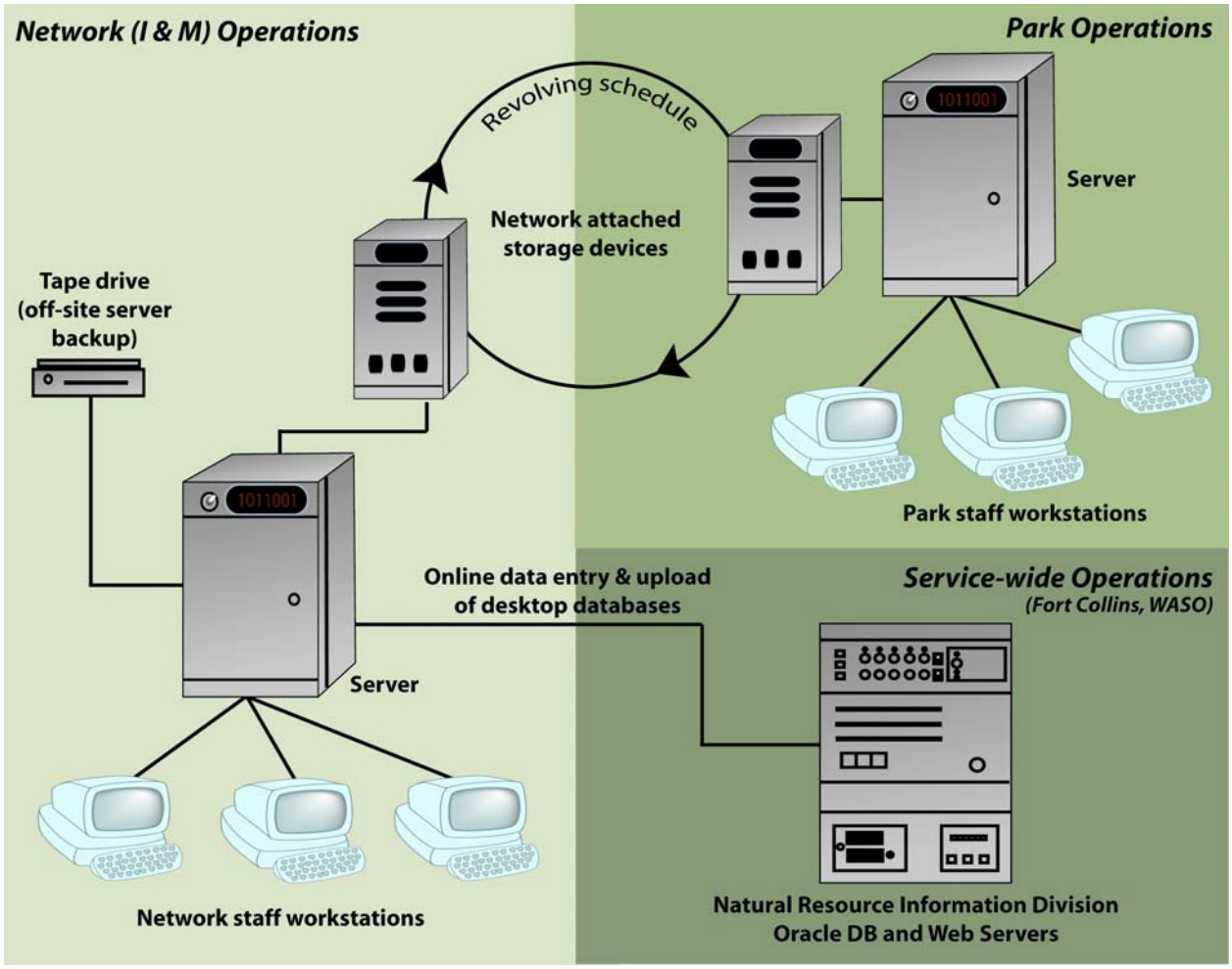


Figure 4.1. Schematic representing the logical layout and connectivity of the PACN infrastructure. The bulk of the PACN information system is maintained on servers at the service-wide and network levels.

4.1.1. Computer Server Infrastructure

Computer servers can be divided into five primary functions:

- *Work File Servers* – Typically team servers or user servers, where anyone can store working files. These servers work similarly to a PC's C: drive, except that they greatly facilitate data backups and file sharing
- *Data File Servers* - Read-only repository of data files. Files stored here do not typically change often and need a consistent location, such as in the case of GIS data, imagery, or final reports. Write access is under strict control, whereas all NPS employees have read access. A contact person monitors what is stored in this location to ensure proper archiving and documentation
- *Database Servers* – Uses special database software, such as SQL Server, Oracle, and ArcSDE. Write access is under strict control. Databases on these servers are for long-term data storage, and require quality database design, documentation, and administration
- *Application Servers* –Provide the applications, or "front end", to accessing the data. These are closely tied to the data file and database servers, but should be located on a different

physical server. Files on these servers change frequently and backups need to support these changes. Application servers are often Internet servers as well

- *Internet Servers* – Provide access to applications and files via the internet and subject to security restrictions. Special software may be used on internet servers, such as ColdFusion for website pages and database applications, Blue Angel for clearinghouse search engines, and ArcIMS for internet interactive mapping

The PACN infrastructure takes advantage of existing infrastructure established at the service-wide, regional, and park levels (Figure 4.2). Each level hosts some portion of PACN information.

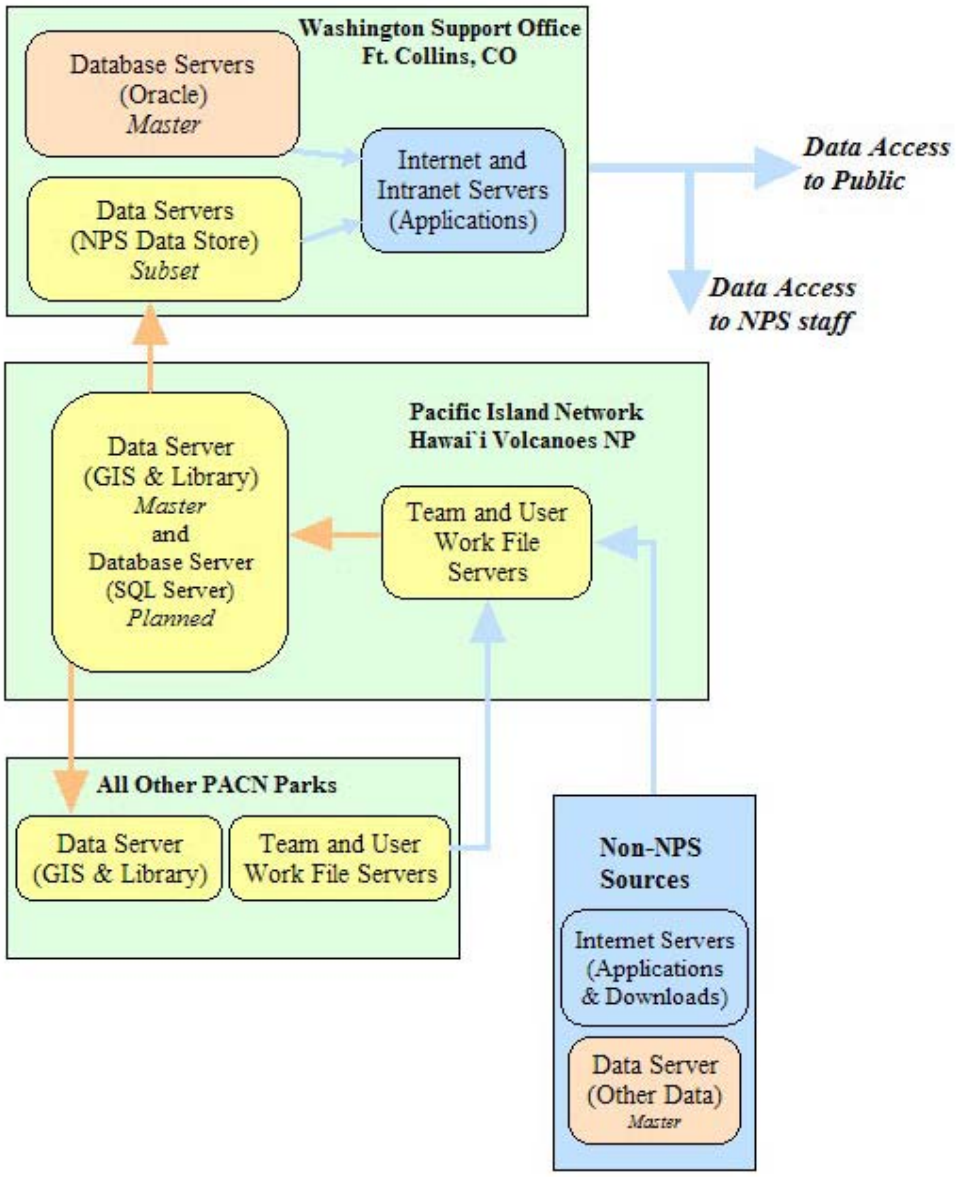


Figure 4.2. Schematic representing the park, regional, and service-wide server infrastructure.

4.2. PACN System Architecture

PACN offices are located at Hawaii Volcanoes National Park (HAVO), and staff workstations access park, regional, and service-wide servers via the regional wide-area network (WAN). Offices of the Pacific West Regional Office – Honolulu (PWRO-HNL) IT Specialists are also located at HAVO, which is where most of the PWRO-HNL servers reside. While both PWRO-HNL and PACN use some park servers to host redistributed copies of digital libraries, GIS data, and metadata, PACN maintains digital libraries and network and project working files on the PACN file server. WAN connectivity between park, network, and PWRO-HNL servers allows active project files to be worked on locally while enabling park and network staff to readily share files (see Figure 4.1). All data and information products from completed projects are archived on the PACN server and redistributed in read-only format to parks. Electronic files of final products are posted to the NR-GIS Metadata and Data Store application (NR-GIS Data Store) for public distribution through the NPS Focus clearinghouse.

4.2.1. Team and User Project Files

Working files for both short- and long-term projects are typically stored on work file servers at a park or at the PACN offices. PACN staff maintains a working directory for each project, and project leaders may modify a directory as needed. When working with contractors, a project directory may be "checked out" from the PACN server with the understanding that a similar structure will be delivered as part of the project final products, and that all files will follow PACN guidance regarding file naming and documentation standards.

As protocol development for a given PACN vital sign monitoring project advances, a project directory structure will be established, with appropriate server access permissions granted to the project leader and project staff. All project personnel will have access to PACN guidance regarding product deliverables identified in project contract/agreement, project tracking requirements, and standards for products, documentation, and metadata (see Chapter 8).

4.2.2. Integration of Products to Master Data Libraries

Digital versions of all deliverables, as specified in cooperative agreements, permits, study plans, or other administrative records, will be electronically stored on servers by the following general methods:

1. Project archive on PACN server, with all certified, integral files associated with the project "packaged" together (i.e., bundled or zipped together as a stand alone, easily distributed entity). Types of information to be included are listed in Chapter 1, Table 1.1
2. Product Library on PACN server, where like products are stored with other products (such as the final report stored with other reports)
3. Final products and long-term datasets posted for access by a broad audience (e.g., NR-GIS Data Store), and integrated into WASO Oracle database servers (e.g., NPSpecies and NatureBib) and/or fed as an annual network dataset to a service-wide database (e.g., NPSTORET)

These files will be centrally stored in the master data libraries on the PACN server using a generalized directory structure (Table 4.1). Files will be reviewed for quality and standards prior to storage in the data libraries.

Table 4.1. *General directory structure and their roles in maintaining master data libraries. Exact directory structure is subject to change.*

Location	Description	Custodian
M:\GIS_Data	Storage of GIS data, metadata, and Theme Manager information. Best access to this information is through the Theme Manager in ArcGIS.	PWRO-HNL GIS Specialist
I:\libraries\archived_projects	Electronic archive of "packaged" products. Includes administrative records as well as raw and certified versions of data. Extraneous files have been deleted.	PACN Data Manager
I:\libraries\ \reports \photos \data products	Storage of final products with other like final products, such as final reports with all other final reports.	PACN Data Manager
Redistribution to Parks M:\GIS-Data I:\libraries	Data stored in above directories are redistributed to the parks for more direct and faster access.	PWRO-HNL GIS Specialist PACN Data Manager PWRO-HNL IT Specialists
NR-GIS Data Store \GIS_Data \Bibliography_PDFs \Biodiversity_Data	Copies of a project's final products will also reside on the WASO Data Stores	WASO

4.2.3. Application Development

Applications are the "front end" or interface with which users interact with data. The "back end", consisting of the actual data tables, may be stored on a database server or may be distributed as a stand-alone MS Access database.

PACN will use one or both of these database sharing methods:

1. Desktop Microsoft Access front end database – where the desktop front end database provides the interface to forms, reports or standard queries, and the back end data is in another Microsoft Access database file that may reside on a server or also be locally stored on the desktop computer
2. SQL Server Application – where the back end data tables are stored on a database server using SQL Server softare, allowing a variety of user interface (forms, queries, and reports) formats to be employed via wide-area network (WAN) connections, and serving as a robust and secure platform from which to update master datasets on WASO servers and generate data summary exports in various formats for distribution and PACN website postings

Stand-alone desktop MS Access databases require strict procedures for data check out and for data uploading in order to maintain the integrity of the master database. These desktop databases are used by service-wide applications such as NPSpecies and NPSTORET. User interface applications drawing on SQL Server data tables allow user access to the most current, centrally stored information, and allow for data access via network (WAN) connections and internet connections. Both MS Access and SQL Server allow for integration of data into the GIS environment. The PACN system architecture employs a geodatabase design structure to serve spatial data to network users and to establish permission controls for user edits to the data.

4.2.4. Internet Servers

Applications, products, and links to the clearinghouses discussed in this chapter will be available on the PACN internet website. Information in draft or deemed sensitive will be made available on the PACN intranet website as applicable. PACN internet and intranet webpages are hosted on the WASO servers.

4.3. LAN and WAN Infrastructure

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

4.4. IT Network Security

Information Technology security is managed by park, regional, and national IT Specialists. PACN staff will comply with all required security training and procedures as advised by the PWRO-HNL IT Specialists.

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

5. DATABASE DESIGN OVERVIEW

5.1. NPS and Program Standards

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

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

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

Each database must ultimately meet the needs of the project leaders and network staff. Considerations for these needs may include interactions with other agencies and ease of use, maintenance, integration, and customization. PACN models and documents databases as described in this Data Management Plan, uses the above recommendations in database design as guidelines, and standardizes attributes across databases where feasible.

5.2. Introduction to Data Models

Communication is a vital part of developing a suitable data design for individual projects. PACN strongly encourages collaborative development of data models by project leaders and PACN data management staff. Data models combine diagrams with associated descriptions. Data modeling is completed in three stages: conceptual, logical, and physical.

5.2.1. Conceptual Data Models

Conceptual data models (CDM) are constructed to graphically portray the processes specifically related to the implementation phase of a project – especially those that involve data acquisition, processing, QA/QC, and data reduction (Figure 5.1). These conceptual models are software-independent and free of database details, and instead focus upon capturing all of the information needed to accurately express the project data design. Conceptual data models (CDMs) are created as the precursor to logical data models (LDMs). For less complex projects, a CDM may sufficiently convey data entities and relationships, negating the needs for a LDM.

Conceptual Data Models should contain the following:

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

5.2.2. Logical Data Models

A logical data model (LDM) is an abstract representation of a set of data entities and their relationships, and should include the key attributes for each entity (Figure 5.2). The logical data model is intended to facilitate analysis of the function of the data design, and is not intended to be a full representation of the physical database. It is produced early in system design, and serves as a precursor to the physical data model that documents the actual implementation of the database.

Logical data models are made up of five main components:

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

5.2.3. Physical Data Models

The physical data model (PDM) conveys the actual database design, depicting data tables, fields and definitions, and relationships between tables (Figure 5.3). Though the logical and physical data models are similar, the logical data model only provides enough detail to communicate the information to be stored in the database. The physical data model provides very specific details and definitions, such as primary keys and field types.

5.2.4. Developing the Models

Four elements are required to develop the data models:

1. People – Communication among integral project staff and PACN staff is required for developing models that effectively convey data collection processes and database design. Discussions may begin between a few, with review from a larger group or start with a large group working towards the specifics with a smaller group. Affected people may include the network coordinator, scientists, partners, field crews, biometricians, and data managers
2. Protocols – The protocols for the vital sign provide the greatest substance to the models. The protocols provide the goals, objectives, methods, standards, analysis, and reporting
3. Reference materials – Reference materials such as field forms, drawings, mock-up reports, references to classifications to be used play a significant role in the data models
4. Frequent interactions -- Frequent discussions between the project leader, data manager, and others integral project staff are needed to develop successful data models. Detailed review of the protocols and reference materials allows accurate descriptions of the data entities, relationships, and flow of information to be developed

Data models are not flexible by themselves. It is how these techniques are applied that count. Not everything can be thought of from the start. To support change, data modeling should be iterative and interactive. A variety of software products can meet the data modeling needs of a project. PACN makes use of Smart Draw and Microsoft Powerpoint software for project data modeling.

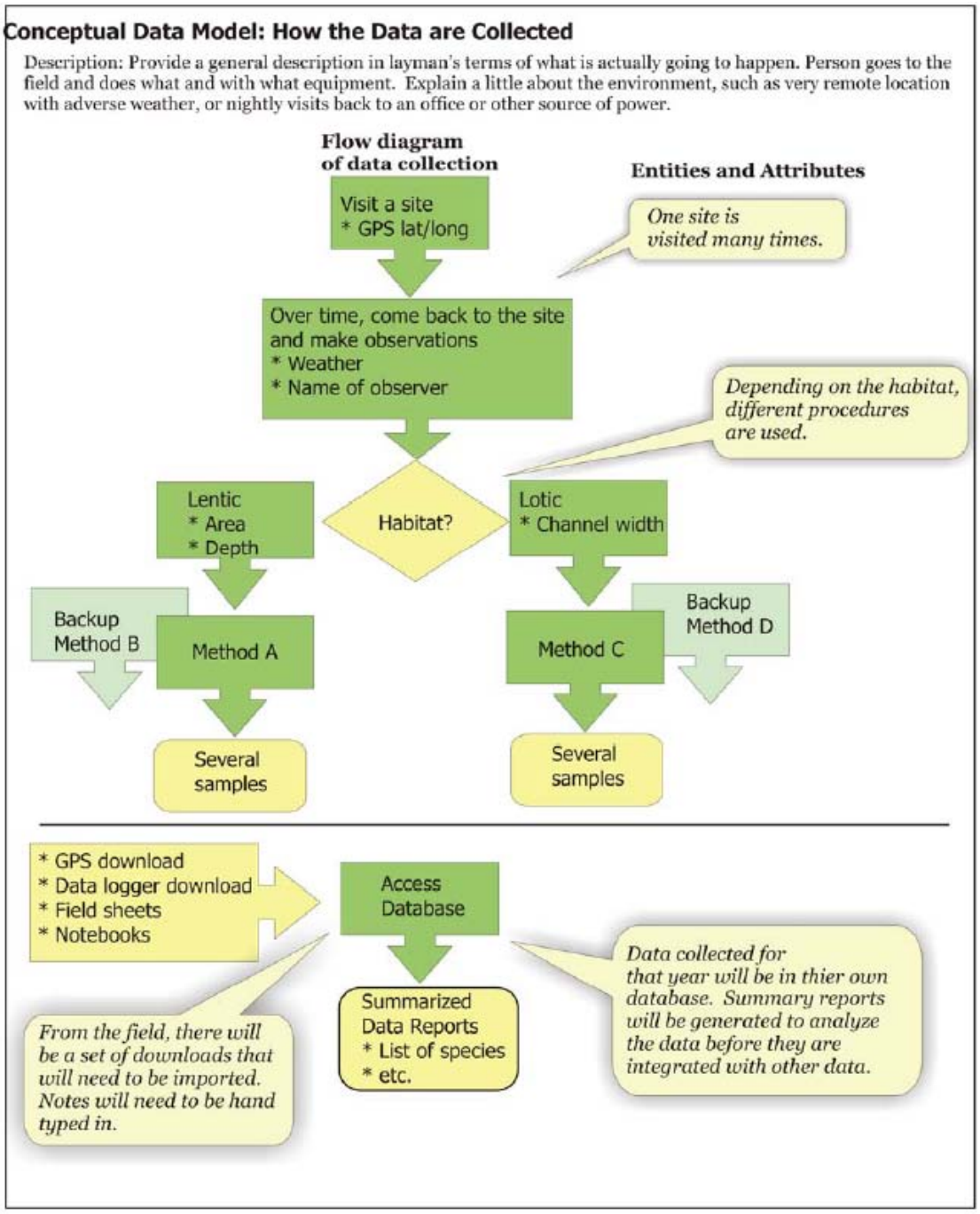


Figure 5.1. Schematic of Conceptual Data Model.

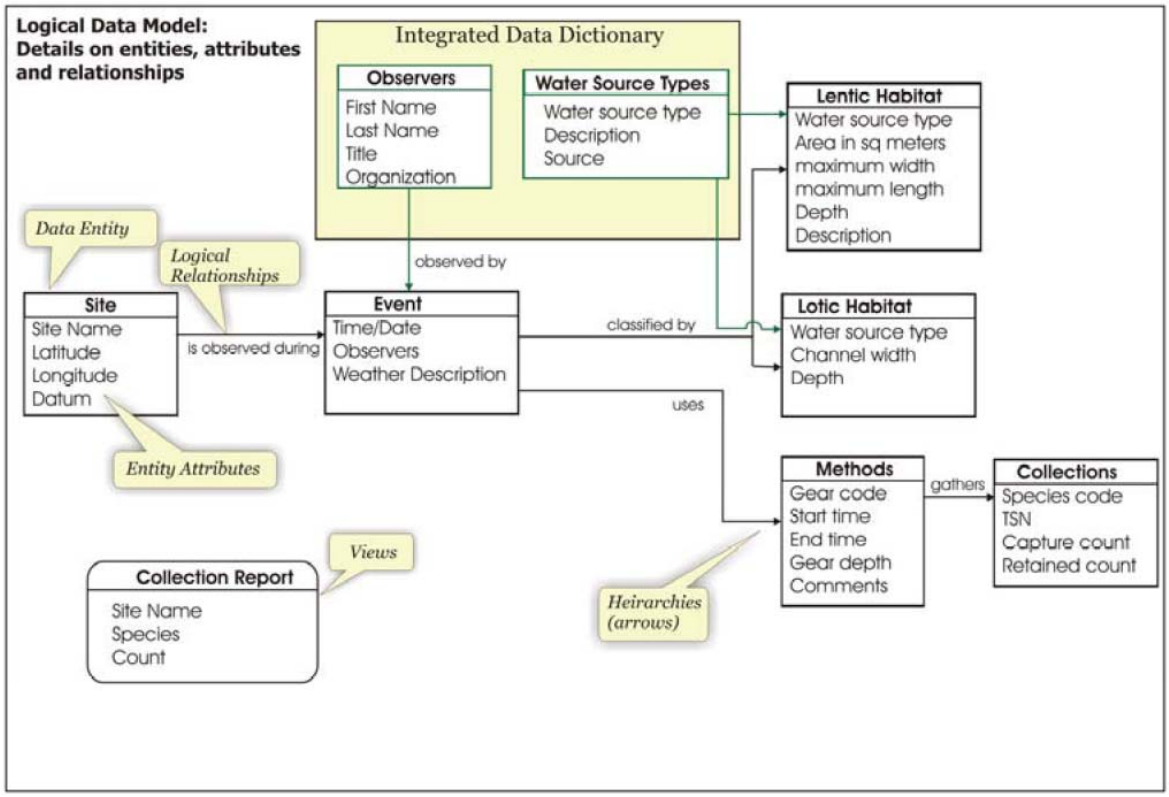


Figure 5.2. Schematic of Logical Data Model.

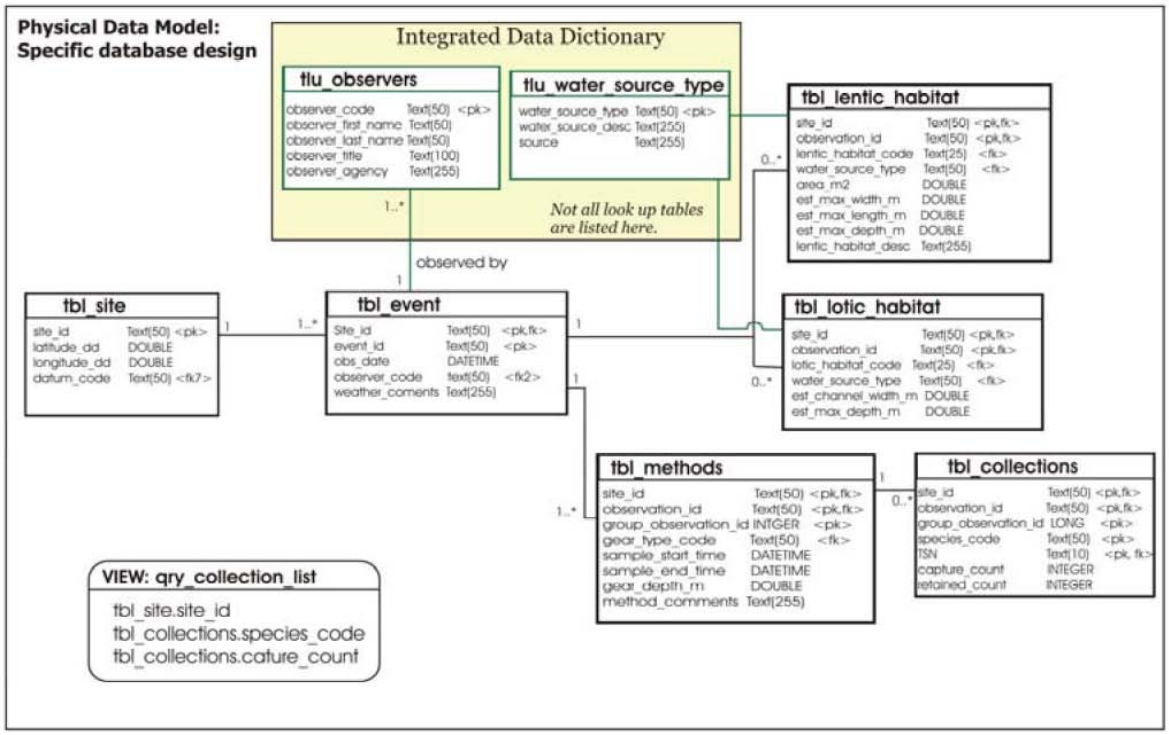


Figure 5.3. Schematic of Physical Data Model.

5.3. Fundamental Database Structure

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

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

The NRDT has been extensively reviewed by the I&M Program data managers and has been applied to some of the biological inventory projects in the Program. It has not, however, been applied extensively to monitoring. Although the fundamentals of this database structure are applied to PACN databases, specific mandatory tables may need to be reconsidered as the sampling methodology and data analysis needs of the PACN vital sign monitoring evolve.

5.4. Integrating Databases

PACN presently relies on modular, stand-alone project databases developed using MS Access software. PACN plans to develop a more integrated and centralized database system using SQL Server software, which will continue to allow for modular, project-specific database applications to be distributed for cyclical data collection and entry. This more centralized approach will provide greater flexibility with user access to data (e.g., stand-alone front end applications as well as network or internet connections) and a more robust and secure platform for storing PACN master data, for generating data summary exports, and for updating master datasets on WASO servers. This more centralized infrastructure will also support PACN plans to develop a master look-up table database, or Integrated Data Dictionary (also referred to as a Buffet of Fields and Tables by WASO), as a means of maximizing consistency across projects and across disciplines regarding data definitions and values. Likely standard attributes and values to be contained in this master look-up table database include park codes, watershed codes, species identifiers, and vegetation or land type classifications.

6. DATA ACQUISITION AND PROCESSING

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

Steps for data acquisition are as follows:

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

In the process of developing the PACN Monitoring Plan, the network data discovery efforts focused on park data mining in order to describe and document the current state of knowledge regarding park resources and past inventory, monitoring, and research work. Over a three year period (2002-2005), through work by cooperators and PACN staff, this park data mining was used to populate service-wide I&M databases (NatureBib, NPSpecies, and Dataset Catalog) with records pertinent to the bibliographic, species, and dataset resources of the PACN parks. As the PACN implements the Monitoring Plan, data acquisition and processing will concentrate on the needs of monitoring and inventory projects, and this chapter focuses on this project-oriented perspective. Chapter 10 presents a list of resources that serve both as repositories for acquired data and as dissemination tools (see table 10.1).

6.1. National Standards

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

6.2. PACN Network Standards

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

6.3. Data Sources

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

1. Programmatic Data – any data produced from projects that are initiated (funded) by the I&M Program or projects that in some way involve the I&M Program

2. Non-programmatic Data – includes data collected from NPS sources and data produced by external non-NPS sources
    - Non-programmatic NPS Data – any data produced by the NPS that did not involve the inventory and monitoring program, such as park visitor use information
    - Non-programmatic External Data – any data produced by agencies or institutions other than the National Park Service, such as the Pacific Island Ecosystems Research Center of the USGS and the state of Hawaii's Invasive Species Council

PACN vital sign monitoring will likely use a combination of both programmatic and non-programmatic data to meet its programmatic goals.

6.4. Data Discovery

Data discovery or data mining is the process of searching for existing data/information that may be useful to the I&M Program mission and that is related to the natural resources of the network parks. This is a continual process that begins with the collection of background information and data that assist in the development of project methodologies and designs. The process involves reviewing many different sources for varying types of information. Data mining is an important part of any I&M project, and some vital sign monitoring projects may depend largely on data collected by other agencies or institutions and harvested via data mining endeavors. Many of the data sources listed below are accessible via the internet, but some require visiting local archives, research or academic institutions, museums, or local parks in order to search reference material. While some of the I&M Program databases (e.g., NatureBib and NPSpecies) serve as a repository for results from data mining efforts, they also serve an important data discovery role for NPS staff, the research community, and the public.

Bibliographic/Literature

- National NPS Databases (e.g. NatureBib)
- Online literature databases (e.g. First Search or Biosis)
- Park archives through ANCS+

Geographic Data

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

Biological/Natural Resources Data

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

Relevant information collected during a data discovery process is maintained at PACN either electronically or in hard copy format depending on how it was collected. Any data collected

during data discovery should be accompanied by as much documentation (metadata) as possible. For datasets, both tabular and spatial, to be truly useful, they must be documented with metadata. PACN uses the NPS Metadata Tools and Editor application for creating and maintaining metadata records, and ensures that metadata is FGDC compliant whenever possible.

6.5. Data Harvesting

The process of harvesting data from other sources should be standardized as much as possible. Considerations for harvesting are as follows:

- Does the data source organization know of the I&M Program and its data needs?
- Does a Memorandum of Understanding (MOU) need to be in place?
- Are there copyright or ownership issues?
- What is the contingency if this data source is no longer available?
- Can downloads or requests for data be consistently exported/imported and scheduled? And are complete data downloads needed, or will links to the data source suffice?
- Will the downloaded data be stored and integrated into the I&M Program? If so, by what means and with what schedule for data updates?
- How should errors be addressed?
- Is the data source organization interested in integrating I&M Program data, where appropriate?
- Is the documentation adequate?

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

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

6.6. Data Collection for Projects

Biological inventories and monitoring projects are the most common examples of projects conducted by the PACN I&M Program. The project leader is responsible for ensuring that data collection, data entry, verification, validation, storage, and archiving are consistent with PACN standards. Project leaders will make use of PACN general standard operating procedures (SOPs) and guidance documents as applicable. In addition, protocol-specific SOPs may be developed to address data collection methods and procedures particular to a given protocol. The data manager will work closely with the principal investigator and network staff to develop protocol-specific SOPs for the collection, storage, and maintenance of project data. This may range from detailing the proper usage of data entry forms or databases to outlining calibration procedures for automated data loggers. See appendices for examples of PACN data management SOPs.

Some of the tools available for field data collection are listed below. This list serves as a guide for consideration as the vital sign protocols are developed. Details on how these tools may apply to protocol-specific SOPs will be contained in the individual protocol documentation.

- **Field Forms** – are the most common method of recording field data. Use a formatted, project-specific data sheet as opposed to a field notebook. Field notebooks are important for entering additional notes and observations. It is recommended to use acid-free paper to prevent fading and subsequent data loss. Some circumstances may warrant the use of paper and writing implements that can withstand moisture, dust, and other extreme environmental conditions (e.g., "Rite in the rain" paper). Standardized data sheets that identify the pieces of information to be recorded and forms that reflect the design of the computer data entry interface will help ensure that all relevant information is recorded and subsequent data entry errors are minimized. Data sheets should contain as much basic preprinted project information as possible and sufficient space for recording relevant metadata such as date, collectors, weather conditions, etc. They should clearly specify all required information, using examples where needed to ensure that the proper data are recorded. Data recorders should adhere to the following guidelines:

    - All information added to the data sheet must be printed and clearly legible.
    - If alterations to the information are necessary, the original information should be crossed out with a single line and the new information written next to the original entry. Information should never be erased and old information should not be overwritten.

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

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

- **Cameras** – Photographs provide an excellent visual record of field visits. Cameras are useful for capturing photo point records of long-term study sites. They also serve well for automated data collection by remotely recording information using web cams or trip cameras. Generally, PACN recommends using digital cameras, but some projects may have sound reasons for using analog cameras.

- **Field Computers** – increase data collection and data entry efficiency by eliminating the need for paper field forms. Data can be downloaded directly from the field computers to the office desktops thereby eliminating the data entry step and reducing the likelihood of error because QA/QC checks can be built into the field database. Field computers can be inefficient if copious amounts of notes or comments need to be recorded in the field.

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

  - **Palm-top computers (Personal Digital Assistants)** – the small size and relative low cost of these devices make them attractive options for entering field data. Good for small field projects but not powerful enough for large data intensive field projects. PDAs can be ruggedized fairly easily and at a relatively low cost. Most run either Windows CE or Palm operating systems that may require additional processing/programming to transfer/create the database structure in the field units.
  - **Tablet PCs** – same properties as most laptops and provide the user with the convenience of a touch screen interface. They are bulkier, more expensive and harder to ruggedize than PDAs but are more powerful as well. Good for field projects that are very data intensive. Because these units run Windows XP (Tablet Edition) the project database can be directly transferred from desktop units to field units without additional programming steps.

- **Automated Data Loggers** – mainly used to collect ambient information such as air or water quality information. These devices can be calibrated and programmed to automatically record data and store them for later download directly to a computer, thereby eliminating the possibility for manual data entry errors. Must be properly calibrated and maintained so field crews must receive proper training and SOPs outlining the calibration/unit maintenance procedures. SOPs are currently in development and will be distributed to all field crews using these devices.

  - **Permanently deployed devices** – often cost prohibitive. Data from these devices must be retrieved and batteries changed on a regular basis. These intervals should be defined in the protocol.
  - **Portable hand-held devices** – deployed for sampling only during site visits. Generally less expensive than units that are permanently deployed in the field. Prior to and following field visits device components should be inventoried to ensure that all necessary equipment is accounted for.

- **GPS Units** – Two types of GPS units are often used during field work in PACN parks to collect location information.

- **Handheld Units (e.g., Garmin)** – good for collecting general position information. Not recommended for obtaining high accuracy location information.
- **Trimble GPS Receivers** – good for collecting highly accurate (submeter) location information.

6.7. Data Collection for Remote Sensing

Remote sensing technologies can be a powerful tool for characterizing and analyzing landscape data, as well as readily capturing data within areas of low accessibility. Considerations for selecting remote sensing imagery are as follows:

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

Each remote sensing product is unique. It is imperative the user fully understands the product he or she is using. Products should be accompanied with well documented metadata. Imagery not rectified cannot be used for measuring distance or area, but may provide a low cost and timely overview of the landscape. Any protocols involving remote sensing should involve consultation with a professional remote-sensing specialist, and should consider the trade-off between accuracy and costs among different imagery sources (Table 6.1).

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

The PACN's network GIS currently holds good coverage for all PACN parks in base cartographic layers, digital orthophoto quadrangles, and natural resource theme layers, as well as satellite imagery coverage for most of the parks (Appendix E). As individual vital sign monitoring protocols identify additional remote-sensing imagery needs, PACN will pursue cost-efficient options for acquisition of imagery in suitable formats. PACN uses ESRI ArcCatalog software to manage remote sensing image metadata, and plans to develop a Microsoft Access application to allow users to browse through a catalog of PACN imagery.

Table 6.1. *Satellite resolution, swath width, area, cost, and history of various remote-sensing imagery sources.*

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

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

6.8. Changes to Data Collection Procedures

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

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

6.9. Data Compilation, Processing, and Integration

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

After compilation and QA/QC is completed for data from a single field season or project milestone period, these data should be integrated with master datasets. Some extraction of information may be required. For example, species information may need to be extracted from various project databases for the purpose of updating the master NPSpecies database.

7. QUALITY ASSURANCE AND QUALTIY CONTROL

The success of I&M networks is dependent on the quality of the data they collect, manage, and disseminate. Analyses performed to detect ecological trends or patterns require data that are recorded properly and have acceptable precision and minimal bias. Poor quality data can limit detection of subtle changes in ecosystem patterns and processes, can lead to incorrect interpretations and conclusions, and could greatly compromise the credibility of the I&M Program. To ensure that PACN produces and maintains data of the highest possible quality, procedures have been established to identify and minimize errors at each project stage associated with the data life cycle (Figure 7.1). Data quality related to sampling design is covered in Chapter 4 of the PACN Monitoring Plan and in individual protocols, and is not discussed here.

7.1. NPS Standards

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

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

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

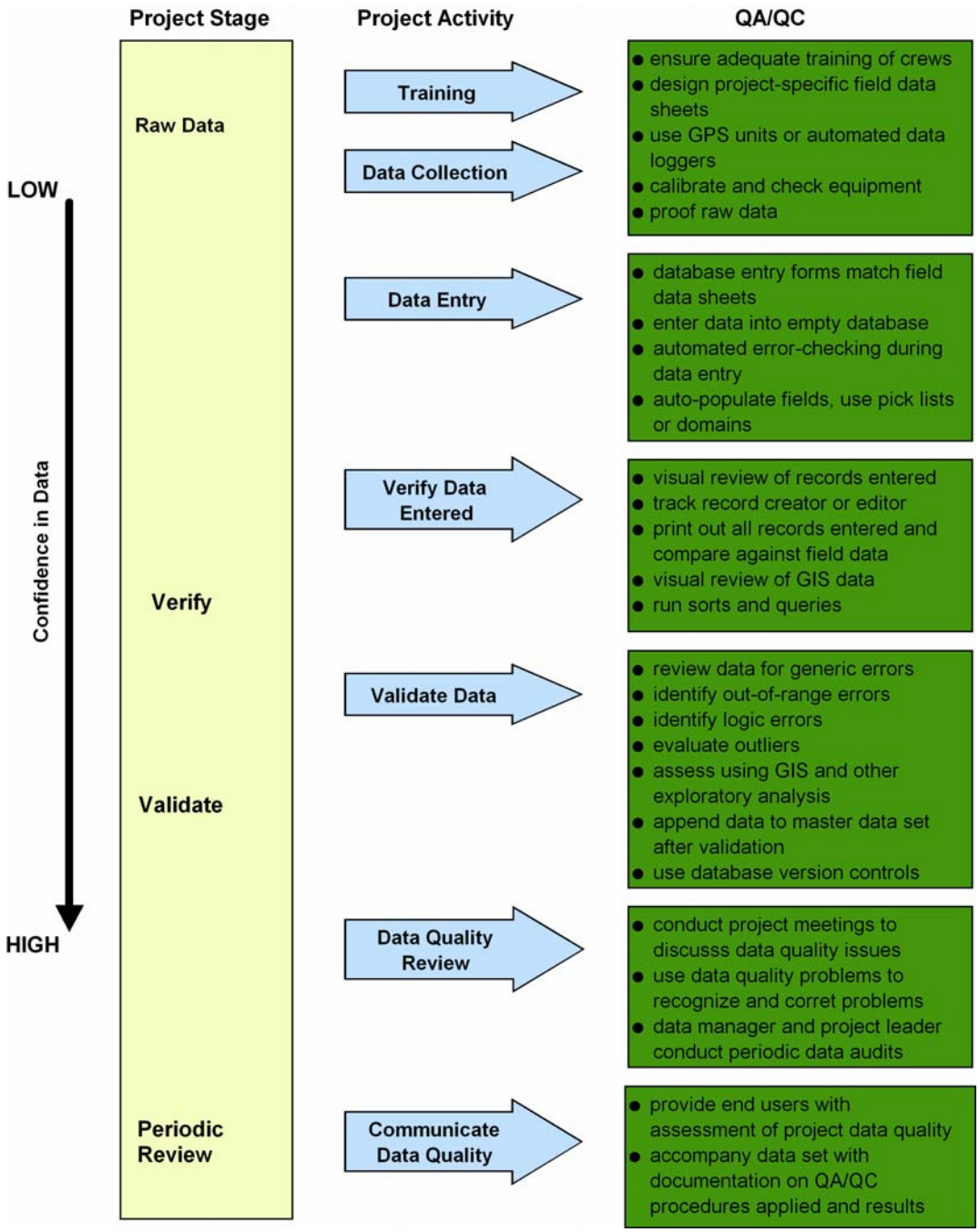


Figure 7.1. Schematic of the Quality Assurance/Quality Control procedures to be carried out during the project stages associated with the typical data life cycle.

7.2. Quality Assurance and Quality Control Mechanisms

Palmer (2003) defines quality assurance (QA) as "an integrated system of management activities involving planning, implementation, documentation, assessment, reporting, and quality improvement to ensure that a process, item, or service is of the type and quality needed and expected by the consumer." He defines quality control (QC) as "a system of technical activities to measure the attributes and performance of a process, item, or service relative to defined

standards." Quality assurance procedures maintain quality throughout all stages of data development. Quality control procedures monitor or evaluate the resulting data products.

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

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

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

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

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

7.3. Roles and Responsibilities

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

The data manager is responsible for:

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

Project leaders must:

- be aware of quality protocols and convey their importance to technicians and field crews
- ensure compliance with the protocols

- validate data after the verification process is complete
- review all final reports and information products

Project technicians must:

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

7.4. Quality Assurance/Quality Control Goals and Objectives

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

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

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

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

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

7.5. Data Collection Quality Assurance/Quality Control

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

Before the data collection phase of a project begins, the project leader and data manager determine data collection and storage protocols. Field sheets and field data recording procedures must be documented in the protocol SOPs and reviewed by the data manager. The project leader, in turn, ensures that field crews understand the procedures and closely follow them in the field. The project leader is responsible for training of field crews, with the data manager providing training assistance as needed and as time permits. Field technicians are responsible for proofing

raw data forms in the field, ensuring their legibility, and verifying and explaining any unusual entries. They are expected to understand the data collection forms, know how to take measurements, and follow all procedures established in the protocol and SOPs.

7.5.1. Methods for Reducing Collection Errors

Ensure that field crews receive proper training

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

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

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

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

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

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

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

Use an electronic device for data collection whenever possible

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

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

Use automated data loggers where appropriate

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

Consider calibration, maintenance, and training requirements of field equipment

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

Be organized and keep a log

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

Perform quantitative assessments of data quality

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

7.6. Data Entry

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

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

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

7.6.1. Methods for Reducing Data Entry Errors

Enter or download data in a timely manner

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

Design efficient data entry forms

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

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

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

Track record creation and edit details in the database

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

Build automated error checking features into the database

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

- Auto-filled fields. Whenever possible, the data in a field should be autofilled by the computer. For example, if a location ID is comprised of a park code, project code, and a unique number, those elements are automatically inserted into the location ID field, ensuring that the record always contains a unique identifier.
- Range limits. Where the appropriate values for a particular field span a finite range, the data entry program can check the entered value against the specified minimum and maximum values for that parameter. When a value is outside the accepted range, a warning message appears and asks the user to reenter a valid value. For some fields, values outside a specified ‘normal’ range may be acceptable. In this case, the warning message asks the user to verify the entry before continuing.
- Pick lists. The data entry application may also use pop-up pick lists for standardized text items where spelling errors can occur. For example, rather than typing in a species code or name (where a misspelling generates a new species in the database), the code or name is

selected from a list of valid species codes or scientific names and automatically entered into the species field. A pick list may also be used when only certain entries are acceptable. Lists are not appropriate for all written fields but should be used when appropriate.

- Unique constraints. Duplicate and incorrect data entry can often be caught with the application of unique constraints on data entry fields. These constraints are particularly useful when importing data from other applications.

Provide a clean, organized work environment

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

If possible, use two data entry technicians for data entry

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

7.7. Verification and Validation Procedures

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

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

7.7.1. Methods for Data Verification

Data verification immediately follows data entry and involves checking the accuracy of the computerized records against the original source (usually hard copy field records), and identifying and correcting any errors. PACN procedures are to verify all records entered against their original source. When the computerized data are verified as accurately reflecting the original field data, the paper forms are set aside for use later in the data validation process. The following five verification methods will be used by PACN for all project data:

Visual review at data entry

The data entry technician verifies each record after input and immediately corrects any errors. This method is the least complicated since it requires no additional personnel or software. Its reliability depends entirely upon the person keying data and thus, is probably the least reliable data verification method.

Visual review after data entry

All records entered during a data entry session are printed in a format that closely matches the original data source. Each data element on the printout is compared with the original values from the hard copy, preferably by a second person who did not perform the data entry. Errors are clearly marked and corrected in the database as soon after data entry as possible.

Sorting and summary queries

Each project will have queries that can be run to detect broad errors such as inconsistent, duplicate, omitted, or unlinked records.

Visual review of spatial data

Any spatial data that are collected as part of the project will be converted to GIS and visually inspected for accuracy (e.g., points located outside park boundaries, upland locations occurring near sea level).

Duplicate data entry or verification of record subset

The data entry person completes all data input, as normal. Random records are selected (every **n**th record) and entered into an empty replica of the permanent database, preferably by someone other than the person keying the permanent data. A query is run to automatically compare the duplicate records from the two datasets and report on any mismatches of data. These disparities are manually reviewed and corrected if necessary. This method involves the overhead of retyping the selected records, as well as the creation of a comparison query (which requires additional effort, but is not time-consuming). This method becomes increasingly successful as the value of **n** decreases. An alternative is to retyping data into the database is to have a second observer read through the randomly selected records, comparing them against the original source (e.g., hardcopy datasheets) and noting errors via a customized database feature.

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

7.7.2. Methods for Data Validation

Although data may have been correctly transcribed from original field notes or forms, they still might be inaccurate or illogical. For example, values may be outside the physical range of possible values, or may fall outside the logical range of reasonable values. The process of reviewing computerized data for range and logic errors is called *validation*, and it can accompany data verification *only* if the reviewer has comprehensive knowledge about the data. Validation is a separate operation carried out *after* verification by a project specialist who can identify generic and specific errors in particular data types.

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

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

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

Data entry application programming

Certain components of data validation are built into data entry forms. The simplest validation during data entry is range checking, such as ensuring that a user attempting to enter a pH of 20.0 gets a warning and the opportunity to enter a correct value between 1.0 and 14.0 (or within a narrow range appropriate to the study area). Not all fields, however, have appropriate ranges that are known in advance, so knowledge of what are reasonable data and a separate, interactive validation stage is important.

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

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

Outlier Detection

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

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

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

Other exploratory data analyses

Palmer and Landis (2002) suggest that in some cases, calculations for assessments of precision, bias, representativeness, completeness, and comparability may be applicable and that for certain types of measurements, evaluation of a detection limit may also be warranted. Normal probability plots, Grubb's test, and simple and multiple linear regression techniques may also be used (Edwards 2000).

7.8. Version Control

Version control is the process of managing copies of changing files over the course of a project. Change includes any alteration to the structure or content of the files. Good version control practices allow for full recovery of a dataset as it existed prior to changes being made. Before making any major changes to a file, a copy of the file should be saved with a unique file name. This process is particularly important when appending newly-validated data to a master dataset, or when upgrading a database to a new version. In addition, proper controls and communication are required to ensure that only the most current version is used for data entry or analysis.

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

These general version control guidelines apply to files maintained on the PACN file server. They do not apply to enterprise-level server applications such as the I&M Program's service-wide Oracle databases; updates to these master databases are accomplished either through online data entry/edit or through file submission to I&M Program WASO staff. General file naming guidelines are as follows:

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

Where:
<project_file> = the name of the established base file name
<QA/QC_status> = the status of the information. For example, " _RAW_"; "_VERIF_"; "_VALID_"
<date> = The date of the file, as YYYYMMDD.
<ext> = the file extension, such as .mdb, .xls

7.9. Data Quality Review and Communication

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

7.9.1. Value of Feedback from QA/QC Procedures

Quality assurance procedures may need revision to improve quality levels if verification and validation processes reveal an unacceptable level of data quality. Quality checks should not be performed with the sole objective of eliminating errors, as the results may also prove useful in improving the overall process. For example, if the month and day are repeatedly reversed in a date field, the data entry technicians may require retraining about the month/day entry order. If retraining is unsuccessful in reducing the error's occurrence, the computer program may need to be rewritten so that month and day are entered separately, field length limits are enforced, or a pick list is created. In this manner, the validation process will serve as a means of improving quality.

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

7.9.2. Monitoring Conformance to Plans and Standards

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

Audits include verification of the following:

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

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

7.9.3. Communicating Data Quality

The PACN will use data documentation and metadata to notify end users, project managers, and network management of data quality. A descriptive document for each dataset/database will provide information on the specific QA/QC procedures applied and the results of the review. Descriptive documents or formal FGDC-compliant metadata will document quality for spatial and non-spatial data files posted on the Internet.

8. DATA DOCUMENTATION

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

8.1. NPS Standards

Department of Interior direction for documentation at the departmental and NPS levels is as follows:

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

8.2. PACN Network Standards

PACN standards for documentation are as follows:

- Project documentation will be tracked using the project tracking database (see Appendix B)
- Tabular and spatial data documentation will be completed using the full standards set by the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM). PACN uses the NPS Metadata Tools and Editor for documentation of data.
- Vital sign protocol documentation will be completed using
    - "Guidelines for long-term monitoring protocols" (Oakley et al. 2003)
    - Master Versioning Table, and additional vital sign documentation described in this chapter

8.3. Project Documentation

Projects should be well organized in a project directory structure or an existing repository. In the initial project planning stage, the project leader will work with the network coordinator and the network data manager to establish a list of products to be delivered by the project. A 'PACN Project Deliverables' guidance document (see Appendix A) provides project leaders with general guidance on organizing project data and information. And the project study plan will identify specific product deliverables.

The PACN uses a project tracking database (see Appendix B) to organize basic documentation on each project. The project tracking database is designed to be flexible for the I&M Program operations, and tracks the following information:

- Contact information
- Dates
- Links to Program goals, associated parks, vital signs, permits, investigators annual reports

- Abstracts
- Status, repository, and notes on deliverables

Reports can be produced from this database to:
- document the project
- list the status of deliverables
- transfer necessary project information to park curators for archiving purposes

8.4. Spatial Data Documentation

Spatial data will use the FGDC, Content Standard for Digital Geospatial Metadata (CSDGM). Metadata records created by the networks are integrated into the National Park Service, NR-GIS Data Store. Associated datasets will be stored in the NR-GIS Data Store. The locations of the associated datasets are directly linked in the metadata record. Metadata record management, accomplished using the ArcCatalog Metadata Extension within the ArcGIS software, will be coordinated with the PWRO-HNL GIS Specialist to ensure quality and consistency. For geodatabases, metadata is created for individual feature datasets as well as for the Microsoft Access .mdb file as a whole. Geodatabase metadata records will include standardized language that will assist end users in using ArcObjects to generate diagrams of the various geodatabase elements and their inter-relationships.

8.5. Tabular Database Documentation

Tabular databases will use the minimum FGDC CSDGM standard for documentation. This allows data discovery through the use of FGDC clearinghouses. It is difficult to document a relational database in adequate detail using this standard alone. Therefore, for each network project, PACN data management staff will generate a database document that will include:
- Revisions to the database
- Overview of the use and purpose of the data
- Illustration of the entity diagrams
- Data dictionary of tables, attributes, and relationships
- Explanation of queries, forms and reports

8.6. Vital Signs Protocol Documentation

8.6.1. Master Version Table

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

The Northern Colorado Plateau Network (NCPN) has suggested using a database to track the narrative and SOP version numbers in a Master Version Table (MVT). The MVT contains a

Version Key Number that designates the narrative and SOP versions that are in use at a specific time (Table 8.1). Every protocol contains an SOP entitled "Revising the Protocol Narrative and SOPs." The MVT is contained in this SOP and is required to be updated when any protocol revisions are made. The protocol narrative, SOPs, and data will not be distributed independently of the MVT. The PACN may adopt this MVT strategy for documenting revisions to protocols and SOPs. Incorporation into individual protocol databases, so that data records can be associated with a specific SOP version, will require active database links that assign the SOP version number(s) to each sampling event.

Table 8.1. *Example of the Master Version Table used to track changes in a vital sign protocol narrative and/or standard operating procedures (SOP).*

Version Key #	Version Key Date	Narrative	SOP #1	SOP #2	SOP #3	SOP #5	SOP #6
VK1	12/15/2004	1.00	1.00	1.00	1.00	1.00	1.00
VK2	5/10/2005	1.00	1.01	1.01	1.00	1.01	1.00
VK3	11/18/2005	1.00	1.01	1.01	1.01	1.02	1.00

8.6.2. Additional Vital Sign Metadata

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

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

9. DATA ANALYSIS AND REPORTING

Providing meaningful results from data summary and analysis is a cornerstone of the I&M Program and characterizes the program's data management mission to provide useful information for managers and scientists. Chapter 7 (Data Analysis and Reporting) of the PACN Monitoring Plan contains the background and overall approach to data analysis and reporting for the network. The associated data management objective is to provide valid data in formats that support scheduled and ad hoc display, query, analysis, summary, and reporting of data. Routine and scheduled data summary and analysis requirements and procedures are identified in each vital sign monitoring protocol. This chapter discusses data management activities related to using GIS and database application software for data summary and analysis, and to prepare data for analysis using statistical software applications.

9.1. Periodic and Annual Reporting

The data manager will work with the project and network statisticians, ecologists, and others to specify and design or adapt database objects, fields, and values to support the formats and functions necessary for analysis using statistical software applications such as SAS (commercial) and R (freeware). Some basic summary and reporting functions required by a vital sign monitoring protocol can be developed within Microsoft Access database applications where data are stored. Examples include descriptive statistics (mean, standard deviation, sample size). The statisticians, ecologists, and others will use existing and custom data conversion and export functions in MS Access to prepare datasets for import into other software applications. Spatial analysis and maps will be produced by network and/or park affiliate staff. Ad hoc queries and reports will be handled on a case by case basis due to their dynamic nature. PACN will enhance its web site over time to deliver reports and provide supplemental background data and information.

9.2. Long-term Trends and Analyses

Most long-term data analysis will involve statistical software applications, for which required data formats often involve arrays of binary or discrete values that represent one or more parameters. Data analysts and the network data manager will identify and develop the data conversion routines necessary to generate these formats for analysis.

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

10. DATA DISSEMINATION

PACN data management will ensure that:

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

The PACN will continue to provide links to PACN public data products via the PACN public website. This will be the network's portal for the most current data distribution as the Monitoring Program develops. Distribution instructions for each dataset will be provided in the respective metadata.

10.1. Mechanisms for Distribution

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

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

Access to PACN data products will be facilitated via a variety of means that allow users to browse, search and acquire network data and supporting documents (Table 10.1). These means include, but are not limited to:

- Links to public data products maintained on the PACN public website
- NR-GIS Metadata and Data Store. Distribution instructions for each dataset will be provided in the respective metadata
- Service-wide databases, such as NPSpecies, NatureBib, and NPSTORET
- Regional, Network, or Park data servers protected with read-only access
- External repositories such as the US Geological Survey, University of Hawaii, US Forest Service, Western Regional Climatic Center, and many others
- FTP sites, CDs or DVDs, as appropriate

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

Table 10.1. *Primary repositories for PACN information and associated specimens.*

Item	Repository
Reports (public) digital	NPS Focus, Data Store, PACN website
hard copy	Park and network libraries, park archives
bibliography	NatureBib
Network-generated digital datasets and data products (public, non-sensitive) • certified data and data products (including photos) • metadata	NR-GIS Data Store, PACN website, NPSpecies, NPSTORET
Network-generated digital datasets and data products (NPS staff, sensitive) • raw, validated, and analyzed data • metadata • submitted reports • digital photos • digital presentations	PACN intranet website; selected vital sign data may be housed externally with an established MOU.
Project product materials • specimen vouchers • photograph film	Park archives, Bishop Museum, or other curation facility (according to project protocol)
Project administrative records or miscellaneous items (hard copy)	PACN office

The national water quality database (NPSTORET) maintained by the NPS Water Resources Division provides a good example of how data collected by a network project is transferred to an on-line master database, from which data and summary products can be made widely available via a web-based clearinghouse service (Figure 10.1).

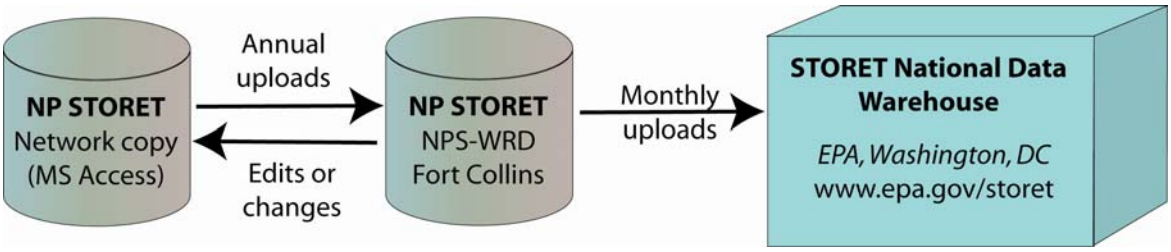


Figure 10.1. Schematic of data flow for transferring network water quality data to the national master database, and for network and public access to that master database via an on-line clearinghouse.

10.2. Ownership, FOIA, and Sensitive Data

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

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

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

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

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

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

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

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

Classification of sensitive data will be the responsibility of network staff, park superintendents, and project leaders. Network staff will classify sensitive data on a case-by-case, project-by-project basis and will work closely with project leaders to ensure that potentially sensitive park resources are identified, that information about these resources is tracked throughout the project, and that potentially sensitive information is removed from documents and products that will be released outside the network.

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

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

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

10.2.1. Access Restrictions on Sensitive Data

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

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

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

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

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

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

Thus, access to data on sensitive park resources can be limited to network staff or research partners. However, limits to how these data are subsequently released must also be clearly defined. It is crucial that the person uploading records to the online applications (repositories) is familiar with the procedures for identifying and entering protected information.

10.2.2. NPS Only versus Public

Only data subjected to complete quality control are released, unless necessary in response to a FOIA request. Products of incomplete, poor, or questionable quality (typically legacy data) may not be appropriate for the public and should be managed separately from acceptable quality. These lower quality products may be the only source of information on the natural history of the park, may have been the basis for early management decisions, and are still valuable in-house.

Provided these do not contain sensitive data as described in this chapter, these data may be released to the public upon specific FOIA request. They must be accompanied with qualifying documentation.

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

11. DIGITAL DATA MAINTENANCE, STORAGE, AND ARCHIVING

Effective long-term data maintenance depends on thoughtful and appropriate data documentation. An essential part of any archive is its accompanying explanatory materials (Olson and McCord 1998). The following sub-sections make reference to, and in some cases elaborate on, metadata standards and dataset documentation procedures that are more fully explained in Chapter 8 (Data Documentation).

Data, documents, and all other products resulting from projects and activities using PACN data are crucial pieces of information. To ensure high-quality long-term management and maintenance of this information, PACN will implement procedures to protect information over time. The goals of these procedures are to:

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

11.1. NPS and PACN Network Standards

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

11.2. Electronic Archiving and Storage

While project leaders and PACN network staff will maintain working files within their personal computer directories, the PACN file server will be used to ensure security of draft documents and draft data products and to ensure the longevity and availability of final products and long-term datasets. Access permissions to various levels of the PACN file server directory structure will be controlled by PACN staff, and folder space for working files will be provided to those project leaders who request it. Once project leaders assemble their data products into a stand alone "package," the package will be archived on the PACN server and individual products will be stored with other like products as appropriate. The general PACN file server directory structure accommodates GIS data, project archive packages, and final product libraries (see Table 4.1). The PACN strategies for local data backup and for integration of final data products into the I&M Program's service-wide databases and repositories are discussed in Chapter 4.

Short-term projects

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

1. Project archive on PACN file server, with all certified, integral files associated with the project "packaged" together (i.e., bundled or zipped together as a stand alone, easily distributed entity). Types of information to be included are listed in Chapter 1, Table 1.1

2. Off-line project archive, same information as above, which may include draft files. Stored on CD or DVD
3. Product library on PACN file server, where like products are stored with other products, such as a final report stored with other reports

Long-term projects

Milestone products, such as annual progress reports and annual datasets, will be electronically archived by the same methods applied to short-term projects. In addition, long-term datasets will be inherently archived and stored on the PACN file server and by the networks routine backup procedures.

11.3. Electronic Archiving Process and Workflow

Electronic files will be consolidated and packaged for electronic archival when a project is complete or, for long-term projects, when milestones are reached. PACN project leaders are responsible for packaging electronic files and data for the data manager, who is then responsible for accomplishing product archiving, integration, backup, and distribution (Figure 11.1). Archiving of hardcopy products, accomplished by park or regional cultural resources staff, is discussed in Chapter 12.

The project leader should prepare the electronic files as follows:

- Organize files according to digital file folder structure and strategy agreed upon with PACN staff
- Comply with file naming standards (see Appendix A)
- Comply with documentation standards, specified in protocol and/or study plan
- Include all related documents listed in the PACN Project Deliverables guidance document (see Appendix A)
- Make a clear distinction between draft and final files. Draft files will be included in project backup, but will be deleted for final online project archiving
- Make a clear distinction between public and sensitive information
- Ensure the project tracking database is complete (this will require working with the data manager)

The data manager will:

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

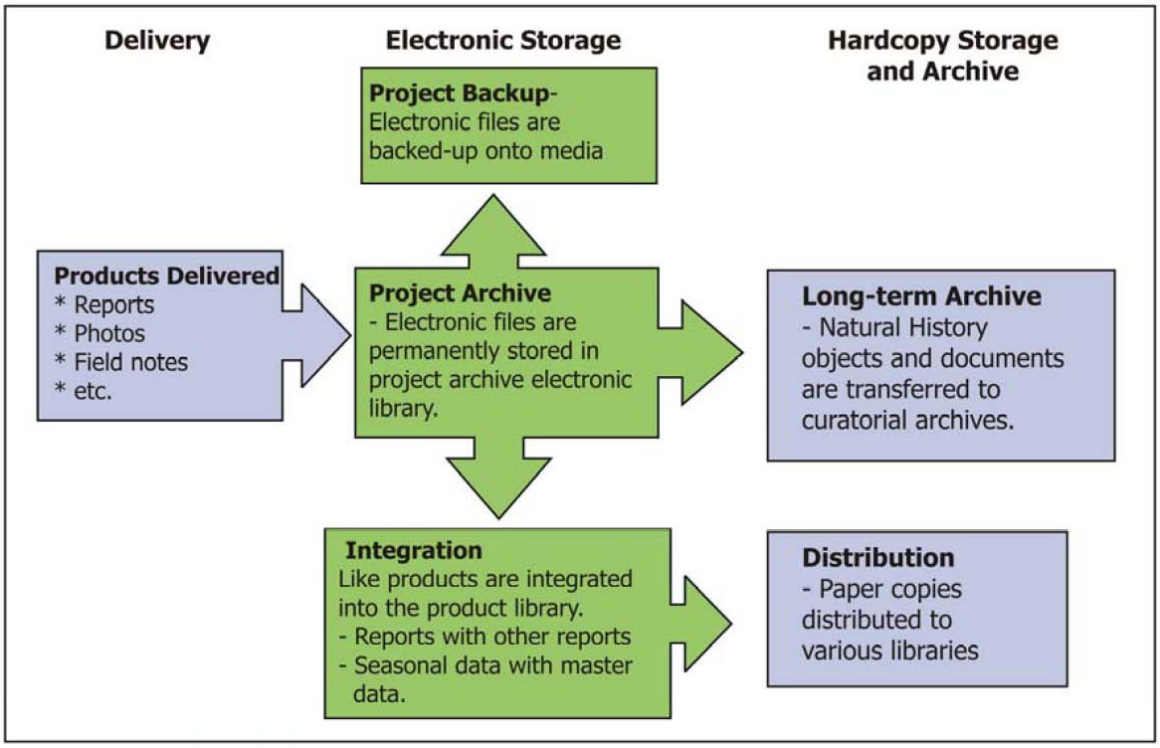


Figure 11.1. Schematic depicting workflow for storage and archiving of electronic and hardcopy products.

11.4. Quality Control for Converted Data

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

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

11.5. Digital Data Backups

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

The risk of data loss from catastrophic events must also be safeguarded against. Such catastrophic events include natural catastrophes (fire, volcanic eruption, hurricane), hardware failure, software failure or corruption, and security breaches or vandalism. Performing regular

data backups and arranging for off-site storage of backup sets are the most important safeguard measures against data loss.

The PACN file server is integrated with PWRO-HNL servers and takes advantage of the network-attached storage (NAS) backup strategy of the regional IT specialists, who use a level 5 redundant array of independent disks (RAID) implementation. In addition to this disk-to-disk backup strategy, PACN employs a tape backup device to create monthly backup sets of the PACN file server data for off-site storage.

Backup of data on desktop computers is the responsibility of the individual assigned to that computer. PACN encourages all network staff to regularly backup desktop computer data, and most staff use external hard drives for this purpose. All critical data stored on desktop computers should also be periodically copied to the PACN server (where individuals can create a personal folder for critical working files) so as to take advantage of the NAS backup strategy.

12. NON-DIGITAL DATA ARCHIVING AND RECORDS MANAGEMENT

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

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

12.1. NPS Standards

Direction for managing many of these materials (as well as digital materials) is provided in NPS Director's Order 19: Records Management (2001) and its appendix, NPS Records Disposition Schedule (NPS-19 Appendix B, revised 5- 2003). NPS-19 states that all records of natural and cultural resources and their management are considered mission-critical records, that is, necessary for fulfillment of the NPS mission. NPS-19 further states:

> *Mission critical records are permanent records that will eventually become archival records. They should receive the highest priority in records management activities and resources and should receive archival care as soon as practical in the life of the record.*

Section N of NPS-19 Appendix B, which provides guidelines on natural resource-related records (including, specifically, products of Inventory and Monitoring Programs), indicates that all natural resource records are considered "permanent" and need to be retained either in an appropriate park museum facility or at the National Archives. It also indicates that non-archival copies of natural resource-related materials are "...potentially important for the ongoing management of NPS resources" and should not, in any instance, be destroyed. The NPS Museum Handbook provides the overarching guidance for archival procedures. In particular, Part II, Appendix A: Mandates and Standards for NPS Museum Collections lists the cultural and natural history laws, regulations and conventions for NPS museum collections and should be reviewed prior to object collections.

12.2. Park Unit and PACN Network Standards

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

The PACN will adhere to the standards described above. All specimens collected as part of PACN projects are considered property of the respective park. Procedures for the collection and processing of specimens or samples will be dictated by curatorial staff at the park or regional level, as appropriate.

12.3. Role of Curators in Curation and Museum Collection Storage

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

An accession number is assigned by the park curator at the time of a park research permit, when there will be field collections, or at the completion/milestone of a project, for "desktop" research. After collection, a range of catalog numbers can be assigned by the curator for the principal investigator to link to individual items (e.g., specimens). The accession and catalog number assignments must be requested by the project leader.

The project leaders will assemble all requested project materials for submission to the park or regional curator, as previously established in their collection permit. The curator will ensure the materials are properly stored, cataloged, and archived. The project leader is ultimately responsible for ensuring this is completed.

12.4. Workflow

PACN project leaders are responsible for preparing materials for the curator. The workflow illustrated here and described in this chapter refers to the general procedure for making the transition between the project leaders and the museum staff (Figure 12.1). The museum staff will follow a much more detailed procedure for actual curation. The workflow schematic begins with the delivery of the products, walks through the electronic storage of these products, and then highlights the storage and archiving of materials (Figure 12.1).

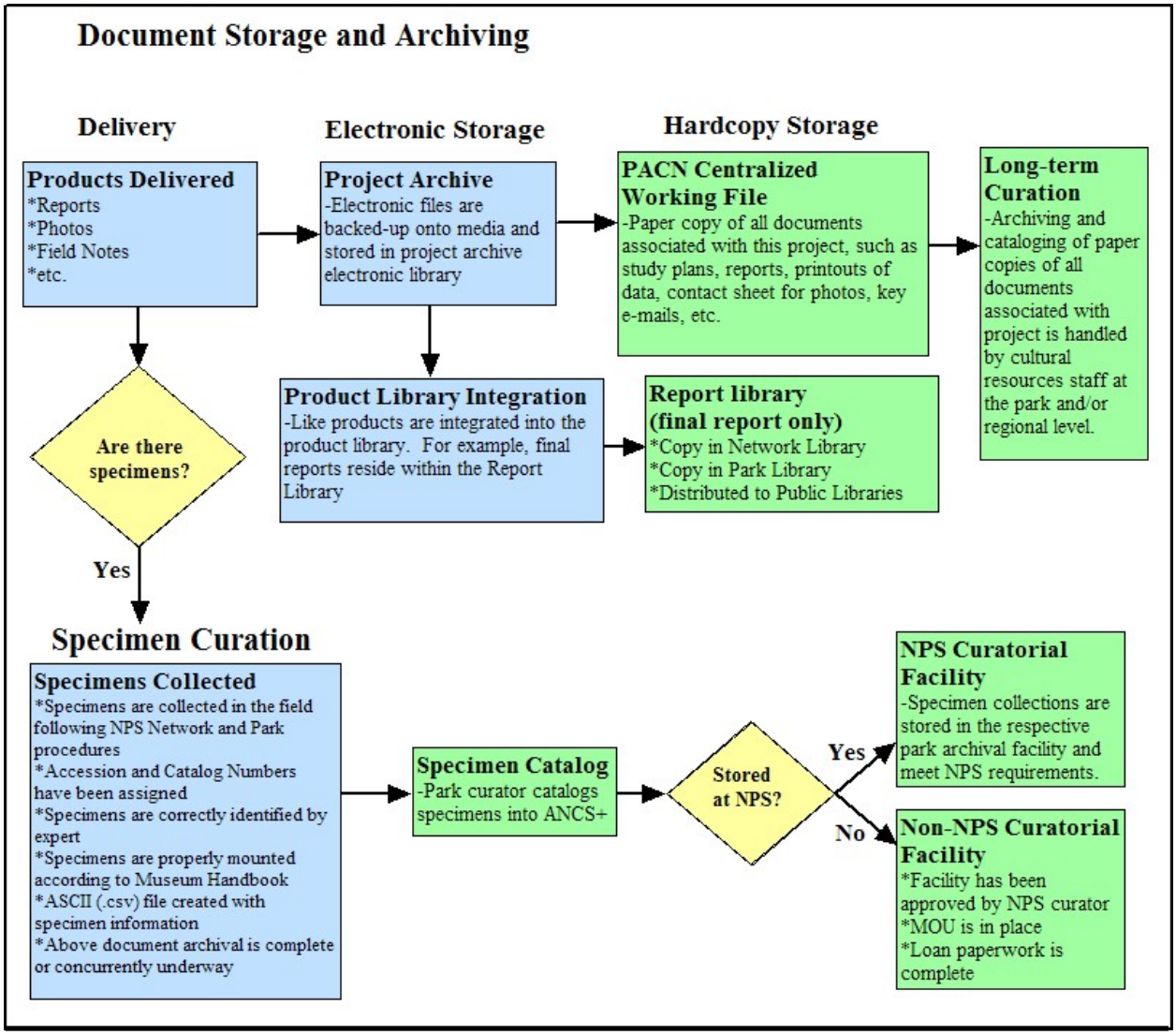


Figure 12.1. Schematic of natural history curation workflow for both hard copy documents and specimens.

12.5. PACN Network Archiving Process

Project materials should be consolidated and packaged for curation when a project is complete or, for long-term projects, reaches a milestone. The project leader should acquire an accession number from the respective curator.

The project leader should package the project information as follows:

- A coversheet, memo, or Collection Catalog Worksheet listing contact information, project abstract and purpose, sensitivity and use of materials, and all materials included in the package
- All materials are clearly labeled with:
  - Park acronym
  - Date, or range of dates

    - Accession number
    - Project number
- Field notes are on acid free paper (preferred) or copy paper and stored in a 3- ring binder or book box
- Other paper materials such as reports and data printouts are also stored on acid free paper (preferred) or copy paper and placed individually in an acid free, labeled folder
- Photos should be printed and documented
- Voucher specimens are properly labeled according to the Museum Handbook
- For specimens not residing in park collection facilities, loan paper work is completed and a copy of the form is stored with the project package
- Specimens not residing in NPS repositories should be stored in a selected facility that meets NPS museum collection standards
- Curation of specimens must be addressed in PACN vital sign protocols and should specify:
    - Name and address of facility
    - Memorandum of Agreement, if applicable
    - List of items to be curated or loaned
- CD or DVD of all electronic materials. Jewel case should be labeled with:
    - Park acronym
    - Date created
    - Range of dates for information
    - Accession Number
    - Project Number
    - List of contents

The museum staff will catalog this information into ANCS+ and will properly store the materials according to the Museum Handbook. If specified in arrangements with the network and each affected the park, a similar package should be prepared as "working copies" for the network and the park(s) for local storage and use.

12.6. Archive and Storage Locations

Many of the PACN parks lack adequate facilities for proper storage and archiving of documents, photographs, artifacts, and specimens from all natural and cultural resource projects. While most PACN parks actively store documents, many look to other institutions for storage and archiving of artifacts and specimens. The principal non-NPS facility used is the Bishop Museum in Honolulu. While storage of artifacts and specimens in non-NPS facilities presents challenges for park staff wishing to access materials, an institution that has ample resources committed to an organized and well maintained collections department serves scientists and researchers well because it represents a centralized location possessing a broad array of materials from a diversity of collection sites.

PACN parks will likely continue their cooperation with the Bishop Museum as a curatorial facility for artifacts and specimens collected in the parks.

12.6.1. PACN Network Working Files and Library

While a PACN project is on-going, all relevant materials (reports, administrative documents, critical e-mails or correspondence, and data print outs) are stored at the PACN offices. Upon project completion, PACN ensures that all final and pertinent project materials are submitted to the appropriate park for long-term storage and archiving. Unless the park lacks the resources to store and archive these materials, PACN generally will not store copies of these materials long-term.

A separate file cabinet for the PACN library will contain annual and final project reports and may contain additional reports that are related to inventories or vital sign monitoring. All materials in this library will be cataloged into the NatureBib database and should be citable. This library should not contain draft materials.

12.7. NPS Automated National Catalog System (ANCS+)

The NPS service-wide museum cataloging system is ANCS+ (using ReDiscovery software customized for NPS). Mostly, staff with cataloging duties are the only staff with this software. Park collections are entered into ANCS+ at the local level, then "rolled up" to the service-wide database.

The cataloging system uses an accession number to identify a collection, and catalog numbers to identify the items within the collection. For a project that collected specimens, the project has a single accession number, but contains multiple catalog numbers (one for each specimen collected). Park staff may access the catalog system by one of three methods:

- Contact a park curator and review together
- Request a copy of the program and data from the park curator
- Access the information online using ReDiscovery Web Catalog

Uploading of a park's ANCS+ records to the online ReDiscovery Web Catalog is provided at an added cost to the park. It should be noted that the online catalog may not represent all collection holdings at a given park.

12.8. Photographs

Some museum curators may be reluctant to fully embrace digital photography due to concern that, with the accelerating rate of technological change, documentary heritage is in danger of being lost in the information age (Cox 2000). PACN, however, recommends use of digital cameras, and will work with project leaders establish appropriate strategies for image file naming and image file maintenance (i.e., a well organized file folder directory structure).

Photographic methods to be used by any given project are developed by the project leader and PACN staff. PACN accepts both digital and analog photographs. Under the project leader's direction, PACN will screen photos quality, redundancy, and adequate documentation. For projects using analog methods, PACN accepts and processes 35mm slides (preferably Kodachrome or Ektachrome), which have a proven long-term stability (Wilhelm and Brower

1993), and 4x6 color prints. Original photographs are a high priority for placing in archival storage conditions.

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

Photographic prints should be stored in individual polypropylene sleeves and within archival boxes. Each photo should be labeled on the back, using archival-quality labels that are either laser-printed or hand-labeled with the same information elements required for slides. If a contractor is submitting photographs, corresponding TIFF files must also be submitted.

PACN employs a photo database (using ThumbsPlus and Microsoft Access software) to catalog photographic images useful for network purposes. Attributes captured for each image include electronic file name, keywords, project, photo description, photographer, date, and location. PACN recommends use of this photo database by project leaders as well, and will assist with database configuration for project leaders who have a need or desire to use the database.

12.9. Specimens

The network assists with planning for monitoring projects that includes budgeting for specimen preparation and records processing and cataloging. Specimens collected under the auspices of PACN will be provided to the respective park in which they were collected for curation, or to a repository approved by a park. Project leaders are responsible for providing park curators with necessary data for cataloging each specimen. PACN data management staff will assist project leaders and curators in ensuring that specimens records are in comma-delimited format (.csv) for automated uploading into ANCS+. Data provided to non-NPS curators will be in an appropriate format specified by the institution.

Many of the PACN parks will continue to use the Bishop Museum as a repository for specimens. Although details on the process for transferring specimens to the museum continue to evolve, the larger parks in Hawaii have established guidelines for collection permit requirements (Appendix J). Storing specimens at the Bishop Museum has the benefit of allowing future researchers access to not only NPS specimens, but to other specimens in their area of interest.

12.10. Other Materials

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

13. IMPLEMENTATION OF THE DATA MANAGEMENT PLAN

The Data Management Plan (DMP) contains practices that may be new to staff and principal investigators. With a few exceptions, however, the DMP does not include any requirements that are new. Almost every requirement comes from law, Director's Orders, or the I&M Program. The DMP helps to put these requirements into context and in sequence, and provides operational guidance for achieving these requirements. Good data management practices will take time. Some vital sign collection procedures and data management practices are already in use and may require minimal revisions. Others may involve several iterations of procedures and databases before reaching their acceptable and functional data reporting formats.

Vital sign monitoring protocols will be the primary focus of PACN data management efforts Integration of data management guidance and standards among these monitoring protocols and associated SOPs is an overarching goal of the PACN data management strategy, and will contribute significantly to the long-term usefulness of the I&M Program and its data products. See the appendices for examples of PACN data management SOPs.

13.1. Common Poor Practices in Data Management

As discussed earlier in Chapter 6 (Data Acquisition and Processing), data mining is an important component of all projects, as a means of discovering existing data and information that may be useful to the project. PACN spent considerable effort in data mining for data and information useful to the I&M Program. This consisted of looking through archives, office filing cabinets, natural resource folders, libraries, and electronic files. Not surprisingly, some data and information was well maintained and documented, and some was not. Examples of common data management problems are:

- *Unfinished projects*. This can happen for any number of reasons beyond the staff's control. Unfinished projects, however, are the most difficult to manage. It takes several times longer to document an incomplete project than a finished one. The best data management practice is to finish the project or reach a logical milestone.
- *Lack of header information*. Often, sheets of information do not have a date, name or subject. It requires reading the content of the material in order to make a determination if the information is relevant. This is extremely time consuming. Often relevant information is discarded because it does not have adequate header information. A good data management practice is to use a memorandum format or a standard header for all files for a project, no matter how insignificant.
- *Lack of documentation*. Most projects and data are inadequately documented. Without documentation, the data are often considered unusable. This DMP describes documentation techniques. Documentation should be on-going throughout the duration of the project.
- *Lack of file naming standard and version control*. In many cases it was impossible to know which file was the final product.
- *No archive or final storage*. Tucking years of data away in an individual's desk or personal computer isn't accessible to anyone else. Often it requires in-house knowledge to know the information even exists. Often when there is a turnover in staff, this information is lost.

13.2. Years 1-5 of Implementation

The first few years of the implementation of this plan will involve learning, testing, and refining. The DMP will be implemented for the first few vital sign monitoring protocols, and data management fundamentals training will be provided to network and park staff. During this period, templates and SOPs will be developed and tested, and these will then be applied to the other vital sign monitoring protocols. Some bottlenecks are anticipate, and identification and elimination of bottlenecks an important component in the implementation of this plan.

Implementation Goals for Years 1-5:

- All network staff understand and adhere to the fundamental guidance established in this plan, including:
  - File management
  - Documentation
  - Quality assurance and quality control
  - Electronic storage
  - Archive storage
- Improve data management practices by implementing:
  - Conceptual Data Models for vital sign monitoring protocols
  - Testing of data entry prior to field work
- Develop common SOPs that can be used for multiple protocols
- Inform and direct the data management aspects of the first vital sign monitoring protocols to be implemented

13.3. Years 5+ of Implementation

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

Implementation Goals for Years 5+:

- The DMP will be revised and streamlined for vital sign monitoring.
- All implemented vitals signs monitoring protocols will follow the DMP guidance.
- Databases and custom reporting will continue to be enhanced.
- Framework and gateway for integration of data with other agencies or networks will be established.

ACKNOWLEDGMENTS

The community of I&M data managers has produced some excellent guidance documents and presentation materials covering the strategies and procedures of sound data management. This community shares information freely, and this plan has borrowed extensively from the work of other I&M networks. The following table attempts to list the appropriate individuals as a way to acknowledge their contributions to each of the chapters in this plan. Undoubtedly some contributors have been left out of the table, and any such omissions are entirely due to the author's oversight. This same community will continue to be involved in review of network Data Management Plans, and future versions of this plan will improve from that collective wisdom and experience.

This plan has also benefited from review and suggestions by staff within the Pacific Islands Network, particularly Fritz Klasner, Leslie HaySmith, Sandy Margriter, and Allison Snyder. And their continued involvement will help raise the quality of writing and clarity of content in future versions of this plan. Additional reviewers of this plan were Gareth Rowell, data manager for the Heartland Network and Prairie Cluster Prototype, and Brent Frakes, data manager for the Rocky Mountain Network.

Acknowledgements Table. *Primary contributors to the content and concepts as presented by chapter in the PACN Data Management Plan.*

Chapter	Contributor	Affiliation
Chap 1. Introduction	Margaret Beer	I&M Program, Washington Support Office (WASO)
	Doug Wilder	Central Alaska Network
	John Boetsch	North Coast and Cascades Network
	Dorothy Mortenson	Southwest Alaska Network
	Sara Stevens	Northeast Coast and Barrier Network
	Alan Williams	Shenandoah National Park
Chap 2. Data Management Roles and Responsibilities	Rob Daley	Greater Yellowstone Network
	Margaret Beer	I&M Program, WASO
	Dorothy Mortenson	Southwest Alaska Network
	Doug Wilder	Central Alaska Network
Chap 3. Project and Data Management Overview	John Boetsch	North Coast and Cascades Network
	Sara Stevens	Northeast Coast and Barrier Network
	Vel Decker	Cape Cod national Seashore
	Doug Wilder	Central Alaska Network
	Dorothy Mortenson	Southwest Alaska Network
	Karen Oakley	USGS, Biological Division
	Sara Wesser	Alaska Region I&M Program
Chap 4. Infrastructure and System Architecture	Joe Gregson	Washington Support Office
	Lisa Nelson	Washington Support Office
	Patrick Flaherty	Appalachian Highlands Network
Chap 5. Database Design Overview	Dorothy Mortenson	Southwest Alaska Network
	John Boetsch	North Coast and Cascades Network
	Angie Southwould	Alaska Regional Office
Chap 6. Data Acquisition and Processing	Geoffrey Sanders	National Capital Region Network
	Jess Grundblatt	Alaska Regional Office
	Dorothy Mortenson	Southwest Alaska Network
Chap 7. Quality Assurance and Quality Control	Debbie Angell	Sonoran Desert Network
	Dorothy Mortenson	Southwest Alaska Network
	Margaret Beer	I&M Program, WASO
	Doug Wilder	Central Alaska Network
	Sara Stevens	Northeast Coast and Barrier Network
Chap 8. Data Documentation	Margaret Beer	I&M Program, WASO
	Bill Eichenlaub	Glacier Bay National Park
	Teresa Leibfreid	Cumberland Piedmont Network
	Bill Moore	Mammoth Cave National Park
	Doug Wilder	Central Alaska Network
Chap 9. Data Analysis and Reporting	Rob Daley	Greater Yellowstone Network
	Dorothy Mortenson	Southwest Alaska Network
Chap 10. Data Dissemination	Sara Stevens	Northeast Coast and Barrier Network
	Wendy Schumacher	I&M Program, WASO
	Doug Wilder	Central Alaska Network
Chap 11. Digital Data Maintenance, Storage, and Archiving	Dorothy Mortenson	Southwest Alaska Network
	Margaret Beer	I&M Program, WASO
	John Boetsch	North Coast and Cascades Network
Chap 12. Non-Digital Data Archiving and Records Management	Dorothy Mortenson	Southwest Alaska Network
	Stephanie Stephens	Alaska Regional Curatorial Center
	Hildy Reiser	Chihuahuan Desert Network
	Rob Daley	Greater Yellowstone Network
	Margaret Beer	I&M Program, WASO
Chap 13. Implementation	Dorothy Mortenson	Southwest Alaska Network
	Alan Bennett	Southwest Alaska Network

REFERENCES CITED

Cox, R.J. 2000. Searching for authority: archivists and electronic records in the New World at the fin-de-siècle. First Monday 5:1. (http://firstmonday.org/issues/issue5_1/cox/index.html , accessed on August 9, 2006).

Edwards, D. 2000. Data quality assurance. Pages 70-91 *in* Michener, W.K., and J.W. Brunt, editors, 2000, Ecological data: design, management and processing. Blackwell Science Ltd. Methods in Ecology Series, Malden, MA.

HaySmith, L., F. Klasner, S.H. Stephens, and G.H. Dicus. 2006. Pacific Island Network Vital Signs Monitoring Plan: Phase III report. National Park Service, Pacific Island Network, Hawaii National Park, HI.

Michener, W.K. 2000. Metadata. Pages 92-116 *in* Michener, W.K., and J.W. Brunt, editors, 2000, Ecological data: design, management and processing. Blackwell Science Ltd. Methods in Ecology Series, Malden, MA.

Michener, W.K., and J.W. Brunt (eds). 2000. Ecological data: design, management and processing. Blackwell Science Ltd. Methods in Ecology Series, Malden, MA.

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

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

Oakley, K.L., L.P. Thomas, and S.G. Fancy. 2003. Guidelines for long-term monitoring protocols. Wildlife Society Bulletin 31(4): 1000-1003.

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

Tessler, S., and J. Gregson. 1997. Draft Data Management Protocol. NPS I&M Program. (http://www1.nrintra.nps.gov/im/dmproto/joe40001.htm , accessed on August 9, 2006).

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

APPENDIX A. PACN PROJECT DELIVERABLES GUIDANCE DOCUMENT

Information Deliverables Specifications for PACN I&M Projects

DRAFT – Pacific Island Network (PACN)

Aug 17, 2004

PROJECT CHECKLIST

Project: _____
Date: _____
Contract: _____

Completed (date)	Deliverable	Required (due date)	Optional (due date)	Notes
	Study Plan			
	Project Log			
	Field Data Sheets			
	Tabular Data			
	GIS Data			
	Photographs			
	Data Collection Protocols			
	Data Documentation / Metadata			
	Information Discovery			
	Final Reports			
	Annual Progress Reports			Dovetail with RPRS reporting
	Collections Management Protocols			TBD with Collections staff at time of Research Permit application
	Research Permit application			Copy for I & M files

Notes/Specifications:

Information Deliverables Specifications for PACN I&M Projects
DRAFT - Aug 17, 2004

Purpose:
Many park management decisions are based on information obtained from inventory and research projects. The information deliverables listed here ensure long-term security and accessibility of information, and help guide and substantiate management decisions now and in the future.

This document describes the products which should appear on a CD or DVD delivered to the Pacific Island Network (PACN) on a yearly basis as the principal component of project deliverables funded by I&M.

Most of the items described below reflect good management practices that would be undertaken in one form or another by most projects. While it is a significant time investment to complete all items, they all become critical at some stage of most projects. It should be expected that approximately 30% of all project efforts be directed toward data management.

To facilitate good project data management, PACN will provide examples, advice, and review of the products described below. When project technicians are given concrete data management tasks at the beginning of a project, producing the important deliverables becomes a relatively simple process. If data management begins at the end of the field season, errors often occur and objectives are not met as readily. To help understand the additional effort involved in producing the deliverables, each section contains an "Efforts" component.

If the documents and data described below are maintained on a regular basis, producing the yearly deliverable product should be as simple as taking the entire project folder structure, copying it to a CD, and delivering it to the PACN Data Manager.

Definitions:
A 'project' is the sum total of all documents and actions taken under the auspices of a NPS Research Permit.

Scientific Data Deliverables
Early in the calendar year following receipt of project funding, the researcher will provide NPS with a CD or DVD containing the following organized and documented information:

- Study Plan
- Project Log
- Field Data Sheets
- Tabular Data
- GIS Data
- Photographs
- Data Collection Protocols
- Data Documentation / Metadata
- Information Discovery (results of project data mining)

- Annual Progress Reports and Final Reports
- Research Permit Application and Collections Management Protocols
- Project Safety Plan

Multi-year projects will deliver data on a yearly basis.

Study Plan

Description:
The Study Plan submitted with project proposal should be considered a static document that defines how a project was envisioned. To the maximum extent possible the Study Plan should outline planned strategies for data collection, data management and documentation, project reporting, and collections management. As necessary, the Study Plan can serve as a starting point for the protocols and deliverables described below.
Effort:
No additional time effort beyond preparation of a thorough Study Plan

Project Log

Description:
The Project Log is simply a single word document that records significant project events such as field data collection trips, project meetings, data processing progress, analysis progress, report writing.
Effort:
One hour per week

Field Data Sheets

Description:
Ideally, field data sheets will be scanned and linked to the appropriate data record in the project database. If a project needs assistance, please contact the PACN data manager. At a minimum, field data sheets should be organized into a collection (e.g., in a 3-ring binder or book-box) on an annual basis.

Scans, copies, or originals of all data sheets will be delivered to the park at project completion or every two years for multi-year projects.
Effort:
Data sheets need to be organized by a project regardless of the deliverables requirement. Providing copies to the park as a deliverable should require 2 days per year of effort to copy, organize and deliver. The preferred medium is a "scan" linked to the database.

Tabular Data

Description:
The preferred storage of tabular field data is in relational MS Access 2000 or higher databases. Whenever possible the database design should utilize an existing database template. Contact the PACN data manager for the most current template.

If there is no appropriate database template, the data structure should follow available guidelines:

- Alaska Support Office. 2002. National Park Service, Database Specifications for Inventory and Monitoring Studies
- Alaska Support Office. 2002. National Park Service, Recommended Database Strategies including I&M Database Templates
- Alaska Support Office. 2002. National Park Service, Recommended Naming Standards.

These guidelines also describe quality control procedures for data entry

Before data entry is initiated, database design should be reviewed and approved by the data manager. The data manager will review the database design for understanding and consistency with NPS database design strategies. If the Principal Investigator is unfamiliar with relational databases and database design, he/she should contact the PACN data manager for assistance in creating/customizing databases.

There are instances where other data formats are acceptable, especially where automated data collection instruments minimize or eliminate the possibility of data entry error. In such cases dbf, delimited text, or other electronic formats may be acceptable. Using MS Excel as a data storage software is unacceptable. All non- MS Access formats must be approved by the PACN data manager prior to initiation of data collection.

All database fields, tables, queries, forms, reports, and modules should have the "Description" portion of the objects "Properties" completed. This simple task serves as the documentation of the database that can be exported to metadata.

Data in each database should be reviewed and corrected using an approved verification method, such that data entry accuracy is 95% or greater. A description of the verification method and results will be included in the Database Design and Description document accompanying this database.

Effort:
Database design is time consuming and critical. It should be scheduled at the very earliest stages of a project to maximize benefit. While database design, using an existing NPS template, can be contracted out, the PACN data manager and database programmer are available to assist with database design. Consult with the PACN data manager for a needs assessment. Complete documentation should occur at the end of the project or at the end of the first 2 years. Documentation efforts should be about 2 days. Minimal additional effort should be required beyond the database design and documentation. Tabular data must be entered and stored electronically regardless of the deliverables requirement.

GIS Data

Description:
All field data with associated location data should be delivered as GIS shapefiles. Note that many large complex projects correctly produce a relatively small GIS layer documenting their data collection site and store the bulk of the field data in a relational database; in this case, there must be a clear unique identifier to link the GIS attribute data to data in other databases. GIS shapefiles should include current year's data. If the project spans multiple years, shapefiles

should present and summarize the data for the entire project dataset. Analysis and summary GIS shapefiles should also be delivered. Shapefiles should use the standard datum of NAD83 and UTM projection, and FGDC compliant metadata must be included for all shapefiles.
All aerial flights over the park should be documented as a GIS flightline layer.
Effort:
GIS data must be entered and documented regardless of the deliverables requirement. Minimal additional effort should be required to comply with the deliverables requirement.

Photographic Data

Description:
Photos collected as part of the Data Collection Protocol should be logically organized in a folder structure. Data photos should be linked to the project database.

Guidelines:
SWAN and SEAN 2004. Digital Photographs Management Specifications for Alaska Inventory and Monitoring Program. National Park Service. Consult PACN data manager for details.

Effort:
The effort to organize and link data photos varies tremendously depending on the number of photos collected and the database they are linked to. If thousands of data photos are to be collected then a substantial effort should be made to construct database forms to facilitate data photo entry.

Data Collection Protocols

Description:
Data collection protocols are a major component of the data deliverables for the FY05 I&M funding. They should provide a complete description of the data collection process for each data field on the data collection forms.

Guidelines:
Oakley, Karen, L. Thomas, S Fancy. 2003. Guidelines for Long-term Monitoring Protocols. This can be found at: http://science.nature.nps.gov/im/monitor/protocols/ProtocolGuidelines.pdf, or through the PACN data manager.

Effort:
Data collection protocols are a major time commitment. Expect to expend 5% of a project time on data collection protocols in the first and second years. Thereafter, effort should decrease substantially.

Data Documentation / Metadata

Description:
All project-related data should be outlined and described in this document. Tabular data must be fully explained, including descriptions/definitions of all tables and data fields (if users of MS Access complete the Description portion for all data fields, tables, queries, forms, reports, and modules, then export of this information to a metadata document is simple and straightforward). Metadata for GIS data should be completed in ArcCatalog or in text files. Additional items to be

included in Data Documentation / Metadata are summaries of data processing steps and/or data analysis steps (describing all software processes used to modify/analyze raw data).

Effort:

Thorough data documentation is time-consuming, and will likely require 5-10% of project time. However, completing small tasks from the onset can significantly simplify the process. For example, completing the description portion for all database objects during database development, completing GIS metadata during shapefile creation, and documenting data processing and analysis steps throughout the process will lay the groundwork for documenting all data components of a project. Generally, documentation of tabular data will occur early in the project, documentation of GIS layers and data processing will occur mid-stream, and documentation of data analysis will occur in the project's latter stages.

Information Discovery

(Literature Review/Bibliography/Data Clearinghouse Search)

A review of existing literature and data should be done at the onset of each project. A summary of the information discovered through this process should be summarized in the format of an annotated bibliography. This may be included in the Study Plan or Report or written as a separate, informal report.

Reports

Description:

Annual progress reports should summarize project progress, including a description of completed project deliverables and a schedule for remaining deliverables. Final reports should summarize the data and review data analysis in light of related scientific data and theory, clearly present key findings, and comment on management implications of project findings.

Copies:

An electronic copy of the report is required in MS Word and any supplemental information in MS Office products. If desired, these may also be provided in Adobe PDF format, consolidated into one document, or a series of documents with logical breaks. The use of bookmarks is encouraged.

For Annual Progress Reports, 2 hardcopies should be provided to the PACN I & M office. For Final Reports, the PACN Network Coordinator should be consulted regarding the number of hardcopies to be submitted (e.g. park libraries, etc.).

Collections Management Protocols and Research Permit Application

Description:

All projects that involve collection of specimens must develop a collections management protocol, which will require collaboration from Cultural Resources staff and approval from PACN staff. If specimens are to be stored at a non-NPS facility, agreements must be arranged through NPS Cultural Resources staff. Cultural Resources staff must be consulted regarding proper storage and cataloging of specimens. These arrangements must meet the needs of both the ANCS+ database (for Cultural Resources) and the NPSpecies database (for the I & M Program). In addition, the PACN office requires a copy of the Research Permit Application for all projects.

Effort:
A Collections Management strategy should be an integral part of the Research Permit. Although planning and collaboration to develop this strategy will not consume much project time, the processing of specimens (identification and proper labeling) may become labor intensive.

Additional Information : Filenaming Conventions

File naming conventions help in data management by clearly separating drafts from most current versions. These guidelines apply to all files:

- No spaces or special characters within the name
- Use date for version control, (YYYYMMDD or YYYYMM)
- Use underscore as delimiters
- Use 50 characters or less

The PACN data manager can be consulted for additional data management recommendations, such as management of digital images and appropriate folder structures for project data and documents.

APPENDIX B. PROJECT TRACKING DATABASE APPLICATION

Summary of the PACN Project Tracking Database application

For the purposes of organizing basic documentation for network projects, the PACN borrowed a project tracking database application from Dorothy Mortenson, Data Manager for the Southwest Alaska Network. The project tracking database was developed using Microsoft Access, and consists principally of a back-end database file that stores the data tables and a front-end database file that stores the forms, queries and reports that comprise the user interface. In this context, a project includes short-term data collection, analysis, and reporting efforts, such as inventories, and long-term efforts such as vital sign monitoring. The intent of the project tracking database application is to document a variety of project-specific information and to provide a convenient means of tracking the progress of each piece of project information that requires development, submission, and review (e.g., project reports, project databases, data analysis summaries, and GIS data layers). Types of project-specific information accommodated by the project tracking database include:

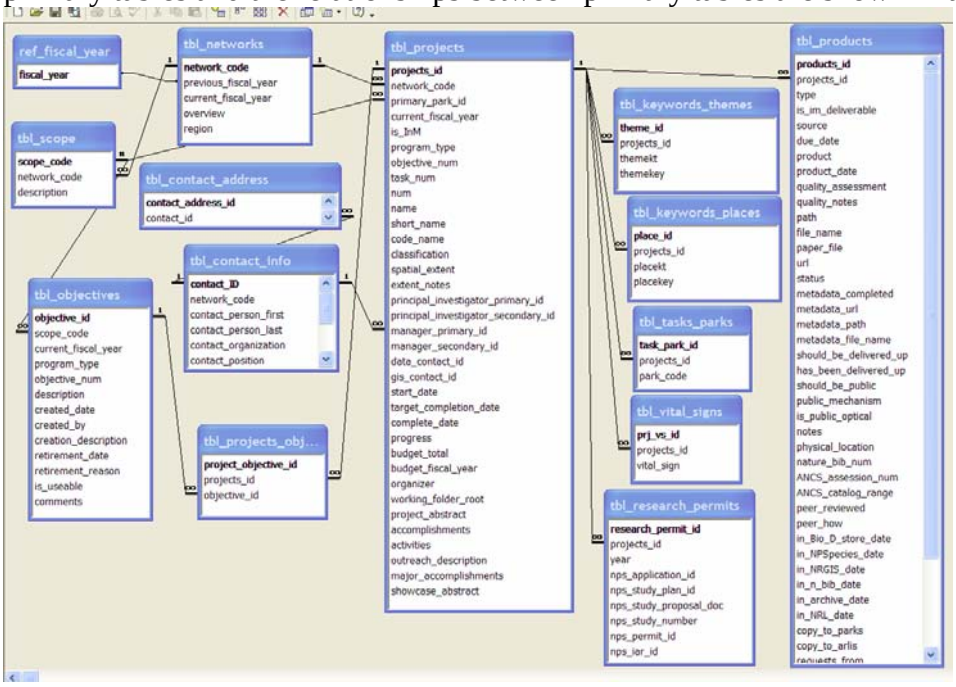
- Contact information
- Dates
- Links to Program goals, associated parks, vital signs, permits, investigators annual reports
- Abstracts
- Status, repository, and notes on deliverables

In addition, the project tracking database can generate reports that help to:

- document the project
- list the status of deliverables
- transfer necessary project information to park curators for archiving purposes

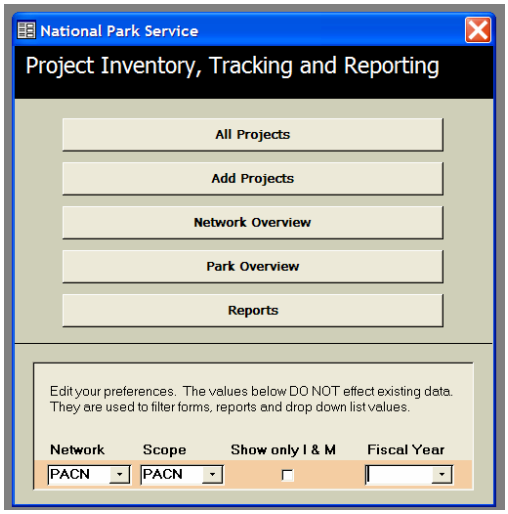
Structure of the back-end database

The back-end database file of the project tracking database application stores the data tables. The primary tables and the relationships between primary tables are shown in the diagram below:

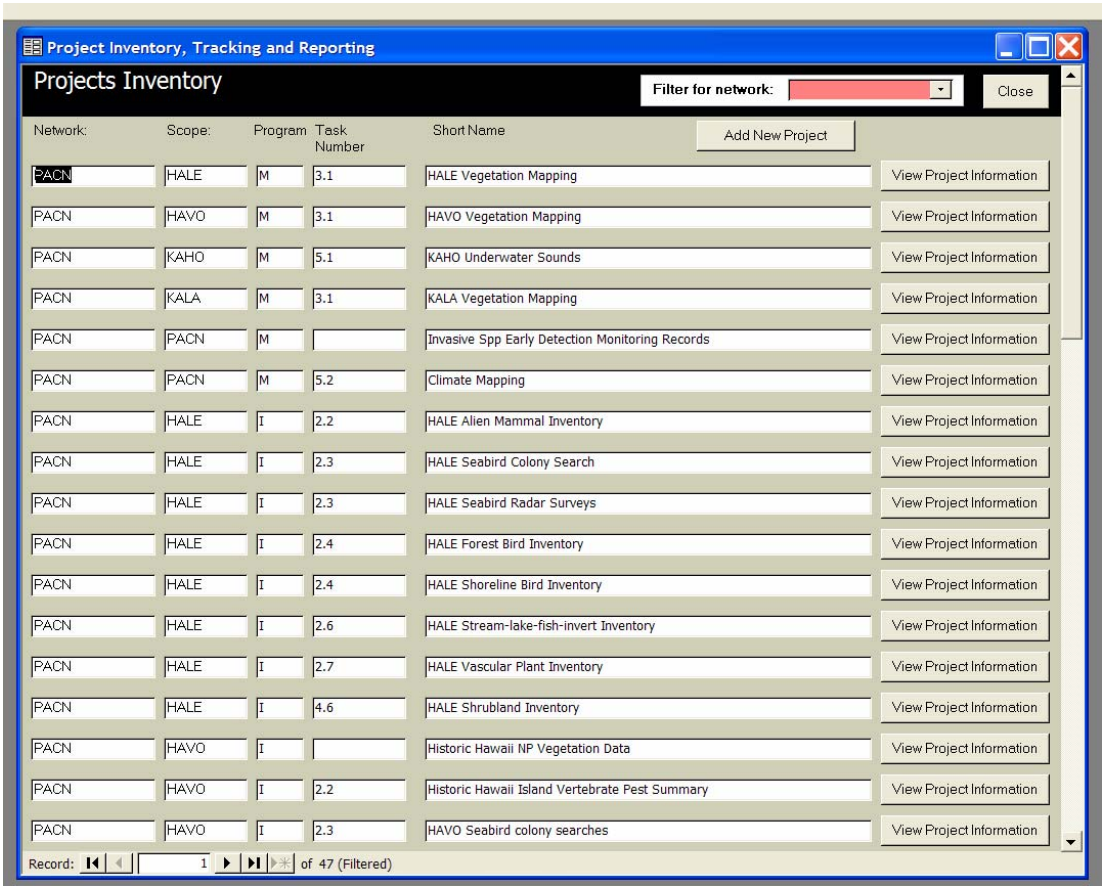


User-interface functionality of the front-end database file

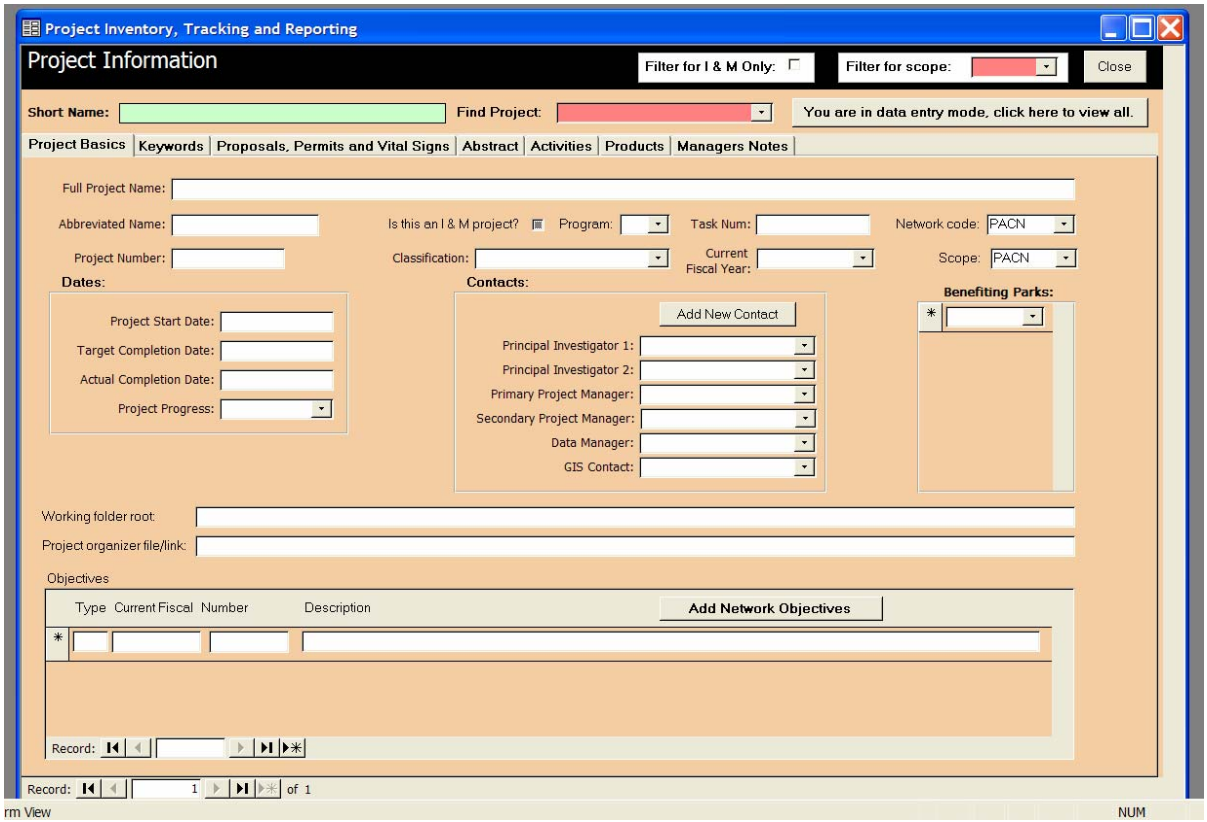
Upon opening the front-end database file, or user interface, the user is presented with the following switchboard form:



From this switchboard, the user can view a summary of all projects:



Or the user can enter a new project:



Finally, the user can also generate reports that summarize each project, including a list of product deliverables and the status of those deliverables. Below is an example of such a report, showing a list of selected PACN inventory projects:

PACN Project Summary

Biological Inventory, Vital Signs Monitoring, and Water Resources

Wednesday, August 09, 2006

Biological Inventory

2. Complete the documentation of 90% of vertebrate and vascular plant species in the parks through targeted field investigations and determine species distribution and abundance for species of concern of the network parks.

AARWP Task:	Project Short Name	Project Manager(s)	PI(s)	Start/End Dates	Status
2.1	**Hawaiian Hoary Bat Inventories**	Leslie HaySmith	David Duffy	02-2005	Complete
		Karin Schlappa	Darcy Hu	09-2005	

Project Log

Date/Person	Comments:
9/13/2005 koz	George Parrish, Heather Fraser, and Vanessa Parker-Geisman hired to conduct field work and write report.

Deliverables:

Type	Deliverable Product/Quality:	Due Date	Status?
Proposals	HI Bat project statement		Completed
Photographs	HI Bat photos	9/15/2005	Completed
Database	HI Bat database	9/15/2005	Completed
GIS Data	HI Bat GIS data	9/15/2005	Completed
Maps	HI Bat maps	9/15/2005	Completed
Report	HI Bat Report	9/15/2005	Completed

AARWP Task:	Project Short Name	Project Manager(s)	PI(s)	Start/End Dates	Status
2.2	**HALE Alien Mammal Inventory**	Leslie HaySmith	David Duffy		Complete
		Kelly Kozar	Cathleen Bailey	09-2005	

Project Log

Date/Person	Comments:
9/12/2005 koz	Cathleen Bailey and crew conducted field work.

Deliverables:

Type	Deliverable Product/Quality:	Due Date	Status?
Proposals	Inventory Alien Mammals in Kaapahu, Haleakala National Park		Completed
Fieldnotes	Kaapahu_Datasheets	9/15/2005	Completed
Database	Kaapahu Mammal Database	9/15/2005	Completed
Photographs	Kaapahu_Photos	9/15/2005	Completed
Maps	Kaapahu_Figures	9/15/2005	Completed
GIS Data	Kaapahu_GIS	9/15/2005	Completed
Report	HALE Alien Mammal Inventory Report	9/15/2005	Completed

AARWP Task:	Project Short Name	Project Manager(s)	PI(s)	Start/End Dates	Status
2.4	**W. HI Shoreline Bird Inventory**	Leslie HaySmith Karin Schlappa	Sharon Reilly Darcy Hu	01-2003 09-2005	Complete

Project Log

Date/Person	Comments:
9/12/2005 koz	Scott Waddington from Ducks Unlimited contracted to conduct surveys and write reports.

Deliverables:

Type	Deliverable Product/Quality:	Due Date	Status?
Proposals	W. HI Shoreline Bird Inventory project statement	9/15/2005	Completed
Photographs	W. HI Shoreline Bird Inventory photos	9/15/2005	Completed
Report	PUHE Shoreline Bird Inventory Report	9/15/2005	Completed
Report	PUHO Shoreline Bird Inventory Report	9/15/2005	Completed

2.4	**HAVO Lowland Bird Inventory**	Leslie HaySmith Karin Schlappa	David Duffy Darcy Hu	01-2005 09-2005	Complete

Project Log

Date/Person	Comments:
9/12/2005 koz	Kathryn Turner hired to conduct survey and write report.

Deliverables:

Type	Deliverable Product/Quality:	Due Date	Status?
Database	HAVO Lowland Bird Inventory database	9/15/2005	Completed
Proposals	HAVO Lowland Bird Inventory project statement	9/15/2005	Completed
Photographs	HAVO Lowland Bird Inventory photos	9/15/2005	Completed
GIS Data	HAVO Lowland Bird Inventory GIS files	9/15/2005	Completed
Report	HAVO Lowland Bird Inventory Report	9/15/2005	Completed

2.4	**KALA Forest Bird Inventory**	Leslie HaySmith Kelly Kozar	David Duffy Darcy Hu	 09-2005	Complete

Project Log

Date/Person	Comments:
9/13/2005 koz	Susan Marshall hired to conduct inventory and write report.

Deliverables:

Type	Deliverable Product/Quality:	Due Date	Status?
Proposals	KALA Forest Bird project statement		Completed
Fieldnotes	KALA Forest Bird databooks	9/15/2005	Completed
Database	KALA Forest Bird database	9/15/2005	Completed
GIS Data	KALA Forest Bird GIS	9/15/2005	Completed
Photographs	KALA Forest Bird Photos	9/15/2005	Completed
Report	KALA Forest Bird Report	9/15/2005	Completed
Maps	KALA Forest Bird Map	9/15/2005	Completed

4. Facilitate completion of additional high-priority biological inventories.

AARWP Task:	Project Short Name	Project Manager(s)	PI(s)	Start/End Dates	Status
4.1	**WAPA Vascular Plant Inventory**	Fritz Klasner Karin Schlappa	David Duffy Dwayne Minton	01-2004 09-2005	Complete

Project Log

Date/Person	Comments:
9/13/2005 koz	Joan Yoshioka hired to conduct inventory and write report.

Deliverables:

Type	Deliverable Product/Quality:	Due Date	Status?
Proposals	WAPA plant project statement		Completed
Database	WAPA plant database	9/15/2005	Completed
Photographs	WAPA plant photos	9/15/2005	Completed
GIS Data	WAPA plant GIS files	9/15/2005	Completed
Maps	WAPA plant maps	9/15/2005	Completed
Report	WAPA plant report	9/15/2005	Completed

AARWP Task:	Project Short Name	Project Manager(s)	PI(s)	Start/End Dates	Status
4.2	**HAVO Bryophyte Inventory**	Fritz Klasner Karin Schlappa	David Duffy Tim Tunison	01-2004 09-2005	Complete

Project Log

Date/Person	Comments:
11/14/2005 koz	Mashuri Waite hired to conduct inventory and write report.

Deliverables:

Type	Deliverable Product/Quality:	Due Date	Status?
Other	Task Agreement		Completed
Report	HAVO Bryophyte Inventory	9/15/2005	Completed

AARWP Task:	Project Short Name	Project Manager(s)	PI(s)	Start/End Dates	Status
4.3	**HAVO Kahuku Inventories, vasc plants**	Fritz Klasner Karin Schlappa	Tim Tunison	 09-2005	In work

Project Log

Deliverables:

Type	Deliverable Product/Quality:	Due Date	Status?
Report	HAVO Kahuku Vascular Plant Inventory Report		In work

APPENDIX C. PACN VITAL SIGN PROTOCOL SOP EXAMPLE – DATA MANAGEMENT SOP FROM BENTHIC MARINE COMMUNITY PROTOCOL

Benthic Marine Community Protocol: SOP #11: Data Management

Draft August 2006

Revision Log

Only changes in this specific SOP will be logged here. Version numbers increase incrementally by hundredths (e.g., version 1.01, version 1.02) for minor changes. Major revisions should be designated with the next whole number (e.g., version 2.0, 3.0, 4.0). Record the previous version number, date of revision, author of the revision, identify paragraphs and pages where changes are made, who approved the revision, and the reason for making the changes along with the new version number.

New Version Number	Revision Date	Author	Changes Made	Reasons for Change	Network Regional Reviewer Approval	Previous Version Number

Purpose

This SOP documents the Benthic Marine database and provides instructions for the development, maintenance, and distribution of monitoring data associated with the Benthic Marine Community protocol for the PACN. General procedures for data handling and quality assurance/quality control for all monitoring protocols implemented by the PACN monitoring program are described in the network's Data Management Plan posted both on the network data server and on the PACN website (http://www.nature.nps.gov/im/units/pacn/index.htm (last accessed 05July06). Microsoft Access is the primary software environment for managing benthic data. Metadata is managed using ESRI ArcGIS software and Microsoft Access software (for the NPS Metadata Tools & Editor application). In addition, ArcGIS software will serve as a tool for exploring the spatial component of the Benthic Marine data.

A master equipment list for the entire Benthic Marine Community Vital Sign Monitoring Protocol can be found in protocol SOP #1 "Before the Field Season." The master

equipment list should be updated simultaneously if any SOP requiring an equipment list is revised.

Benthic Database – Data Model

Marine monitoring at the PACN presently consists of five separate protocols: Benthic Marine, Marine Fishes, Water Quality, Fisheries Harvest, and Erosion and Deposition. The two monitoring protocols, Benthic Marine and Marine Fishes, will be co-located, with data managed within one database and data management framework. Combining the data for these two protocols is essential for effective analysis and interpretation. Although this SOP currently contains information pertaining only to the Benthic Marine protocol, future revisions will incorporate information relevant to the Marine Fishes protocol (particularly regarding database structure) as the Marine Fishes protocol is developed.

The Benthic Marine database employs a front-end/back-end configuration in which the user interface resides in the front-end file and the front-end file is linked to the back-end file which holds the data tables. The user interface includes a Back-End Linking Utility that allows the user to control this front-end/back-end file link. The front-end file (marine_benthic_FE_v1.0.mdb) contains the forms, queries, modules, macros, and reports for the Benthic Marine database application. The back-end file (marine_benthic_BE_v1.0.mdb) contains the data tables. This front-end/back-end configuration allows for continual improvements to the user interface (i.e., the various forms and queries for getting data into and out of the database) without requiring duplication or modification of the underlying data tables. Figure 1 shows the relationships among the primary tables in the database. The database is based on the National Park Service Natural Resources Database Template Version 3. The table “tbl_Sites” contains the park name and any descriptions, notes, etc. Linked to this table is “tbl_locations” which includes location information about transects. The linked table “tbl_events” records each sampling event and therefore includes the date. This allows for easy repeat entry where multiple dates are tied to the same locations or transects. The actual sampling data is maintained through four different tables which are linked to events. These tables are “tbl_benthic_cover”, “tbl_rugosity”, “tbl_recruitment”, and “tbl_growth”. Several sub-tables link from these data tables and contain more detailed information.

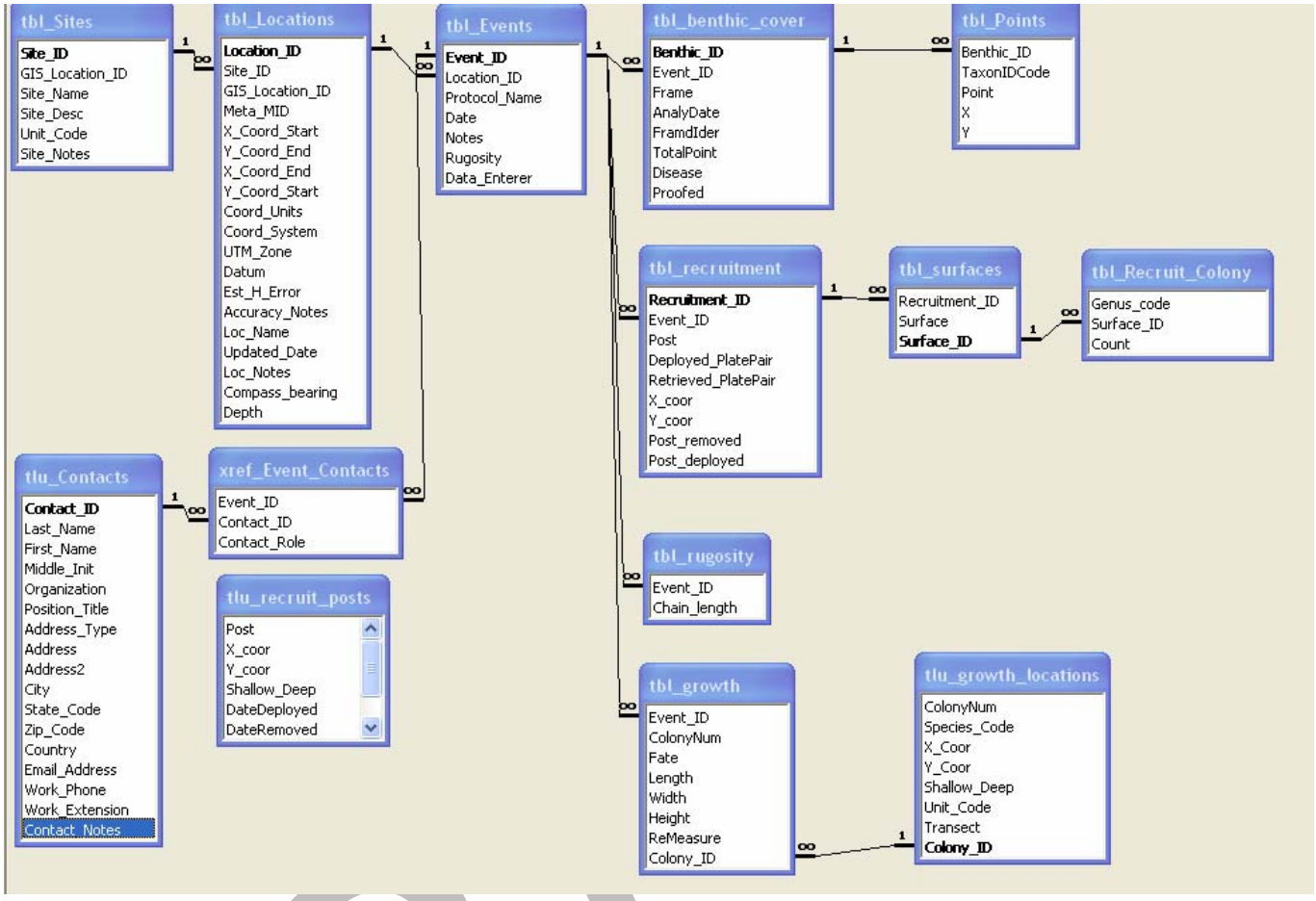


Figure S11.1. Data Model for the example database. There are four main data tables.

Documentation of Database Tables

Table: tbl_Sites

Description: Primary table that stores data about the broader location, which in this case is the park. There is one record for each park.

Field Name	Field Type	Size	Field Description
Site_ID	ReplicationID	NA	Site identifier
GIS_Location_ID	ReplicationID	NA	Link to GIS feature, equivalent to NPS_Location_ID
Site_Name	Text	100	Unique name or code for a site
Site_Desc	Text	255	Description for a site
Unit_Code	Text	12	4-letter Park, Monument or Network code
Site_Notes	Memo	NA	General notes on the site

Table: tbl_Locations

Description: Standardized table for storing information about the sampling location- in this case the location is each transect. Contains UTM coordinates for start and end of transects, as well as other spatial details such as UTM zone and datum. There is one record for each transect.

Field Name	Field Type	Size	Field Description
Location_ID	ReplicationID	NA	Location identifier
Site_ID	ReplicationID	NA	Link to tbl_Sites
GIS_Location_ID	ReplicationID	NA	Link to GIS feature, equivalent to NPS_Location_ID
Meta_MID	ReplicationID	NA	Link to NR-GIS Metadata Database
X_Coord_Start	Double	10	UTM X (easting) of transect start
Y_Coord_Start	Double	10	UTM Y (northing) of transect start
X_Coord_End	Double	10	UTM X (easting) of transect end
Y_Coord_End	Double	10	UTM Y (northing) of transect end
Coord_Units	Double	8	Coordinate distance units
Coord_System	Double	8	Coordinate system
UTM_Zone	Double	8	UTM Zone
Datum	Double	8	Datum of mapping ellipsoid
Est_H_Error	Double	5	Estimated horizontal accuracy
Accuracy_Notes	Memo	NA	Positional accuracy notes
Loc_Name	Text	10	Name of the location (F (fixed) or R (random) + Transect Number)
Updated_Date	Date/Time	10	Date of entry or last change
Loc_Notes	Memo	NA	General notes on the location
Compass_bearing	Single		Bearing of transect from starting point

Table: tbl_Events

Description: Standardized table containing information about each sampling event. There is one record for each sampling event.

Field Name	Field Type	Size	Field Description
Event_ID	ReplicationID	NA	Sampling event ID code
Location_ID	ReplicationID	NA	Link to tbl_Locations
Protocol_Name	Text	10	Protocol version used in this sampling event
Date	Date/Time	8	Sampling Date (Format dd/mm/yy)
Notes	Memo	NA	Any notes for this sampling event
Rugosity	Yes/No	NA	Was rugosity measured during this sampling event?
Data_Enterer	Text	50	Person who entered this event data

Table: tbl_benthic_cover

Description: Table for storing information on benthic cover. Data for this table and linked table "tbl_point" will be automatically imported from a data file which has been exported from the software "PhotoGrid".

Field Name	Field Type	Size	Field Description
Benthic_ID	ReplicationID	NA	Field data table row identifier
Event_ID	ReplicationID	NA	Link to tbl_Events
Frame	Integer	50	Sequence number of frames, as photographed along the transect. Range from 1-50.
AnalyDate	Date/Time	4	Date the photo was analyzed.
Framdlder	Text	50	The person who identified the species at specific points in the frame.
TotalPoint	Integer		Number of points within each photo frame where individual species were identified. In most cases this will be a set number (50); allows calculation of percent cover.

Table: tbl_rugosity

Description: Stores rugosity measurement. There is only one measurement per transect, but this can vary with time. It may not be measured every time a transect is sampled.

Field Name	Field Type	Size	Field Description
Event_ID	ReplicationID	50	Link to tbl_Events
Chain_length	Double	8	Actual horizontal length of transect. Rugosity will be calculated by dividing chain length (in meters) by transect length (25 meters).

Table: tbl_recruitment

Description: Stores recruitment data. Data include colonies that have been recruited to plates. While plates are not on transects, they are associated with them. Only used on fixed transects.

Field Name	Field Type	Size	Field Description
Recruitment_ID	ReplicationID	NA	Field data table row identifier
Event_ID	ReplicationID	NA	Link to tbl_Events
Post	Text	50	Which post was used (holds plate pairs)
Deployed_PlatePair	Integer	NA	Which plate pair was deployed?
Retrieved_PlatePair	Integer	NA	Which plate pair was retrieved?

X_Coor	Double	NA	Post location-distance from transect
Y_Coor	Double	NA	Post location-distance along transect
Post_Removed	Yes/No	NA	Was the above post removed in this survey?
Post_Deployed	Yes/No	NA	Was a new post added during the survey?

Table: tbl_growth

Description: Stores data about growth of coral colonies. Colonies are marked and measured over time. Only used on fixed transects.

Field Name	Field Type	Size	Field Description
Event_ID	ReplicationID	NA	Link to tbl_Events
ColonyNum	Integer	8	The number that identifies the colony
Fate	Text	10	Fate of colon-fission, phoenix, dead
Length	Single	8	Length of colony in cm
Width	Single	8	Width of colony in cm
Height	Single	8	Height of colony in cm
ReMeasure	Text	50	Was this a colony re-measurement?

Table: tbl_Point

Description: This is a sub-table to "tbl_benthic_cover". It is filled by importing data from the software Photogrid. It contains the raw data for each point that was identified within each frame of Photogrid.

Field Name	Field Type	Size	Field Description
Benthic_ID	ReplicationID	NA	Links to tbl_benthic_cover
TaxonIDCode	Text	10	Code for species identified at this point
Point	Integer	8	Point Number (1-50) (of the random points assigned by PhotoGrid)
X	Single	8	X coordinate of the point within the picture frame
Y	Single	8	Y coordinate of the point within the picture frame

Table: tbl_surfaces

Description: Stores which surfaces were analyzed, per plate pair. This links both to tbl_recruitment and tbl_recruit_colony. Allows for multiple surfaces per plate.

Field Name	Field Type	Size	Field Description
Recruitment_ID	ReplicationID	NA	Links to tbl_recruitment
Surface	Text	50	Which surface of plate pair: upper top, upper bottom, lower top, lower bottom, edges
Surface_ID	ReplicationID	NA	Links to tbl_recruit_colony

Table: tbl_Recruit_Colony

Description: Stores data on lowest identifiable taxon per plate surface. Uses tlu_taxon for more in-depth genus or family information.

Field Name	Field Type	Size	Field Description
Surface_ID	ReplicationID	NA	Links to tbl_surfaces
Genus_code	Text	250	Taxon identified on plate surface
Count	Integer	NA	The total count of each taxon identified

Table: tlu_Contacts

Description: Lookup table for observers and data enterers.

Field Name	Field Type	Size	Field Description
Contact_ID	ReplicationID	50	Contact identifier
Last_Name	Text	50	Last name
First_Name	Text	50	First name
Middle_Init	Text	50	Middle initial
Organization	Text	50	Organization or employer
Position_title	Text	50	Title or position description
Address_type	Text	50	Address (mailing, physical, both)
Address	Text	50	Street Address
Address2	Text	50	Address line 2, suite, apartment number
City	Text	50	City or town
State_Code	Text	8	State or province
Zip_Code	Text	50	Zip code (postal)
Country	Text	50	Country
Email_Address	Text	50	E-mail address
Work_phone	Text	50	Phone number
Work_extension	Text	50	Phone extension
Contact_notes	Memo	NA	Contact notes, if any

Table: xref_Event_Contacts

Description: Cross- reference table that links to both tbl_Events and tlu_Contacts. This keeps track of which contact is connected to each event, and how.

Field Name	Field Type	Size	Field Description
Event_ID	ReplicationID	NA	Link to tbl_Events
Contact_ID	ReplicationID	NA	Link to tlu_Contacts
Contact_Role	Text	50	The contact's role in the protocol

Table: tlu_taxon

Description: Lookup table for genus. This will be used in tbl_recruit_colony and tbl_growth. Allows data enterer to choose a code and not enter genus/family information for every record.

Field Name	Field Type	Size	Field Description
Genus_code	Text	NA	Genus code
Genus	Text	NA	Genus Name
Family	Text	50	Family Name

Table: tlu_growth_locations

Description: Lookup table for colonies measured in tbl_growth. This keeps track of permanent information such as location and species.

Field Name	Field Type	Size	Field Description
ColonyNum	Text	10	Colony number
Species_Code	Text	10	Species code
Shallow_Deep	Text	50	Whether the colony is on the shallow or deep side of the transect
X_coor	Single	8	X coordinate-distance from transect (meters)
Y_coor	Single	8	Y coordinate-distance along the transect (meters)

Table: tlu_recruit_posts

Description: Lookup table for recruitment posts, used in tbl_recruitment. This keeps track of more permanent information about the post (such as coordinates). It also keeps track of locations if posts are moved, so that plates on posts in the same location can be compared from year to year.

Field Name	Field Type	Size	Field Description
Post_Loc_ID	ReplicationID	NA	Post identification identifier
Post	Text	NA	Post identifier
SurfaceArea	Double	50	Area of each surface (top, bottom, sides)
X_coor	Single	50	X coordinate-distance from transect (meters)
Y_coor	Single	50	Y coordinate-distance down transect (meters)
Shallow_Deep	Text	50	Records whether the post is on the shallow or deep side of the transect (in case observers start from a different end of the transect each time)
DateDeployed	Date/Time	10	Date the post was deployed
DateRemoved	Date/Time	10	Date the post was removed

Table: tbl_DB_Meta
Description: Database description and links to I&M metadata tools.

Field Name	Field Type	Size	Field Description
DB_Meta_ID	ReplicationID	NA	Local primary key
Db_Desc	Memo	NA	Description of database purpose
Meta_MID	ReplicationID	NA	Link to NR-GIS Metadata Database
X_DSC_GUID	ReplicationID	NA	Link to I&M Dataset Catalog desktop metadata tool

Table: tbl_DB_Revisions
Description: Database revision history data.

Field Name	Field Type	Size	Field Description
Revision_ID	Text	NA	Database revision (version) number or code
Revision_Contact_ID	ReplicationID	NA	Link to tlu_Contacts
DB_Meta_ID	ReplicationID	NA	Link to tbl_DB_Meta
Revision_Date	Date/Time	10	Database revision date
Revision_Reason	Memo	NA	Reason for the database revision
Revision_Desc	Memo	NA	Revision description

Documentation of Database Queries

This SOP will be updated to reflect changes in queries. Basic queries follow.

Update and Append Queries
A series of update and append queries were created to import benthic cover data from PhotoGrid in the correct format. These include the following.

Query: qry_Import2
Description: Appends data from qry_Import1 to non-Temp table.
SQL: INSERT INTO tblCVSTemp ([ID Code], Point, X, Y, [File Name], [Total Points], [ID Date], Institution, [User Name], SurveyDate, FrameID, Loc_Name, Unit_Code, [ID Name])
SELECT tblCSVImportTemp.[ID Code] AS Expr2, tblCSVImportTemp.Point AS Expr3, tblCSVImportTemp.X AS Expr4, tblCSVImportTemp.Y AS Expr5, tblCSVImportTemp.[File Name] AS Expr10, tblCSVImportTemp.[Total Points] AS Expr11, tblCSVImportTemp.[ID Date] AS Expr12, tblCSVImportTemp.Institution AS Expr13, funname() AS Expr14, CDate((Mid(funD(),5,2)) & "/" & (Mid(funD(),7,2)) & "/" & (Left(fund(),4))) AS exp1, Mid([file name],10,3) AS Expr6, Mid([file name],8,2) AS Expr7, Left([file name],4) AS Expr8, tblCSVImportTemp.[ID Name] AS Expr1FROM

tblCSVImportTemp WHERE ((([tblCSVImportTemp].[ID Name]) Not Like "yes" And ([tblCSVImportTemp].[ID Name]) Not Like "no"));

Query: qry_Import_Disease
Description: Appends to non-Temp disease table.
SQL: INSERT INTO tblCVSTempDisease ([ID Code], Point, X, Y, [File Name], [Total Points], [ID Date], Institution, [User Name], SurveyDate, FrameID, Loc_Name, Unit_Code, [ID Name])
SELECT tblCSVImportTemp.[ID Code] AS Expr2, tblCSVImportTemp.Point AS Expr3, tblCSVImportTemp.X AS Expr4, tblCSVImportTemp.Y AS Expr5, tblCSVImportTemp.[File Name] AS Expr10, tblCSVImportTemp.[Total Points] AS Expr11, tblCSVImportTemp.[ID Date] AS Expr12, tblCSVImportTemp.Institution AS Expr13, funname() AS Expr14, CDate((Mid(funD(),5,2)) & "/" & (Mid(funD(),7,2)) & "/" & (Left(fund(),4))) AS exp1, Mid([file name],10,3) AS Expr6, Mid([file name],8,2) AS Expr7, Left([file name],4) AS Expr8, tblCSVImportTemp.[ID Name] AS Expr1 FROM tblCSVImportTemp WHERE ((([tblCSVImportTemp].[ID Name]) Like "yes" Or ([tblCSVImportTemp].[ID Name]) Like "no"));

Percent Cover Queries

A series of queries were created to provide percent covers by species, type, etc. These include the following queries.

Query: qry_Benthic_Cover1
Description: Presents all data for benthic cover- exported from Photogrid. This sets data up for more queries which provide percent cover and comparisons from year to year.
SQL: SELECT tbl_Sites.Unit_Code AS Park, tbl_Locations.Loc_Name AS Transect, Right([Date],4) AS [Year], tbl_benthic_cover.Frame, tbl_benthic_cover.TotalPoint, tbl_Points.TaxonIDCode, tbl_Points.Point, tblTaxon.Type
FROM tbl_Sites INNER JOIN (tbl_Locations INNER JOIN (tbl_Events INNER JOIN (tbl_benthic_cover INNER JOIN (tbl_Points INNER JOIN tblTaxon ON tbl_Points.TaxonIDCode = tblTaxon.TaxonName) ON tbl_benthic_cover.Benthic_ID = tbl_Points.Benthic_ID) ON tbl_Events.Event_ID = tbl_benthic_cover.Event_ID) ON tbl_Locations.Location_ID = tbl_Events.Location_ID) ON tbl_Sites.Site_ID = tbl_Locations.Site_ID;

Query: qry_Benthic_Cover2
Description: totals per frame, by species
SQL: TRANSFORM IIf(IsNull(Count([frame])),0,(Count([frame]))) AS Expr1
SELECT qry_Benthic_Cover1.Park, qry_Benthic_Cover1.Year, qry_Benthic_Cover1.Transect, qry_Benthic_Cover1.Frame, IIf(IsNull(Count([frame])),0,(Count([frame]))) AS TotalCoral, IIf(IsNull(Count([frame])),0,(Count([frame]))) AS Total
FROM tlu_Benthic_Cov_Taxon LEFT JOIN qry_Benthic_Cover1 ON tlu_Benthic_Cov_Taxon.TaxonIDCode = qry_Benthic_Cover1.TaxonIDCode

GROUP BY qry_Benthic_Cover1.Park, qry_Benthic_Cover1.Year, qry_Benthic_Cover1.Transect, qry_Benthic_Cover1.Frame
ORDER BY qry_Benthic_Cover1.Park, qry_Benthic_Cover1.Year, qry_Benthic_Cover1.Transect, qry_Benthic_Cover1.Frame
PIVOT [tlu_benthic_cov_taxon].[Type] & ":" & [tlu_benthic_cov_taxon].[TaxonIDCode];

Query: qry_Benthic_Cover3
Description: Totals per transect, by species.
SQL: TRANSFORM IIf(IsNull(Count([frame])),0,(Count([frame]))) AS Expr1
SELECT qry_Benthic_Cover1.Park, qry_Benthic_Cover1.Year, qry_Benthic_Cover1.Transect, IIf(IsNull(Count([frame])),0,(Count([frame]))) AS TotalCoral, IIf(IsNull(Count([frame])),0,(Count([frame]))) AS Total
FROM tlu_Benthic_Cov_Taxon LEFT JOIN qry_Benthic_Cover1 ON tlu_Benthic_Cov_Taxon.TaxonIDCode = qry_Benthic_Cover1.TaxonIDCode
GROUP BY qry_Benthic_Cover1.Park, qry_Benthic_Cover1.Year, qry_Benthic_Cover1.Transect
ORDER BY qry_Benthic_Cover1.Park, qry_Benthic_Cover1.Year, qry_Benthic_Cover1.Transect
PIVOT [tlu_benthic_cov_taxon].[Type] & ":" & [tlu_benthic_cov_taxon].[TaxonIDCode];

Query: qry_Benthic_Cover4
Description: Totals per transect, by type.
SQL: TRANSFORM IIf(IsNull(Count([frame])),0,(Count([frame]))) AS Expr1
SELECT qry_Benthic_Cover1.Park, qry_Benthic_Cover1.Year, qry_Benthic_Cover1.Transect, IIf(IsNull(Count([frame])),0,(Count([frame]))) AS Total
FROM tlu_Benthic_Cov_Taxon LEFT JOIN qry_Benthic_Cover1 ON tlu_Benthic_Cov_Taxon.TaxonIDCode = qry_Benthic_Cover1.TaxonIDCode WHERE (((qry_Benthic_Cover1.Park) Is Not Null))
GROUP BY qry_Benthic_Cover1.Park, qry_Benthic_Cover1.Year, qry_Benthic_Cover1.Transect ORDER BY qry_Benthic_Cover1.Park, qry_Benthic_Cover1.Year, qry_Benthic_Cover1.Transect PIVOT tlu_Benthic_Cov_Taxon.Type;

Query: qry_Benthic_Cover5
Description: Percent per transect, by type.
SQL: SELECT qry_Benthic_Cover4.Year, qry_Benthic_Cover4.Park, qry_Benthic_Cover4.Transect, qry_Benthic_Cover4.Total, [algae]/[total] AS [Algae%], [coral]/[total] AS [Coral%], [invert]/[total] AS [Invert%], [other]/[total] AS [Other%], [substrate]/[total] AS [Substrate%] FROM qry_Benthic_Cover4 WHERE (((qry_Benthic_Cover4.Park) Is Not Null));

Query: qry_Benthic_Cover6
Description: Mean of percent cover for all transects per year per park. Includes SD. By Type.
SQL: SELECT qry_Benthic_Cover5.Year, qry_Benthic_Cover5.Park, Avg(qry_Benthic_Cover5.[Algae%]) AS [AvgOfAlgae%],

StDev(qry_Benthic_Cover5.[Algae%]) AS [StDevOfAlgae%],
Avg(qry_Benthic_Cover5.[Coral%]) AS [AvgOfCoral%],
StDev(qry_Benthic_Cover5.[Coral%]) AS [StDevOfCoral%],
Avg(qry_Benthic_Cover5.[Invert%]) AS [AvgOfInvert%],
StDev(qry_Benthic_Cover5.[Invert%]) AS [StDevOfInvert%],
Avg(qry_Benthic_Cover5.[Substrate%]) AS [AvgOfSubstrate%],
StDev(qry_Benthic_Cover5.[Substrate%]) AS [StDevOfSubstrate%],
Avg(qry_Benthic_Cover5.[Other%]) AS [AvgOfOther%],
StDev(qry_Benthic_Cover5.[Other%]) AS [StDevOfOther%]
FROM qry_Benthic_Cover5 GROUP BY qry_Benthic_Cover5.Year,
qry_Benthic_Cover5.Park;

Query: qry_%Cov_x_transect
Description: Similar to qrySiteSummary in CRAMP db.
SQL: SELECT qry_Benthic_Cover3.Park, qry_Benthic_Cover3.Year,
qry_Benthic_Cover3.Transect, qry_Benthic_Cover3.Total, [Algae:Coralline algae]/[total]
AS [Coralline algae], [Algae:Macroalgae]/[total] AS Macroalgae, ([Algae:Turf
algae]/[total]) AS [turf algae], [coral:Cyphastrea ocellina]/[total] AS [Cyphastrea ocellina],
[Coral:Fungia scutaria]/[total] AS [Fungia scutaria], [Coral:Montipora capitata]/[total] AS
[Montipora capitata], [Coral:Montipora flabellata]/[total] AS [Montipora flabellata],
[Coral:Montipora patula]/[total] AS [Montipora patula], [Coral:Montipora studeri]/[total]
AS [Montipora studeri], [Coral:Pavona duerdeni]/[total] AS [Pavona duerdeni],
[Coral:Pavona maldivensis]/[total] AS [Pavona maldivensis], [Coral:Pavona
varians]/[total] AS [Pavona varians], [Coral:Pocillopora damicornis]/[total] AS
[Pocillopora damicornis], [Coral:Pocillopora eydouxi]/[total] AS [Pocillopora eydouxi],
[Coral:Pocillopora ligulata]/[total] AS [Pocillopora ligulata], [Coral:Pocillopora
meandrina]/[total] AS [Pocillopora meandrina], [Coral:Porites brighami]/[total] AS
[Porites brighami], [Coral:Porites compressa]/[total] AS [Porites compressa],
[Coral:Porites evermanni]/[total] AS [Porites evermanni], [Coral:Porites lichen]/[total] AS
[Porites lichen], [Coral:Porites lobata]/[total] AS [Porites lobata], [Coral:Porites
rus]/[total] AS [Porites rus], [Coral:Psammocora nierstraszi]/[total] AS [Psammocora
nierstraszi], [Coral:Psammocora verrilli]/[total] AS [Psammocora verrilli],
[Coral:Unknown Coral]/[total] AS [Unknown Coral], [Invert:Diadema paucispinum]/[total]
AS [Diadema paucispinum], [Invert:Echinometra mathaei]/[total] AS [Echinometra
mathaei], [Invert:Echinothrix calamaris]/[total] AS [Echinothrix calamaris],
[Invert:Heterocentrotus mammillatus]/[total] AS [Heterocentrotus mammillatus],
[Invert:Holothuriidae]/[total] AS Holothuriidae, [Invert:Porifera]/[total] AS Porifera,
[Invert:Tripneustes gratilla]/[total] AS [Tripneustes gratilla], [Invert:Tunicate]/[total] AS
Tunicate, [Invert:Zoanthid]/[total] AS Zoanthid, [Other:Other]/[total] AS Other,
[Other:Other/fish]/[total] AS [Other/fish], [Other:Overlapping image]/[total] AS
[Overlapping image], [Substrate:N/C]/[total] AS [N?C], [Substrate:Rubb]/[total] AS Rubb,
[Substrate:Sand]/[total] AS Sand, [Substrate:Silt]/[total] AS Silt, [Coral:Leptastrea
purpurea]/[total] AS [Leptastrea purpurea] FROM qry_Benthic_Cover3 WHERE
(((qry_Benthic_Cover3.Park) Is Not Null));

Query: qry_%Cov_x_transect2
Description: Mean of percent cover for all transects per year per park. By Species.
SQL: SELECT qry_Benthic_Cover3.Park, qry_Benthic_Cover3.Year, Avg([Algae:Coralline algae]/[total]) AS [Coralline algae], Avg([Algae:Macroalgae]/[total]) AS Macroalgae, Avg(([Algae:Turf algae]/[total])) AS [turf algae], Avg([coral:Cyphastrea ocellina]/[total]) AS [Cyphastrea ocellina], Avg([Coral:Fungia scutaria]/[total]) AS [Fungia scutaria], Avg([Coral:Montipora capitata]/[total]) AS [Montipora capitata], Avg([Coral:Montipora flabellata]/[total]) AS [Montipora flabellata], Avg([Coral:Montipora patula]/[total]) AS [Montipora patula], Avg([Coral:Montipora studeri]/[total]) AS [Montipora studeri], Avg([Coral:Pavona duerdeni]/[total]) AS [Pavona duerdeni], Avg([Coral:Pavona maldivensis]/[total]) AS [Pavona maldivensis], Avg([Coral:Pavona varians]/[total]) AS [Pavona varians], Avg([Coral:Pocillopora damicornis]/[total]) AS [Pocillopora damicornis], Avg([Coral:Pocillopora eydouxi]/[total]) AS [Pocillopora eydouxi], Avg([Coral:Pocillopora ligulata]/[total]) AS [Pocillopora ligulata], Avg([Coral:Pocillopora meandrina]/[total]) AS [Pocillopora meandrina], Avg([Coral:Porites brighami]/[total]) AS [Porites brighami], Avg([Coral:Porites compressa]/[total]) AS [Porites compressa], Avg([Coral:Porites evermanni]/[total]) AS [Porites evermanni], Avg([Coral:Porites lichen]/[total]) AS [Porites lichen], Avg([Coral:Porites lobata]/[total]) AS [Porites lobata], Avg([Coral:Porites rus]/[total]) AS [Porites rus], Avg([Coral:Psammocora nierstraszi]/[total]) AS [Psammocora nierstraszi], Avg([Coral:Psammocora verrilli]/[total]) AS [Psammocora verrilli], Avg([Coral:Unknown Coral]/[total]) AS [Unknown Coral], Avg([Invert:Diadema paucispinum]/[total]) AS [Diadema paucispinum], Avg([Invert:Echinometra mathaei]/[total]) AS [Echinometra mathaei], Avg([Invert:Echinothrix calamaris]/[total]) AS [Echinothrix calamaris], Avg([Invert:Heterocentrotus mammillatus]/[total]) AS [Heterocentrotus mammillatus], Avg([Invert:Holothuriidae]/[total]) AS Holothuriidae, Avg([Invert:Porifera]/[total]) AS Porifera, Avg([Invert:Tripneustes gratilla]/[total]) AS [Tripneustes gratilla], Avg([Invert:Tunicate]/[total]) AS Tunicate, Avg([Invert:Zoanthid]/[total]) AS Zoanthid, Avg([Other:Other]/[total]) AS Other, Avg([Other:Other/fish]/[total]) AS [Other/fish], Avg([Other:Overlapping image]/[total]) AS [Overlapping image], Avg([Substrate:N/C]/[total]) AS [N?C], Avg([Substrate:Rubb]/[total]) AS Rubb, Avg([Substrate:Sand]/[total]) AS Sand, Avg([Substrate:Silt]/[total]) AS Silt, Avg([Coral:Leptastrea purpurea]/[total]) AS [Leptastrea purpurea] FROM qry_Benthic_Cover3 GROUP BY qry_Benthic_Cover3.Park, qry_Benthic_Cover3.Year HAVING (((qry_Benthic_Cover3.Park) Is Not Null));

Database Administration

Database files will be distributed to individual parks at the beginning of the field season to facilitate prompt data entry and edits by park staff. At the end of the field season, files will be submitted to the PACN data management staff for consolidation into the master database file. While PACN recommends that parks store the database files on a park server with automatic backups, some parks lack these resources and will store the files on a local computer and employ their own backup strategy. PACN also recommends

that the parks store the database files and backups in the appropriate I&M project folder (see Data Organization section), preferably with the back-end file on a shared server drive and the front-end file on the same server drive or on a local drive. At the beginning of each data entry session, the user must create a backup of the back-end data file, ensuring that the initial data entry starting point can be recovered should irreversible errors or problems occur during the data entry session. The backup will be done using a database utility and a prompt will appear at the onset of a data entry session, checking that the back up has been made. This will not require a name change or revision change. The backup files will be named by adding the current date to the end of the back-end file name (.e.g., marine_benthic_BE_v1.0_20060530.mdb for a backup file created on May 30, 2006). Back-up copies are used for the current field season only and will not be archived.

The Benthic Marine database must be stored in the specified I&M project folder, and all data entry and edits must occur in this database file only. It is the responsibility of the PI or designee(s) to ensure that multiple copies of the database are not created, with data entry and edits erroneously occurring in multiple database files. The PACN data management staff will work with the PI or designated staff to design appropriate exploratory data analysis tools (e.g., specific queries and reports tailored to the project's exploratory data analysis needs) and incorporate these tools into updated versions of the Benthic Marine database application. Such changes will be fully documented in the protocol or SOP edit logs and reflected in the protocol database filename (e.g., in version identifier) and metadata.

Data Entry from Data forms

Data entry should occur as soon as possible after data collection is completed and before the next sets of observations occur. Data entry should be completed by the person who collected the data or someone who is familiar with the project and data. The primary goal of data entry is to transcribe the data from paper and photographic records into the computer with 100% accuracy. If any sensitive data is collected (e.g., location of rare organisms or spawning aggregations) then make a note of this on the original data form and send a copy to the database manager to flag the data as sensitive.

Data entry will occur at multiple locations. At each park, a single database back-end should be used for all data entry and manipulation. Site information, coral recruitment, rugosity, and coral growth data will be entered at each individual park by park staff. Photogrid comma-separated-value (CSV) files containing data on benthic cover and disease will be generated (see SOP #13) by the Marine Biological Technician. These Photogrid CSV files will be imported into the Benthic Marine database back-end file(s) at the end of the field season after QA/QC corrections have been made. Waiting until the end of the field season will facilitate matching the transect data from Photogrid to the sampling event for the other data components (site information, coral recruitment, rugosity, and coral growth) collected at the same time and place as the digital photos. Import of the Photogrid CSV files will be accomplished collaboratively by the PACN data

management staff and the Marine Biological Technician. Photo and Photogrid CSV files will be backed up, with a copy retained at each park after these data are collected.

At the end of the field season, the Principle Investigator or designee, the Marine Biological Technician, and PACN data management staff are collectively responsible for ensuring that all validated data have been migrated into a single back-end database file representing all work at all parks for that field season. This final verified database file will then be archived by the PACN data manager. Read-only copies of this final database file will be sent to each park for their data summary needs and their local archiving.

A database user manual is in development, and will provide step-by-step instructions accompanied by screen-capture images. This user manual will not be completed until after the initial phases of data entry following protocol implementation. Eventually, the user manual will be included as an appendix to this SOP. Ultimately, it is the Principal Investigator's and park lead's shared responsibility to ensure that all data entry staff understand how to enter data and follow all applicable SOPs. Data entry technicians are responsible for becoming familiar with the field data forms, the database software, database structure, Photogrid export file formats, and any standard codes for data entry.

Data Maintenance

Any editing of archived data is accomplished jointly by the PI or designee and PACN data manager. Every change must be documented in the edit log and accompanied by an explanation that includes pre- and post-edit data descriptions (Tessler and Gregson 1997). All data collected using this protocol are subject to the following three caveats:

Only make changes that improve or update the data while maintaining data integrity.

Once archived, document any changes made to the dataset through an edit log. At end of each fiscal year, the database manager will update the central database and will send out read-only versions.

Mistakes can made during editing so updates must be compared with the original data form prior to validating the data.

Data Organization

To help make file management intuitive and straight-forward, PACN recommends that each park store digital files related to the Benthic Marine protocol in a standardized file folder directory structure (Figure S11.2). This directory structure should be used on a shared server where files are automatically backed-up on an established schedule. If necessary, the directory structure can be used on a local computer drive with file back-ups accomplished by the PI or appropriate staff.

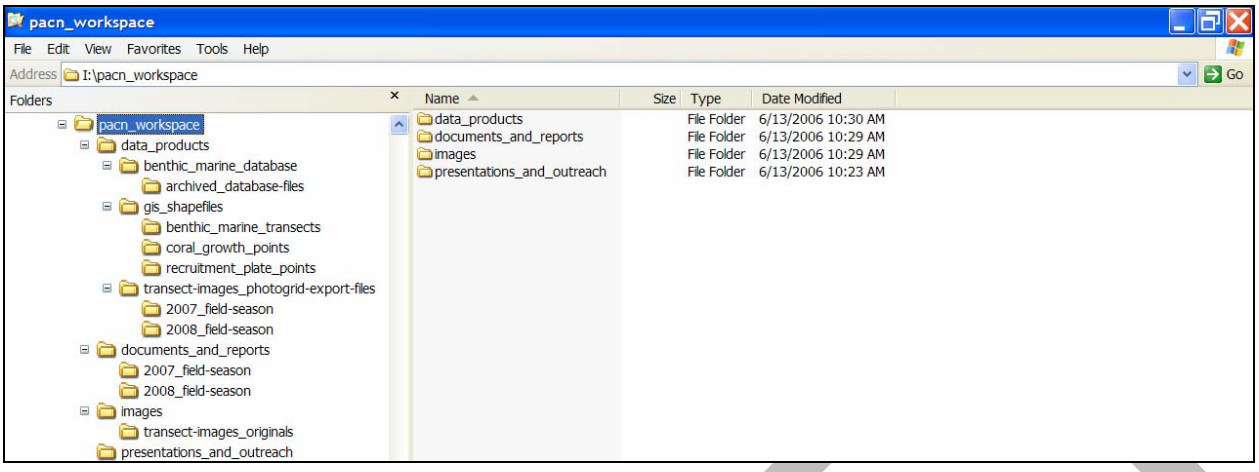


Figure S11.2. Recommended file folder directory structure for storing Benthic Marine digital files.

Version Control

Prior to any major changes of a dataset, a copy is stored with the appropriate version number to allow for tracking of changes over time. Versioning of archived datasets is handled by adding a three digit number to the file name, with the first version being numbered 001 (e.g., marine_benthic_BE_v1.0_validated-20071121_v001, for the first version of a back-end data file validated by the PI and data manager at the end of the 2007 field season). Each additional version is assigned a sequentially higher number. Frequent users of the data are notified of the updates, and provided with a copy of the most recently archived version.

Data Logs and Backups

Once the data are archived, any changes made to the data must be documented in an edit log. Original field forms will not be altered. Field forms can be reconciled to the database through the use of the edit log. Secure data archiving is essential for protecting data files from corruption. Once a dataset has passed the QA/QC procedures specified in the protocol, a formal metadata record is created and the PACN project tracking database is updated to reflect the validation of that year's dataset. The validated dataset is archived on the PACN I&M server in read-only format, and copies are distributed to the individual parks for their local read-only backup. At a minimum, the metadata record for the validated dataset is posted to the appropriate National Park Service online data store (e.g., the NR-GIS Data Store, http://science.nature.nps.gov/nrdata (last accessed 05July06). If there are no sensitive data issues, then the dataset file will also be posted to the online data store.

Verification of Data Entry

Quality control of data is a critical step in data management. Verification of data (ensuring data on field sheets match data entered into a database) is the responsibility of the PI and designees. See recommendations outlined in Chapter 4 of the Benthic Marine protocol. The specific data verification processes are described in the applicable Benthic Marine protocol SOPs, which will be updated as improvements are made to these data verification processes. Any modifications to the Benthic Marine database will be described in the edit log and the functionality of the verification routines will be explained in detail in the Benthic Marine database user manual (in development).

Data verification will occur in three steps.

Step 1: Park staff, working closely with the PI and Marine Biological Technician, will be responsible for reviewing 10% of all benthic image analysis records each field season. Review of benthic image analysis records will be accomplished using the Photogrid software, and edits will be made prior to generating Photogrid export files for import into the Benthic Marine database (see details in the Benthic Image Analysis SOP). This review will serve as both a data verification process and a data validation process. For the purposes of data verification, if more than 10% of the reviewed records are affected by data entry errors, then all of the records will be reviewed by the park lead and corrected by the marine biological technician as needed.

Step 2: Using record review routines built into the Benthic Marine database, park staff will compare 10% of all transect site information, rugosity data, coral growth data, and coral recruitment data against the original hardcopy data sheets. Corrections to data entry errors will be tracked in order to quantify data entry error rates. If more than 10% of the reviewed records are affected by data entry errors, then all of the records will be reviewed by the park lead and corrected by the marine biological technician as needed.

Step 3: The geospatial component of the Benthic Marine database will be verified each field season. A database utility linking the database to ArcGIS software will be used to evaluate all geospatial coordinates. Any corrections will be tracked, and an error rate greater than 10% will require a detailed review of the coordinate fields for all records. As this database functionality is finalized, it will be fully described in this Data Management SOP and in the database user manual.

Validation of Datasets

At the end of each field season, the PI, designees, and the PACN data management staff are collectively responsible for finalizing a validated dataset for that field season. Park staff, working closely with the PI and the Marine Biological Technician, will complete all data validation. Some validation (ensuring that the data make sense) methods have been incorporated into the benthic marine database. Other, more specific validation routines will be worked out with the protocol PI and/or project staff and incorporated into the database as appropriate. These modifications will be described in

the edit log and the functionality of the validation routines will be explained in detail in the Benthic Marine database user manual (in development).

Validation of benthic image analysis records will be accomplished using the Photogrid software. A second observer, by comparing their own benthic cover interpretations for 10% of the images against the interpretations of the original observer, will validate the original observer's records. If there is disagreement with more than 10% of these reviewed records, then the two observers will review the benthic cover interpretation guidelines and the original observer will re-interpret all of the images. In this event, the validation of benthic image analysis records would then be repeated. This process of validating benthic image analysis records will be completed before that field season's Photogrid export files are imported into the Benthic Marine database.

Validation of coral recruitment analysis records will also be accomplished by park staff, in collaboration with the PI and Marine Biological Technician. As with the benthic image analysis records, park staff will select 10% of the coral recruitment plates to be re-read by a second observer. If there is disagreement with more than 10% of the records, for the coral taxon identified and the number of recruits counted, then the two observers will review the guidelines for identifying and counting coral recruits, and the original observer will re-analyze all of the plates. In this event, the validation of 10% of the plates by a second observer would be repeated.

The PI, in conjunction with all relevant staff, will be responsible for evaluating all of the database records for a given field season in order to provide final approval for a validated dataset for that field season. PACN data management staff will work closely with the PI and PICRP personnel to provide any needed database queries, reports, graphs, or export file formats to assist with this overall validation. Once a given field season's dataset has been validated, the PACN data manger will be responsible for archiving the validated dataset, posting the metadata (and dataset, if appropriate) to the appropriate NPS online data store, and distributing read-only copies of the validated datasets to the parks.

File Management

Current Files

The master versions of all digital files relating to the Benthic Marine protocol are stored on the PACN file server, with regular file back-ups accomplished automatically. Presently, the main files include the protocol narrative, the SOPs, and the Benthic Marine database files. These files are stored according to the file folder directory structure recommendations outlined in the Data Organization section of this Data Management SOP.

Local Archiving of Files

An overview of the processing and disposition of project data is provided in Table S11.1.

Digital Data

Any time a revision of protocols requires a revision to the database, a complete copy of the database will be made and stored in an archive directory (see Version Control, above). In addition to this copy in its native database format, all tables will be archived in a comma-delimited ASCII format that is platform-independent by using the Access_to_ascii.mdb utility developed by CAKN. An abbreviation of benthic marine monitoring, the database version number, and the term Archive are incorporated into the filename of archives, with ascii files including a .txt file extension and Access files including an .mdb file extension. These files are saved in the 'version_archive' directory, with a subdirectory created for each version. Example:

\version_1.0\marine_benthic_BE_v1.0_Archive_001.mdb
\version_1.0\marine_benthic_BE_v1.0_Archive_001.txt
\version_1.04\marine_benthic_BE_v1.04_Archive_009.mdb
\version_2.02\marine_benthic_BE_v2.02_Archive_013.txt

At the end of a field season when all data have been entered, verified and validated, an annual archive copy of the front and back end database files will be made (see Version Control section). Archived datasets will be managed by PACN, and made available to the parks, PI, and to cooperators as needed. These archived files will be stored on the PACN file server, regularly backed-up, including an off-site storage rotation, and will be distributed in read-only format.

Hard Copies and Originals

Hard copies of all datasheets, maps, and protocols will be archived with the PI, at the park, and within the park museum/library. Table S11.1 identifies the processing and disposition of the project data.

Table S11.1. Processing and disposition of project data.

Data item	Action	Database	Project Binder	Archive
Benthic data form	Make two copies	Enter data	Park copy	Investigator keeps original. Museum keeps one copy
Growth data form	Make two copies	Enter data	Park copy	Investigator keeps original. Museum keeps one copy
Recruitment data form	Make two copies	Enter data	Park copy	Investigator keeps original. Museum keeps one copy
Digital photographs of quadrats		Enter digital file name and storage location in database	Park copy of list of digital file names, date, and file size	Electronic copy on backed-up park server on hard disk. If unavailable then marine ecologist is responsible for storage.
Maps	Download GPS data	Enter new location data	Original	Museum keeps an annual copy
Protocol update		Create archive copy (both Access and ASCII) before update		Save on backed-up server or hard drive; send copy to PACN
End-of-season verified/validated database		Create archive copy		Save on backed-up park server or hard drive. If unavailable then marine ecologist is responsible for storage. Send copy to PACN
Metadata	Complete or update dataset catalog record		Park copy	Send updated record to PACN

APPENDIX D. PACN VITAL SIGN PROTOCOL SOP EXAMPLE – CORAL RECRUITMENT ANALYSIS SOP FROM BENTHIC MARINE COMMUNITY PROTOCOL

Revision Log

Only changes in this specific SOP will be logged here. Version numbers increase incrementally by hundredths (e.g., version 1.01, version 1.02) for minor changes. Major revisions should be designated with the next whole number (e.g., version 2.0, 3.0, 4.0). Record the previous version number, date of revision, author of the revision, identify paragraphs and pages where changes are made, who approved the revision, and the reason for making the changes along with the new version number.

New Version Number	Revision Date	Author	Changes Made	Reasons for Change	Network Regional Reviewer Approval	Previous Version Number

Purpose

This SOP details the processing of CRAs using bleach and the methodology for locating, identifying, counting and documenting coral recruits. This SOP also includes procedures for cleaning the tiles in preparation for the next deployment. Data handling and quality assurance/quality control for all monitoring protocols implemented by the PACN monitoring program are detailed in the program's Data Management Plan posted on drive I: at the HAVO data server. Microsoft Access is the primary software environment for managing benthic data and the associated metadata. ESRI ArcInfo 9 serves as a tool for validation of spatial data residing in Access.

A master equipment list for the entire Benthic Marine Community Vital Sign Monitoring Protocol can be found in protocol SOP #1 "Before the Field Season." The master

equipment list should be updated simultaneously if any SOP requiring an equipment list is revised.

Laboratory Analysis of Coral Recruitment Plates

This SOP is to be followed once Coral Recruitment Arrays (CRAs) have been collected from the field, properly stored, and returned to the laboratory for analyses (see protocol SOP #8 “Conducting Coral Recruitment Sampling”). The time required to complete this procedure takes approximately two days, including preparation time. The equipment needed for the preparation and analysis of the recruitment plates can be found in the protocol (included as an appendix to this SOP). The coral recruitment data forms (see protocol SOP #6 “Selecting and Marking Subtidal Sampling Transect Locations”) used during the collection of the CRAs should also be available for consultation.

Preparing Plates

The preparation of the recruitment plates is an important step and should be completed, with no delay, immediately as part of the post dive procedures. Do not allow the plates to sit overnight before processing. Handle the plates as little as possible and only by the edges. Excessive handling can damage the coral recruits.

Step 1: Prepare a 10% bleach solution by mixing 1 part of standard household bleach (Sodium Hypochlorite) with nine parts of water. Pour enough solution into plastic trays so the tiles will be completely immersed.

Step 2: Place the plates in trays filled with the bleach solution.

Step 3: Leave plates in the bleach solution for at least 24 hours. Remove from the bleach and rinse off in fresh water to remove algae and sediment.

Step 4: Optional: Take photos of each plate after bleaching. See protocol SOP #12 “Data Management – Photographs” for more information on photo storage.

Step 5: Allow the plates to air dry. Ensure that site code is legible. If necessary retouch the site code with an ultra fine Sharpie marker.

Step 6: Wrap each tile in bubble wrap and store in box in a safe place until microscopic examination is possible.

Locating, Identifying, and Counting Coral Recruits

The following procedure outlines the routine used to find, identify, and count coral recruits on the dried terra cotta plates. The amount of time needed to complete one plate varies, and could differ by season and by park. Pay particular attention to crevices/cracks and edges of tiles – Montiporids tend to occur in the crevices.

Step 1: Set up a dissecting microscope mounted on a stable “arm” stand according to its instructions. Set up in a comfortable location. Make sure lighting is adequate.

Step 2: Using the dissecting microscope at 6-10X magnification, systematically search the top then bottom plate surfaces. “Raster scan” entire tile twice, once side-to-side and once top-to-bottom. Keep track of where you are on the tile using the permanent marker to make hash marks on the periphery.

Step 3: For each coral recruit encountered, identify it to the lowest possible taxonomic level (typically genus) and record it in the appropriate location on the datasheet (see example datasheet in protocol, included as an appendix to this SOP). Make sure that the appropriate site code, CRA number, plate number, and surface (upper top, upper bottom, lower top, lower bottom, and side; see details in protocol SOP #8 “Conducting Coral Recruitment Sampling”) has been noted for each observed recruit. Record the approximate dimensions of each recruit (in millimeters) using a calibrated ocular micrometer for particular magnification settings.

Step 4: To keep from recounting recruits, circle each recruit that has been tallied with a blue sharpie fine point pen.

Step 5: Once all plate surfaces have been searched, set the plate aside. The plate will need to be cleaned prior to redeployment.

Step 6: Review your work by quickly scanning plates, taking care to note site locations, CRA numbers, species identifications, and plate surface for each coral recruit.

Entering Data

For detailed procedures of data entry into the Benthic Marine Database, including quality control and verification, please see protocol SOP #11 “Data Management.”

Cleaning Plates

Once all plates have been searched, data has been entered and verified, plates should be cleaned and stored for re-deployment at a future date.

Step 1: Completely submerge plates in household vinegar for at least 24 hours or until all CaCO3 material has dissolved from the plate surface. This can be confirmed through a thorough visual inspection of the plate surface using a microscope or a hand lens.

Step 2: Use a toothbrush or other soft plastic bristle brush to clean difficult spots. Rinse in freshwater. Dry for 24 hours in a clean place.

Step 3: Wrap each tile in bubble wrap and store until needed.

Create Photo-Guide of Coral Recruits

It is advisable to create a photographic guide of each park's coral recruits. This guide should be assembled during the first year of monitoring and serve as a living document that is updated over time. The guide should include photographs of each coral recruit taxon observed over the course of the year. In most cases, multiple individuals of each species should be photographed to account for natural variation. The photos should be assembled into a printable guide. Each photograph should be large enough to illustrate the recruit and be accompanied by a caption detailing the taxon name, the time period the individual was collected, and an appropriate scale bar. Prior to including any photos in this guide, confirm the species identification by using the published scientific literature or a local expert. Because coral recruits can be easily confused with other marine organisms, it is helpful to include examples of the common bryozoans and foraminifera. This guide is invaluable for training new observers and confirming identifications. Please see protocol SOP #2 "Training Observers" for an updated list of park photographic guides. Protocol SOP #12 "Data Management – Photographs" shows an example of coral recruits from Hawaii.

Data Verification

Prior to sending the working copy of the database to the PACN database manager it is important to check both the number of recruits and taxon identifications on at least 10% of the tile pairs. This effort equates to five tile pairs annually. Typically the marine ecologist for each park conducts the initial examination of the tile pairs. Therefore, the marine ecologist is responsible for selecting another natural resources staff member to complete the re-examination of the tile pairs.

Step 1: After the marine ecologist has finished examining the 45 tile pairs for their park he randomly selects five tile pairs for re-examination by another natural resources staff member (usually a marine biotechnician).

Step 2: Random selection can be done using the Excel "RANDBETWEEN" function or simply picking points out of a hat. For example, in Excel assign the 45 tile pairs to 45 cells and label them sequentially from one to 45 in another column. Then in a third column highlight five cells and enter the formula =Randbetween(1,5). Select the tile pairs whose number is generated in the five cells.

Step 3: The person conducting the re-examination verifies that the correct number of recruits have been recorded on the datasheet (see protocol SOP #12 "Data Management – Photographs") corresponding to the premarked tiles. In addition, taxon identification should also be validated.

Step 4: If the measurement error is less than 10% for both the incorrect number of recruits AND the number of misidentified recruits then proceed to step #7.

Step 5: If the measurement error is greater than 10%, either for the incorrect number of recruits OR the number of misidentified recruits, then the initial examiner must re-examine the remaining tile pairs.

Step 6: New sets of five tile pairs will be subsequently selected and re-examined by the natural resources staff member until the measurement error is less than 10%.

Step 7: All corrections should be completed on the working database copy for that park which will subsequently be sent off to the PACN database manager.

APPENDIX E. TABLE OF PACN GIS COVERAGES

***PACN GIS Coverage Table**. A list of GIS coverages held by PACN, with indication as to which parks are covered.*

GIS Layer	West Pacific		South Pacific	Hawaii							
	WAPA	AMME	NPSA	USAR	KALA	HALE	ALKA	PUHE	KAHO	PUHO	HAVO
Base Cartography											
USGS Quads (DRGs v1)	X		X	X	X	X	X	X	X	X	X
USGS Quads (DRGs v2)	X	X		X	X	X	X	X	X	X	X
Park Boundary	X	X	X	X	X	X	X	X	X	X	X
Hydrography	X	X	X	X	X	X	X	X	X	X	X
Digital Elev Model (DEM)	X	X	X	X	X	X	X	X	X	X	X
Hypsography (scanned contours)		X		X	X	X	X	X	X	X	X
Transportation (roads & trails)	X	X	X	X	X	X	X	X	X	X	X
Shoreline	X	X	X	X	X	X		X	X	X	X
Nautical Charts			X	X	X	X		X	X	X	X
Boundaries											
Land Ownership	X			X	X	X	X	X	X	X	X
Ahupua`a (land divisions)					X	X3		X3	X3	X3	X3
Natural Resource Data											
Vegetation (not to Nat'l Standards)	X	X	X		X1	X1, X2	X		X1	X	X
Fire											X
Wetlands	X	X	X2	X	X	X	X	X	X	X	X
Rainfall				X	X	X	X	X	X	X	X
Temperature							X	X	X	X	X
PRISM Climate Data	X2	X2	X2	X	X	X		X	X	X	X
NPS Climate Stations											X
Air Quality Stations											
Geology (surficial)	X	X	X4					X	X	X	X
Soils	X	X	X	X	X	X	X	X	X	X	X
Landcover	X	X	X2	X	X	X	X	X	X	X	X
Land Use		X	X	X	X	X	X	X	X	X	X
Anchialine Pools									X		
Tsunami Inundation				X	X	X	X	X	X	X	X
Flood	X		X	X	X	X	X	X	X	X	X
Lava Flow Hazard								X	X	X	X
Watershed		X	X		X	X		X	X	X	X

GIS Layer	West Pacific		South Pacific	Hawaii							
	WAPA	AMME	NPSA	USAR	KALA	HALE	ALKA	PUHE	KAHO	PUHO	HAVO
Satellite Imagery											
IKONOS	X		X		X	X2		X	X	X	X2
SPOT (10m Pan, 20m MS)				X	X	X	X	X	X	X	X
Landsat (30m MS)				X	X	X	X	X	X	X	X
Quickbird (USFS, Fall 2002)			X4					X	X	X	
Multispectral – 1m (Terrasystems)								X	X	X	
Orthophoto Quads											
DOQQs	X	X	X	X	X	X	X	X	X	X	X
EMERGE				X	X2	X2	X2	X	X2	X2	X2
Aerial Photos (georegistered)											
NOAA Coastal Survey (2000)								X	X		
True Color 1:1200 (2002)			X2								
Infrared 1:12000 (1994)			X								
True Color 1:8000 (2002)			X4								
Infrared 1:24000 (2003)			X4								
Natural Color 1:5000 (2004)								X4	X4	X4	
Natural Color 1:10000 (2004)							X				
Color IR 1:5000 (2003)									X		
Natural Color 1:5000 (2003)									X		

X1 = Data may be outdated or inaccurate.
X2 = Partial coverage.
X3 = needs to be compiled from USGS DLG Boundary data.
X4 = Expected to be delivered by contractor or partner.

APPENDIX F. PACN SOP EXAMPLE – CREATING METADATA

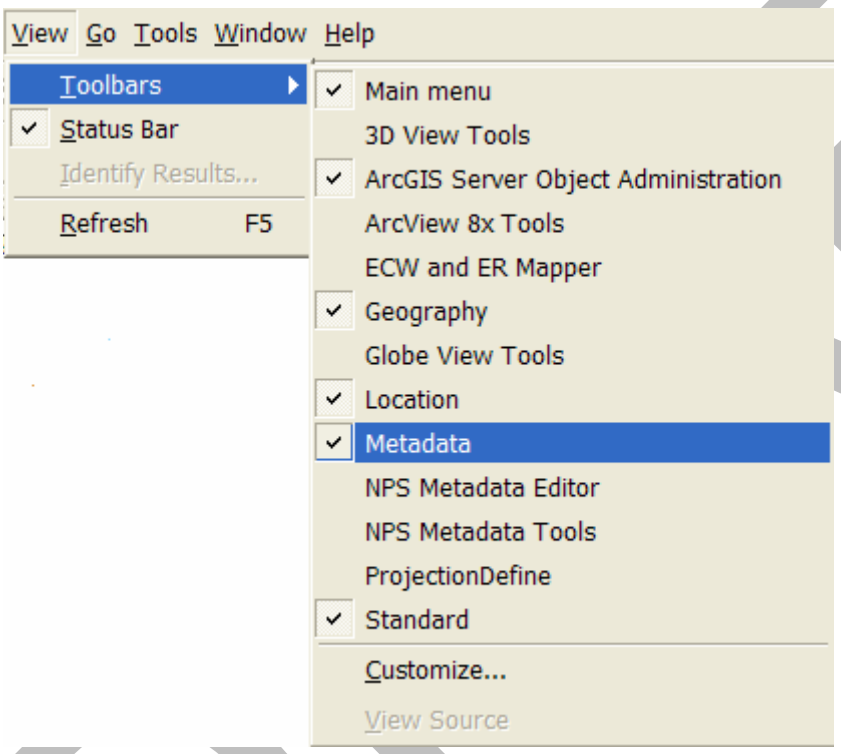
Creating Metadata SOP – Part 1 (Using ArcCatalog)

DRAFT March 2006

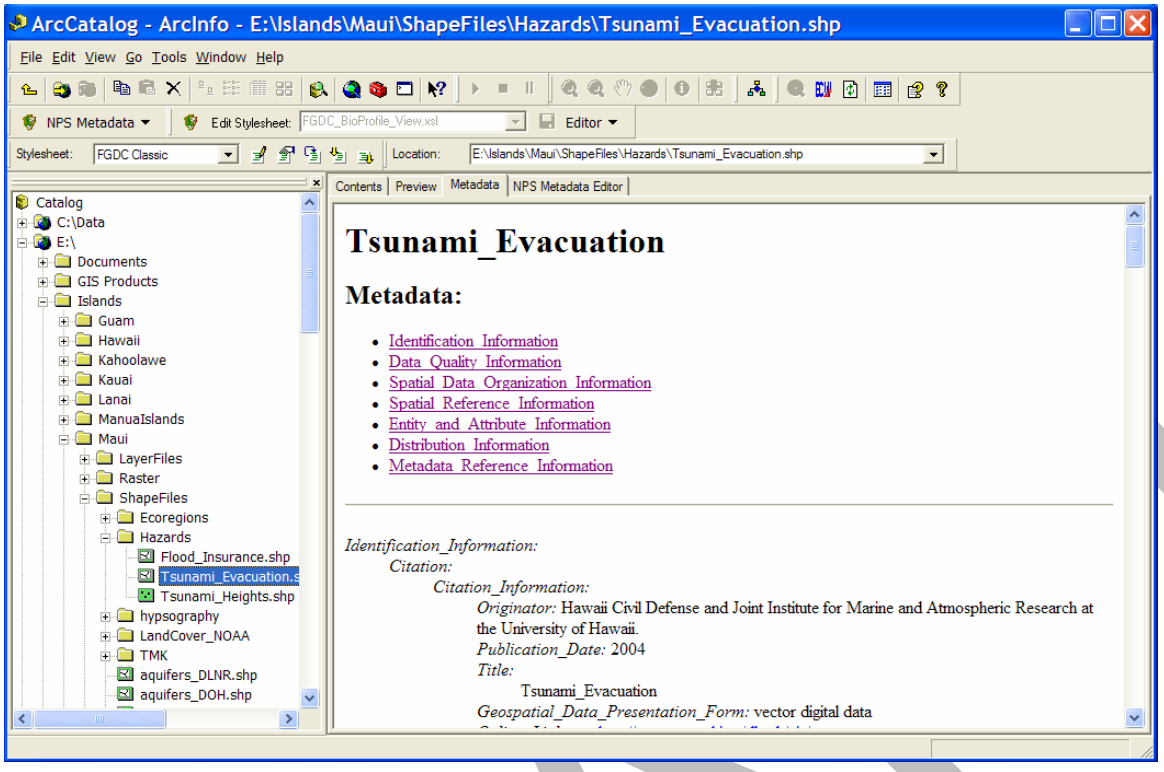
Developed by Sandy Margriter, GIS Specialist, PACN/PWRO-Honolulu

Before proceeding check to make sure that the Metadata toolbar (shown below) is available from within ArcCatalog.

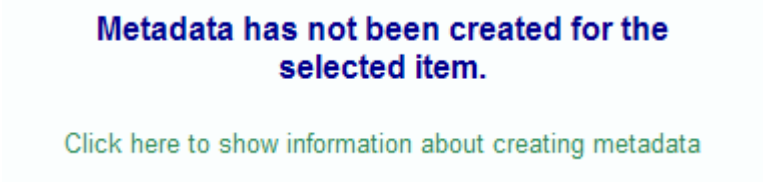
If the toolbar is not present, click View | Toolbars | Metadata (as shown below).



When you click on a filename in the "Catalog Tree" or Table of Contents in ArcCatalog, the metadata for that file will appear in the main ArcCatalog window on the right (as shown below.) Note that the table of contents can be toggled on and off by selecting Window | ArcCatalog Tree from the Menu Bar.



If no metadata exists for the selected file then you will send the following:



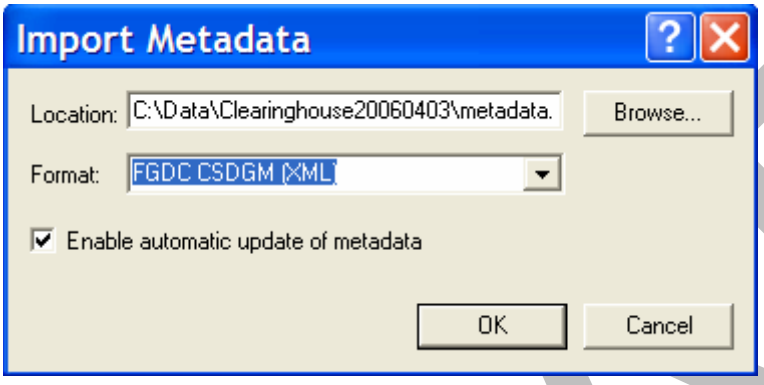
Importing Metadata

When creating metadata from scratch it's best to start by first importing the following NPS "Data Distribution Liability Statement."

> The National Park Service shall not be held liable for improper use of the data described and/or contained herein. These data and related graphics ("GIF" format files) are not legal documents and are not intended to be used as such. The information contained in these data is dynamic and may change over time. The data are not better than the original sources from which they were derived. It is the responsibility of the data user to use the data appropriately and consistent within the limitations of geospatial data in general and these data in particular. The related graphics are intended to aid the data user in acquiring relevant data; it is not appropriate to use the related graphics as data. The National Park Service gives no warranty, expressed or implied, as to the accuracy, reliability, or completeness of these data. It is strongly recommended that these data are directly acquired from an NPS server and not indirectly through other sources which may have changed the data in some way. Although these data have been processed successfully on a computer system at the National Park Service, no warranty expressed or implied is made regarding the utility of the data on another system or for general or scientific

purposes, nor shall the act of distribution constitute any such warranty. This disclaimer applies both to individual use of the data and aggregate use with other data.

In the Metadata Import dialog box select FGDC CSDGM (XML) as the format type.

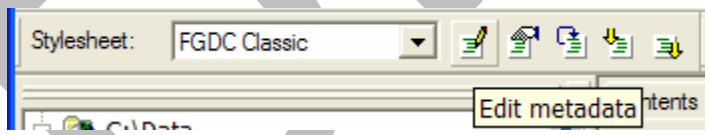


Click the "Browse" button and select the MetadataDistribution.xml file from the M:/Documents/SOPs/Metadata/Templates directory. Note the Drive letter (in this case M:) will depend on how you "mapped" your computer to the PWR GIS data directory.

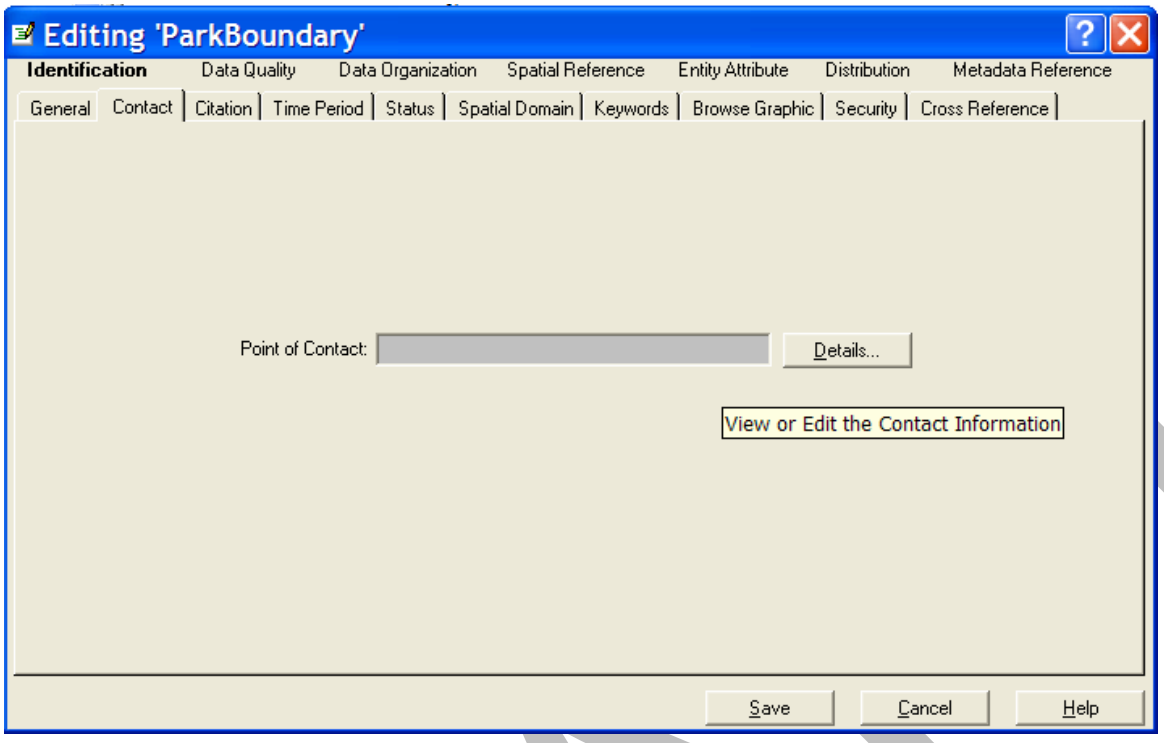
Leaving the "Enable automatic update of metadata" checked will automatically update your metadata with the title, spatial reference information (if the projection has been defined), and the names of the attribute fields in the table.

Edit Metadata

Edit your metadata in order to add your contact information in the ….



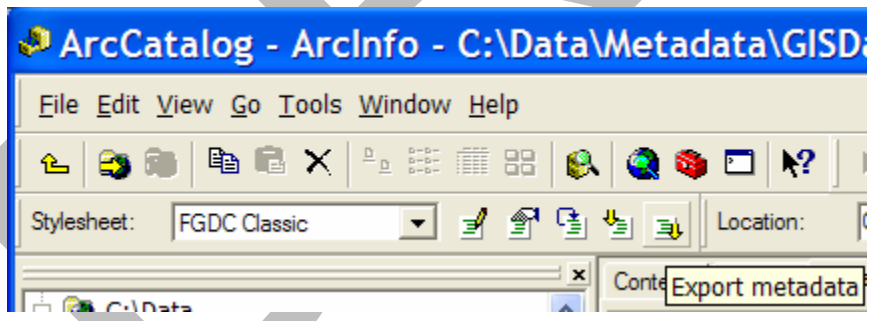
When the Editing dialog box appears, click the "Contact" tab, then click the "Details" button.



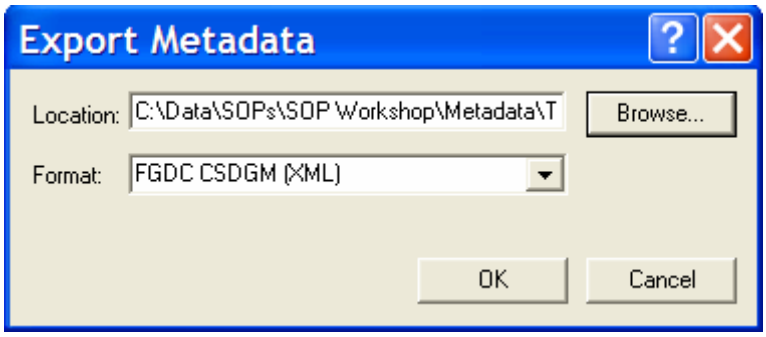
Enter your contact information in the fields provided (name, organization, telephone number, email address and mailing address). Click the "Address" tab to add your mailing address. Click OK, then Save.

Export Metadata

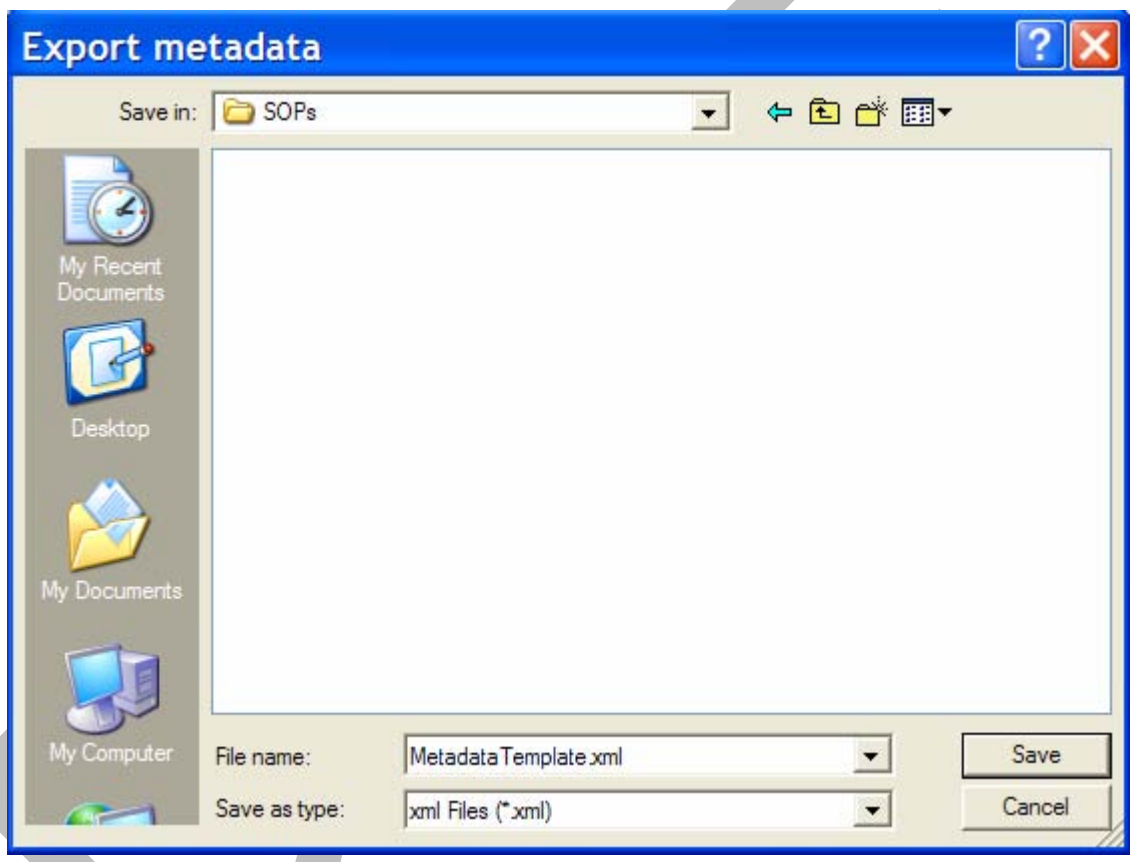
ArcCatalog can be used to export metadata to the following formats: XML, HTML, SGML, and TXT. The file you export can be used as a template when creating metadata in the future.



Select FGDC CSDGM (XML) as the Format.



Click the Browse button and navigate to the directory where you would like to store the exported file (in the “Save in” dropdown menu) and type in the name of the file (in “File name” at bottom of screen.



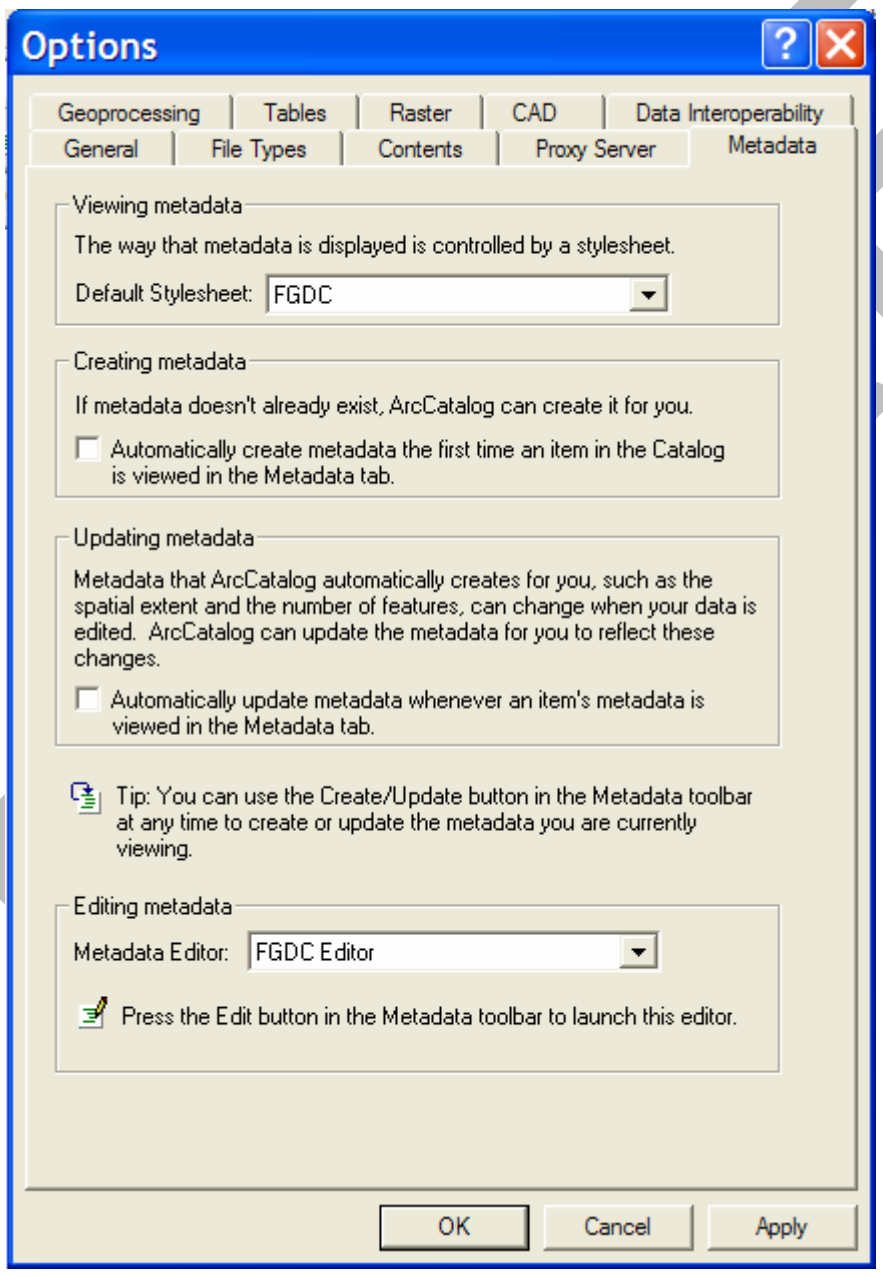
Creating Metadata SOP – Part 2 (Using NPS Metadata Tools and Editor)

DRAFT March 2006

Developed by Sandy Margriter, GIS Specialist, PACN/PWRO-Honolulu

Creating Metadata

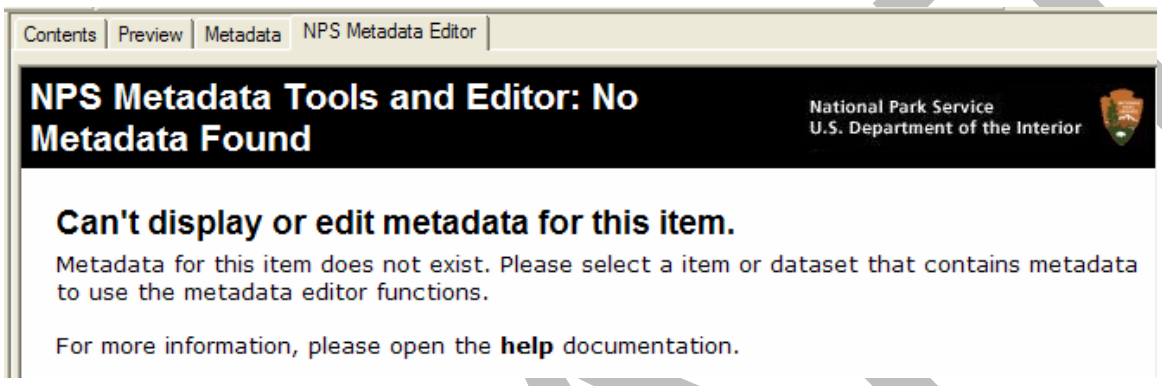
Change the default metadata options. Click on Tools, then Options, then on the Metadata tab. Uncheck the boxes to the left of "Automatically create metadata the first time an item in the Catalog is viewed in the Metadata tab" and "Automatically update metadata whenever an item's metadata is viewed in the Metadata tab." This will give you more control over creating and updating metadata. Also, go ahead and choose "FGDC" as your Default Stylesheet, as shown below.



Create a Personalized Template

The template you will create, using the steps below, can be used in creating all future metadata records. It will help you to avoid having to repeatedly enter contact information each and every time you create new metadata.

Select a GIS layer that does not have metadata (at this point it does not matter which file you select, as long as it does not have metadata). Your metadata stylesheet should now look like this:



We will first add the following NPS "Distribution Liability" statement to our metadata:

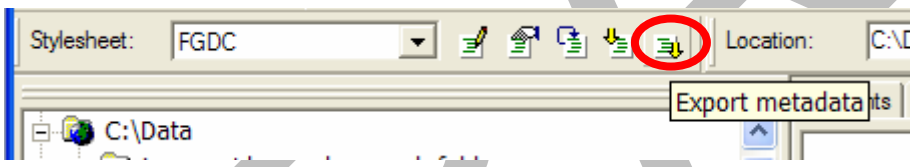
> The National Park Service shall not be held liable for improper use of the data described and/or contained herein. These data and related graphics ("GIF" format files) are not legal documents and are not intended to be used as such. The information contained in these data is dynamic and may change over time. The data are not better than the original sources from which they were derived. It is the responsibility of the data user to use the data appropriately and consistent within the limitations of geospatial data in general and these data in particular. The related graphics are intended to aid the data user in acquiring relevant data; it is not appropriate to use the related graphics as data. The National Park Service gives no warranty, expressed or implied, as to the accuracy, reliability, or completeness of these data. It is strongly recommended that these data are directly acquired from an NPS server and not indirectly through other sources which may have changed the data in some way. Although these data have been processed successfully on a computer system at the National Park Service, no warranty expressed or implied is made regarding the utility of the data on another system or for general or scientific purposes, nor shall the act of distribution constitute any such warranty. This disclaimer applies both to individual use of the data and aggregate use with other data.

Import the standard NPS "distribution liability" clause using the "Import Metadata" tool in ArcCatalog.

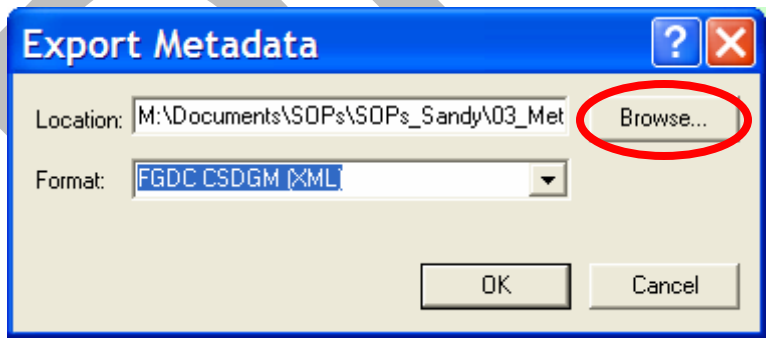
Select the MetadataDistribution.xml file in the SOP Workshop/Metadata/Templates folder and click “Open”. You should now see a screen like the one shown below. Click OK.

You are now ready to export your template for use in creating metadata in the future. There are a few options for exporting metadata. The NPS Metadata Toolbar “Export” option will automatically name the metadata file with the same name as the shapefile (but with a different extension). You can optionally use the “Metadata Template” tool to export the metadata, but this will limit what gets exported to the template. In any case, the file(s) created during the export procedure will become your metadata template file(s).

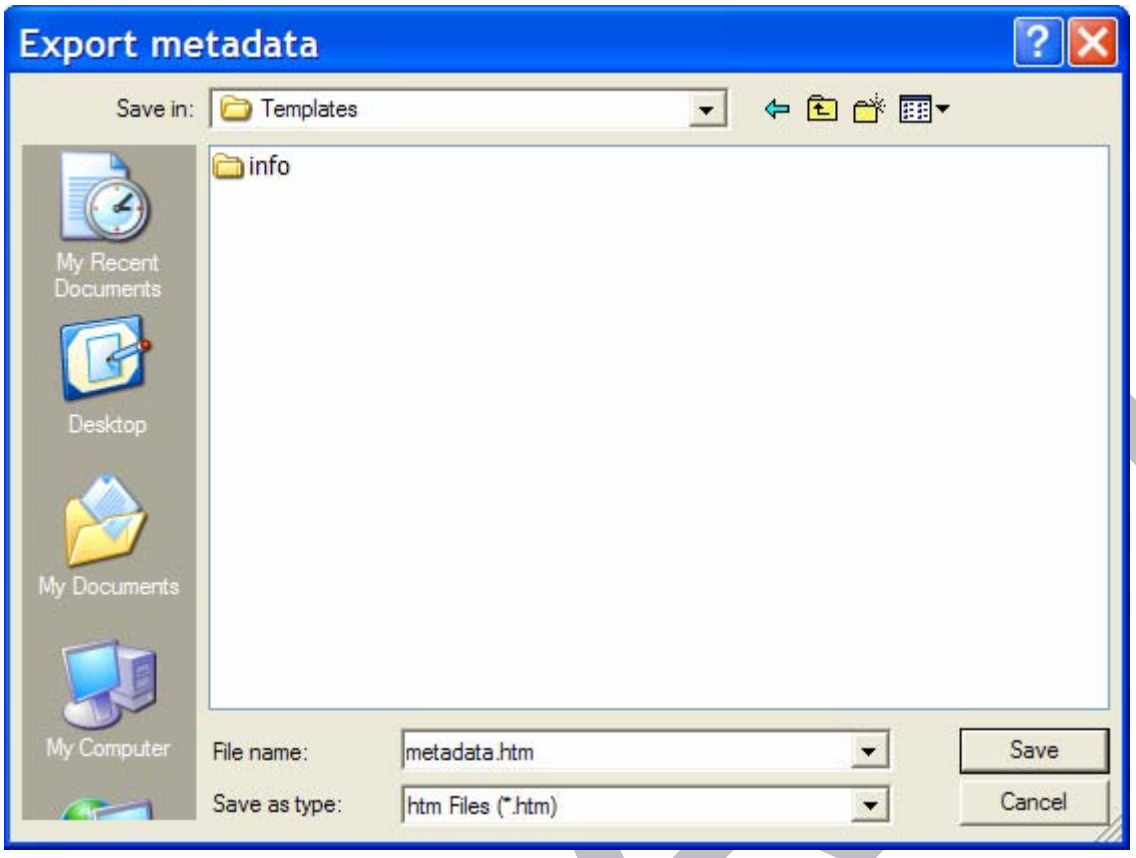
Instead, we will use the standard Metadata Toolbar in ArcCatalog to export the metadata. Using this toolbar we can provide a name for the export file. Click on the “Export” button shown below.



When the Export Metadata window pops up, use the down arrow key to select “FGDC CSDGM (XML) as the export file format.click. Then click on the “Browse” button to select a directory that will store the metadata template you are exporting. Save the file in a location that will be easy to find in the future.



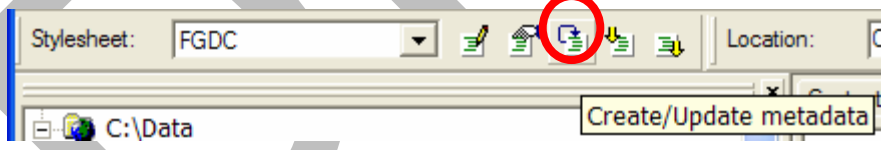
For today’s workshop…
select the SOP Workshop/Metadata/Templates directory. Type the metadata filename in the File Name dialog box.



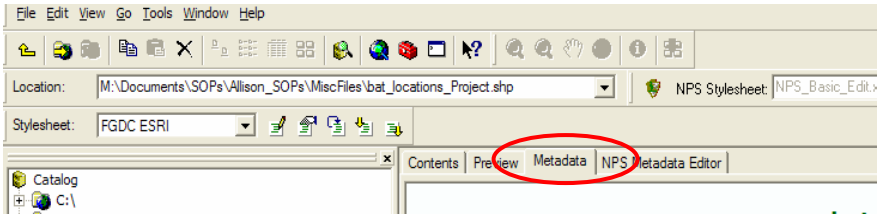
Create Metadata Using your Personalized Template

Click on the GIS layer you need to create metadata for. If metadata does not already exist for the spatial data layer, you will first need to create a metadata file that will then be updated with your personalized information. If the projection has been defined for your spatial data, this button automatically fills in the spatial reference sections of your metadata. In addition, the "Entity and Attribute" section of your metadata will be populated with the field names contained in the spatial layer's table.

Click on the "Create/Update metadata" button shown below.



Before completing the rest of your metadata, click on the Metadata tab.



Complete your metadata by filling at least the minimal amount of metadata (who, what, when, where, and how).

Using the NPS Metadata Toolbars

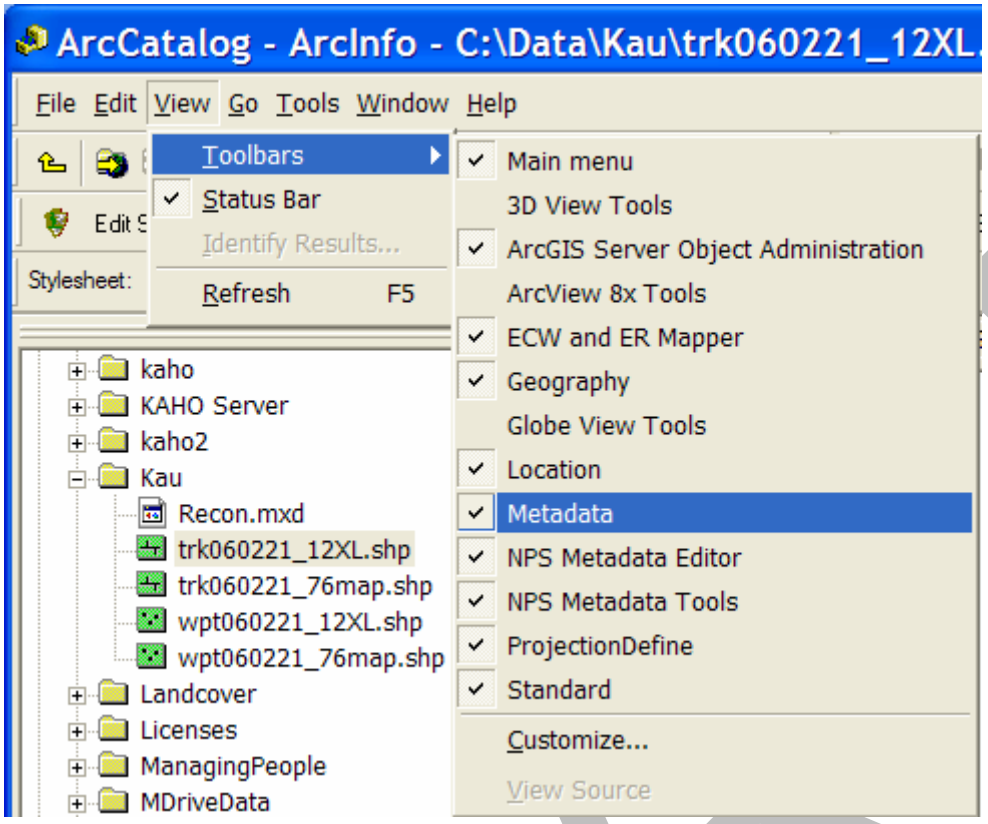
NPS Metadata Tools **NPS Metadata Editor**

The NPS Metadata Tools and Editor toolbars are designed to provide a comprehensive desktop metadata editing application. It runs both as an extension to ArcCatalog within ArcGIS and as a stand-alone application. It expands considerably on ArcCatalog's basic functionality and provides a stand-alone tool for creating, editing and managing metadata outside of the ArcCatalog environment. The application is intended to be the primary editor for metadata that will be uploaded to the NR-GIS Data Store. Currently, there are two components: the Metadata Tools and Metadata Editor (see graphic above). Unless noted (in the Help file), the features of both components are available in both ArcCatalog and the stand-alone interface.

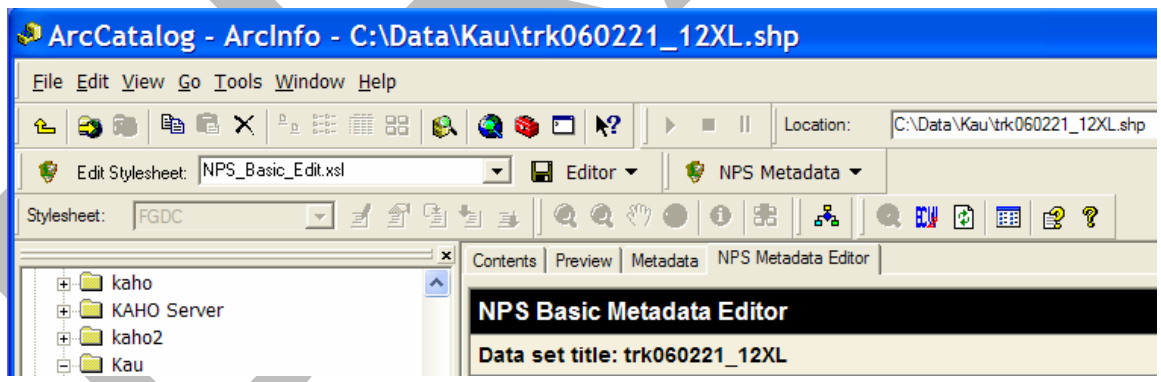
Before proceeding make sure all of the metadata toolbars have been added to ArcCatalog.

Open ArcCatalog

Select View→Toolbars and check the boxes to the left of the following: Metadata (provided with ArcGIS), NPS Metadata Editor, and NPS Metadata Tools. All checked toolbars will be available for use in ArcCatalog.



In the ArcCatalog window, click the NPS Metadata Editor tab. Then choose the "NPS_Basic_Edit Stylesheet" in the NPS Metadata Editor toolbar. Note: there are 7 stylesheets to choose from: 4 provide editing capabilities and 3 provide different ways of viewing the metadata. The NPS_Basic_Edit Stylesheet provides a simplified (but incomplete) form for creating/editing basic metadata contact and distribution information.

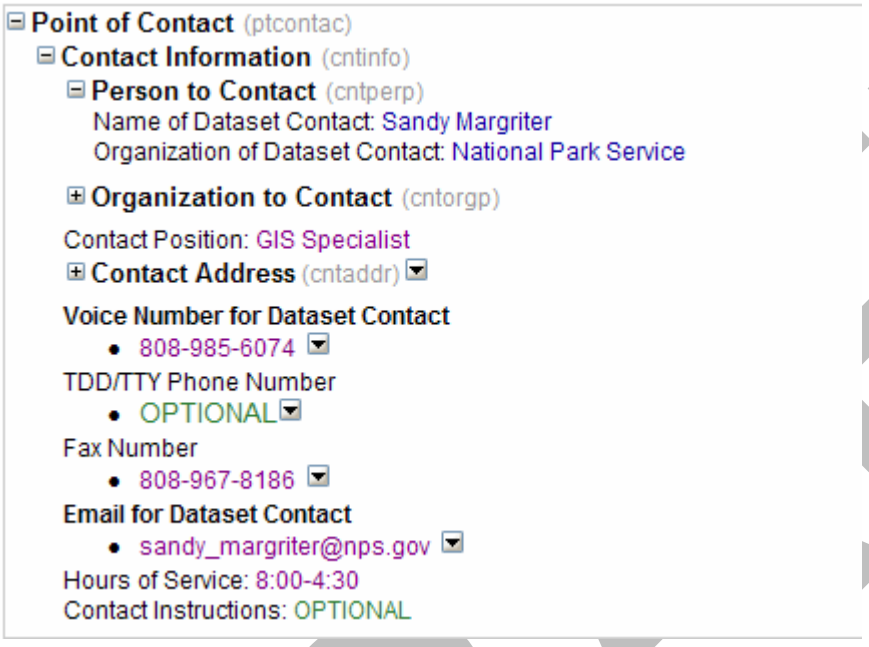


You are now ready to add contact information to complete a metadata template for future use. First simplify the Metadata style sheet by clicking "Collapse All" within the metadata window. Your metadata style sheet should now look like the one below:

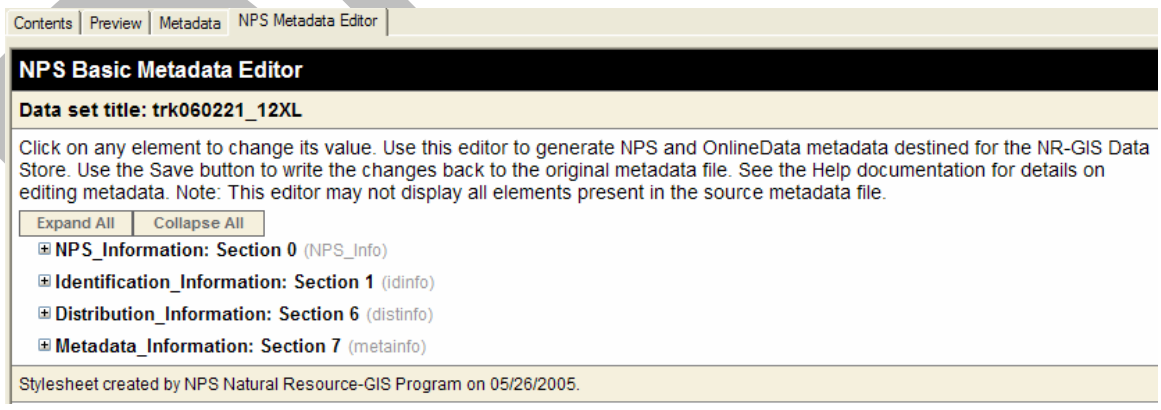
For now we are going to skip filling out the NPS Information section, which only needs to be filled out for metadata records that will be uploaded to the NR-GIS Data Store. Instead, click on

the “Identification Information” tab to expand this section. Then expand the “Dateset Citation” section, expand “Dataset Citation Details” section and type in “National Park Service” as the “Originator of Dataset.”

Edit the contact information by first expanding the “Point of Contact”, “Contact Information”, and then “Person to Contact” sections. Click the box next to “Person to Contact” and type in your name (or position title). Continue filling in your contact information. (See example below)



Follow the same procedure and fill out the Distribution_Information, Data Distributor, Data Distrubiton Contact Information and the Metadata_Information, Contact for Metadata, Metadata Contact sections with your contact information. You may later need to update the distribution information, if you plan on distributing the data on the NR-GIS Data Store internet site.



To upload data to the NR-GIS Data Store, there are certain fields that must be populated (e.g., Metadata Purpose, NPS Unit, Data Category, Data Site, and Data Steward).

APPENDIX G. PACN SOP EXAMPLE – UPLOADING GPS DATA TO GIS

Uploading GPS Data to GIS SOP – Part 1 (DNR Garmin and ArcGIS)

DRAFT March 2006

Developed by Joel Cusick, Cartographic Specialist, ASKO
and Sandy Margriter, GIS Specialist, PACN/PWRO-Honolulu

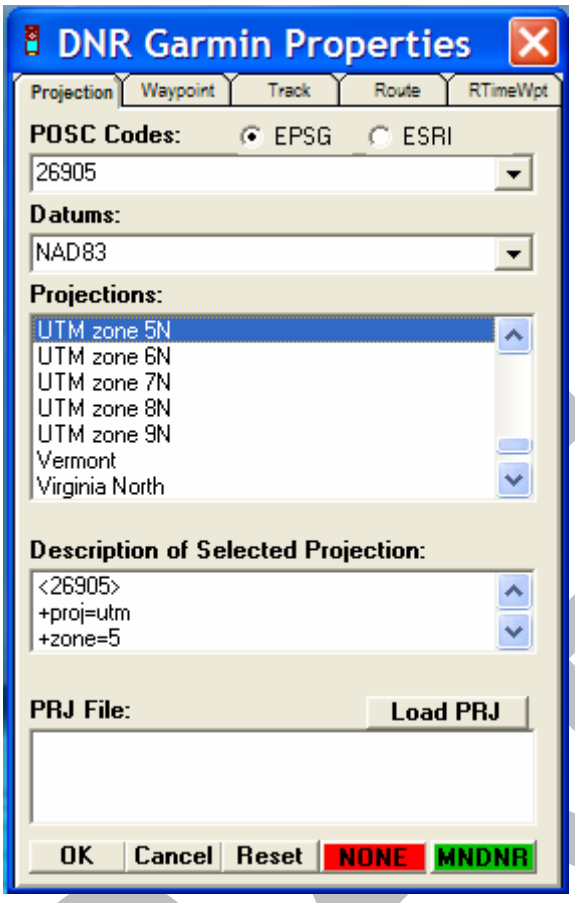
Background: DNRGarmin is free software that is the Pacific Islands NPS standard for incorporating Garmin GPS data into GIS. This document outlines the steps for creating a shapefile for display in the ArcGIS environment. Instructions for installing DNRGarmin are at the end of this document.

Hookup GPS receiver to PC – do this first to minimize com errors.

1) Attach the Garmin Cable to the COM1 (Serial) port of the PC.
2) Attach the other end of the Garmin Cable to the back of the Garmin GPS
3) Turn on the GPS by pressing and holding the red lantern button
4) Press Enter button twice to confirm the messages on screen.
5) Press the Menu button and select the "Use With GPS Off".
6) Press Enter and the Garmin should be in Simulator MODE (GPS unit no longer is "looking" for satellites).
7) Press the Page button ~ 4 times until you get o the Main Menu
8) Go into Setup | Interface and select Garmin as interface format.

Download Garmin GPS points to ArcGIS.

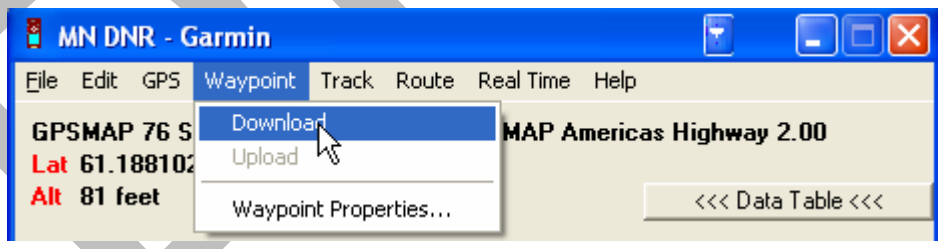
1. Launch DNRGarmin from the desktop or loading from Programs| DNRGarmin. You should see a splash screen announcing version 5.0.4.
2. When DNR Garmin is run for the first time on a computer, you will be asked to accept or change the default projection of UTM - 1983, Zone 15. Select **NO** to change the default projection.
3. The DNR Garmin Properties box appears.
4. Select the ESRI (or EPSG) radio button. In the POSC code list type in 26905 or select 26905 from the pull-down list and press ENTER. The datums/projection list will automatically change to UTM_Zone5N_NAD83. Before pressing OK, check the parameters are set as follows.



These settings are particular to Hawaii Island.

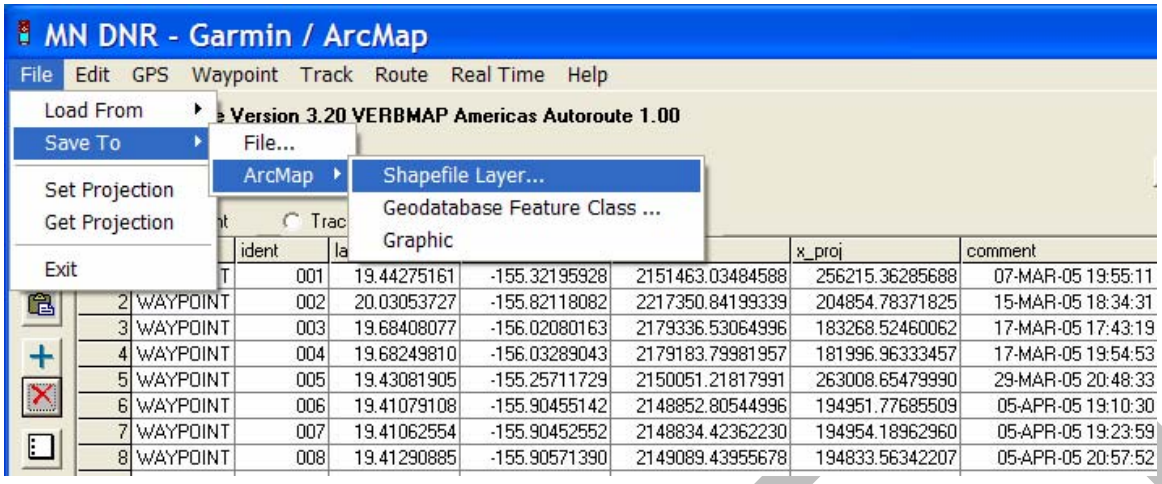
Contact your GIS specialist and ask what projection and Datum you must use

5. When Finished Press OK. These projection settings can be checked at any time by selecting File | Set Projection.
6. Check that your Garmin is connected to your PC and the DNRGarmin interface is set to Garmin.
7. Download the waypoints into DNRGarmin by selecting Waypoint | Download. Your waypoints will be downloaded in tabular format.



8. Launch ArcMap in order enable to the save to ArcMap option.

9. You can now save the points by selecting File | Save To | ArcMap | Shapefile Layer.



10. Name the shapefile using the following naming convention:
Project Name_Date of Collection (YYYYMMDD)_Feature Type (pt-point, ln-line, ply-polygon)_Garmin GPS Unit #. For example, bat GPS point data collected on July 25th, 2005 using the I&M Garmin GPS unit # 2 would be named:
bat_20050725_pt_g2.shp

11. Press the Save Button to store the shapefile.
12. Download Tracks. Select Track download.
13. Save tracks by selecting File | Save To | ArcMap | Shapefile Layer. Use naming conventions from step 10.

Installation of DNR Garmin

1. If a previous version of DNR Garmin exists, go to **Step 2**. Otherwise go to **Step 6. Failure to cleanly remove earlier versions can be disastrous.**
2. Uninstall any previous versions of DNR Garmin by using the Start | Control
3. Panel | Add/Remove Programs tool.
4. Perform a search on local drives for all files named dnrgarmin.* If found, delete all files.
5. Navigate to the directory holding the program install. This may be on M:\Software\ArcGIS9.1_Extensions\dnrgarmin51setup_9x (Pacific Islands NPS USERS) or alternatively you may download the **DNR Garmin 5.0.4Setup** program from the Minnesota Department of Natural Resources http://www.dnr.state.mn.us/mis/gis/tools/arcview/extensions/DNRGarmin/DNRGarmin.html
6. Double click the dnrgarminsetup.exe file.
7. Press Next at the Welcome Screen, then accept the license agreement. Press Next to continue.

8. The program will install itself on “c:/program files/dnrgarmin” folder. It is suggested that you keep this location. Press Next to continue.
9. The program will install both a desktop shortcut and a DNR Garmin folder will be added to the Start | Program Files menu. Press Finish and view the readme file supplied with the program. Close Readme File.

Installation Troubleshooting Guide:
The most common problem with GPS software is in the communication between the GPS and PC. Here are some steps to follow if you receive a failure message to communicate with the GPS on COM 1.

1. Check all cables. If the PC has two com ports, ensure you are attached to COM1. If COM1 is taken over by another cable Go to Step 3.
2. Ensure the GPS is in Garmin Interface Mode. On the Garmin, press MENU MENU, select Setup and go to the Interface Tab. Ensure Serial Data format is Garmin.
3. If no connection is made and the COM1 port is taken up by another cable, go to GPS | Set Port and select Port 2. Connection should be found.
4. If you are using a USB cable, try instead using the cable with 4 prongs on one end and a parallel port on the other.
5. If DNR Garmin is not recognizing any port, then this is a fail safe approach. Note, this will require a REBOOT at end of sequence.

FAILSAFE RESET for COM ports. This may be a bit scary at first, but the premise here is WINDOWS operating systems must always have a COM1 port. If all else fails in steps 1-4 above, follow these instructions (written for XP).

1. Close DNR Garmin, ArcView and Turn off GPS. You may leave it connected to the COM1 port
2. Click on Start | Settings | Control Panel | Administrative Tools
3. Double-Click Computer Management
4. Double-Click Device Manager
5. Open up Ports and select COM 1
6. Right mouse over COM 1 and select Disable.
7. Say yes to really disable this.
8. Press OK or Apply all the way out to close all windows.
9. Restart the PC
10. When PC is rebooted, Windows will re-establish the COM1 port. Proceed on steps 1 above

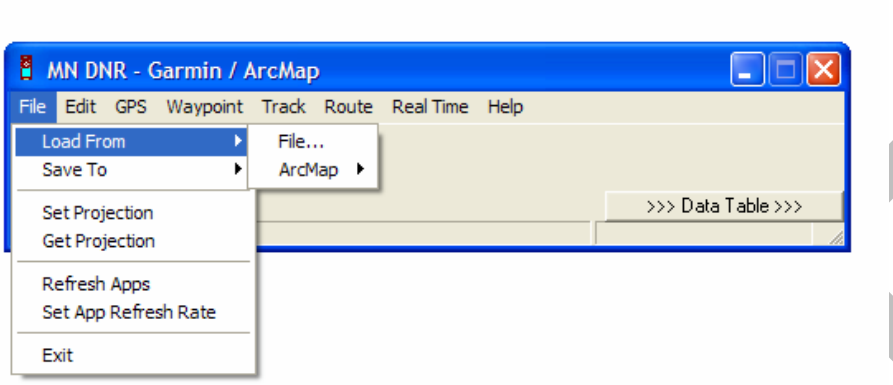
Other Trouble-shooting steps:
1) You must have administrative privileges to install this program in the c:/program files directory. Contact your system administrator if installation was stalled because of installation rights.
2) If you have an earlier version of the DNR Garmin program on your machine you may need to uninstall before installing the new version. Uninstall previous versions by using the Start | Control Panel | Add/Remove Programs tool.
3) If GPS data does not align with NPS base data, then you may be projecting incorrectly. For NPS users in Hawaii using ArcGIS 9.1, ensure the projection in

DNRGarmin is set to ESRI format 26905. Check by opening DNRGarmin, selecting File | Get Projection. You should check that the Projection is set to UTM_Zone5_North_american_NAD83 (for Hawaii Island).

4) For other installation errors please refer to Minn. DNR Help page at http://thoreau.dnr.state.mn.us/mis/gis/tools/arcview/Training/WebHelp/Training.htm#DNRExt/DNRGarmin4/GPSIntro.htm

Adding Points from ArcGIS to Garmin GPS:

Connect GPS to computer. Open DNR Garmin.



From ArcMap:
ArcMap will only be an option if an ArcMap project is open. If ArcMap is open, you can highlight the file to Load. Then in DNR Garmin, go to File→ Load From→ ArcMap→ Layer.

From File:
This will open a browse window where you can search for your shapefile. Be sure to select the correct "Files of Type" since shapefile is not the default. Go to File→ Load From→File.

An Identify Fields dialog may open. Fill in the appropriate fields (???). Click OK and the data will be moved to DNR Garmin. You will now see a list of waypoints, tracks, or routes.

From the Waypoint, Track, or Route menu, select Upload and data will be uploaded to the GPS.

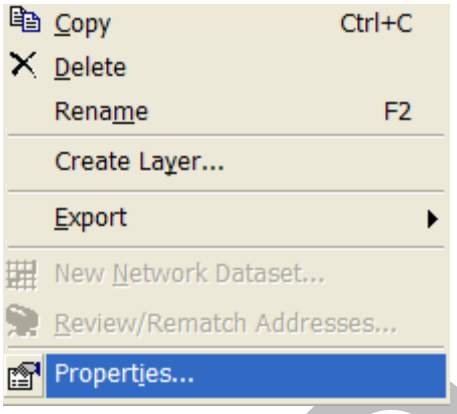
Uploading GPS Data to GIS SOP – Part 2 (Defining a Projection in ArcGIS)

DRAFT March 2006

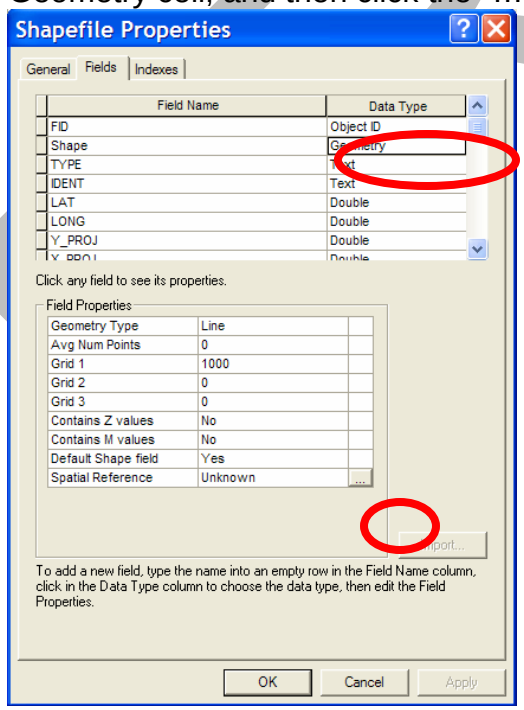
Developed by Sandy Margriter, GIS Specialist, PACN/PWRO-Honolulu

ArcGIS has no way of knowing a file's coordinate system unless a projection is first defined or created for that file. For shapefiles a .prj file is created in the process of defining the projection / coordinate system. In this example, our spatial data layer is stored in the UTM zone 5, NAD83 projection, but the shape file lacks the .prj file and needs to be created using the following steps.

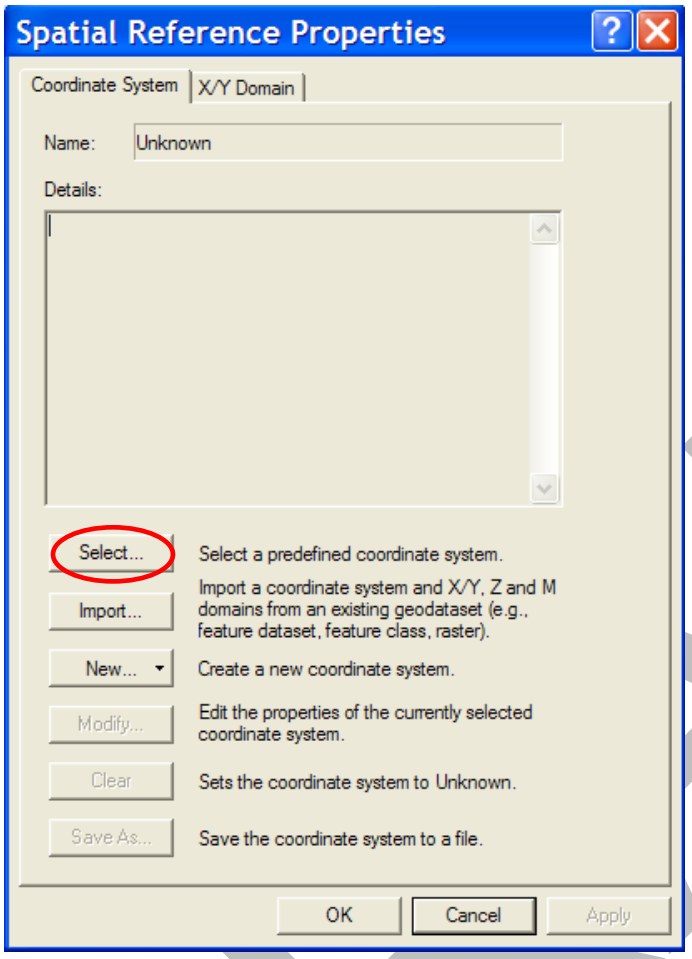
1. Open ArcCatalog and select your shapefile.
2. Right click on shapefile filename and click Properties.



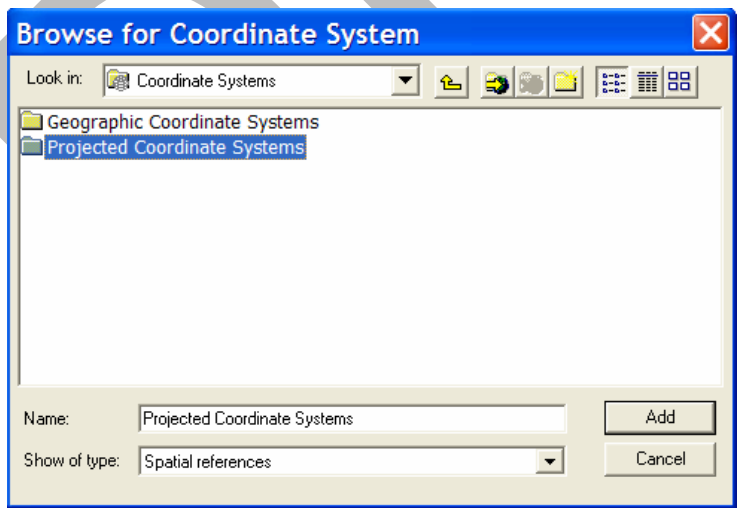
3. In the Shapefile Properties dialog, under the Fields tab, click on the shape's Geometry cell, and then click the "…" button next to "Unknown" Spatial Reference cell.



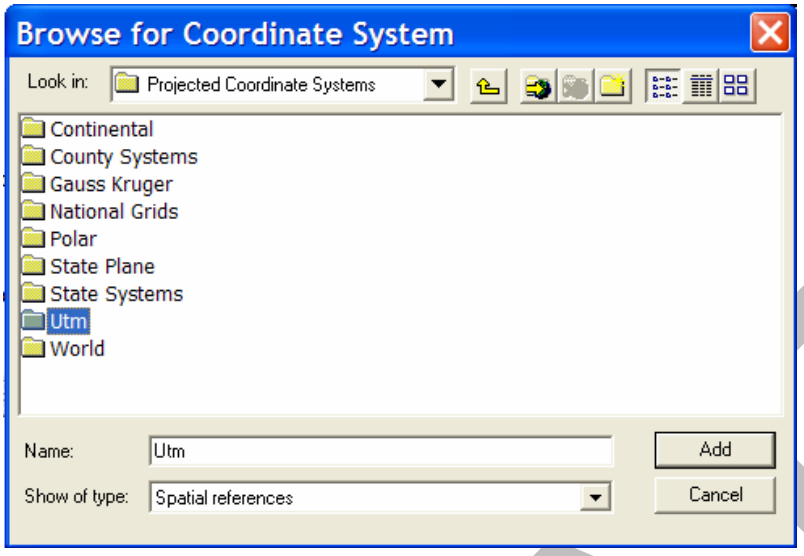
4. In the Spatial Reference Properties dialog, under the Coordinate System tab, click the Select button to select a predefined coordinate system.



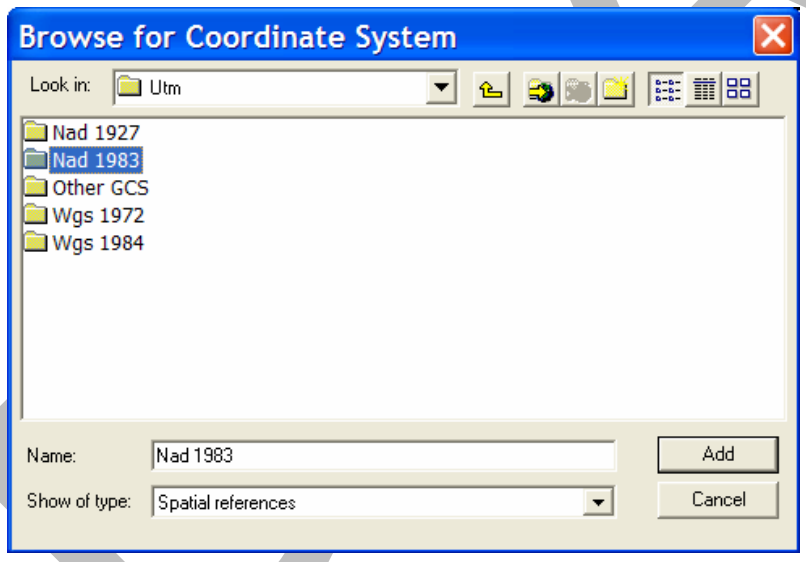
5. Then, in the Browse for Coordinate System dialog, double click “Projected Coordinate Systems”.



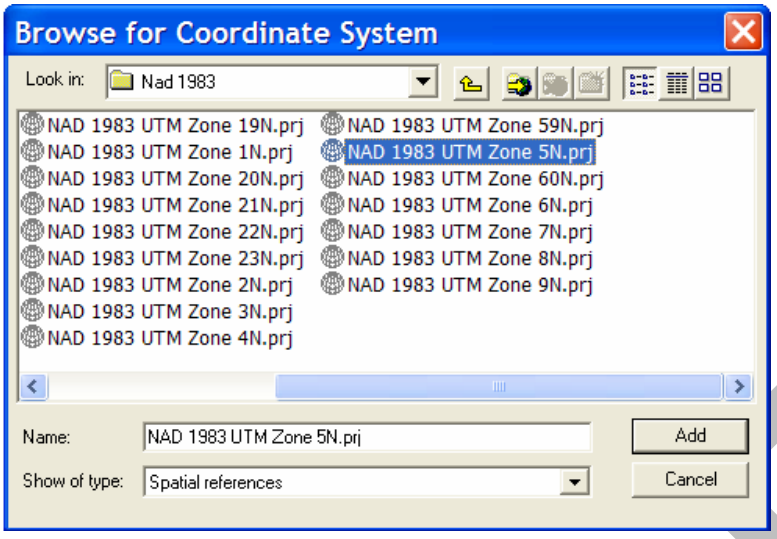
6. Double click on “UTM” folder.



7. Then double click “Nad 1983”



8. Select “NAD 1983 UTM Zone 5N.prj” from the list (as shown below).



9. Click “Add”. Then click “Apply”. And click a series of OK buttons to complete the process.

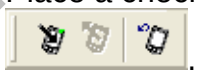
APPENDIX H. PACN SOP EXAMPLE – USING ARCPAD WITH TRIMBLE GEOXT GPS UNIT

Trimble GeoXT / ArcPad SOP

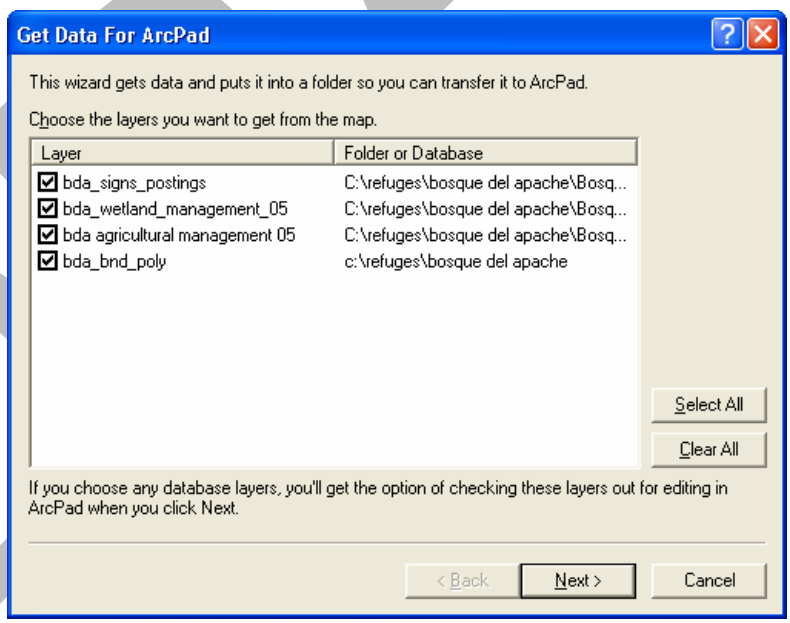
DRAFT March 2006
Developed by Tom Kochanski, GIS Technician, HAVO
and Sandy Margriter, GIS Specialist, PACN/PWRO-Honolulu

Checking out data layers for use in ArcPad and field data collection.

1. Launch ArcMap and open a project or create a new one. This can be done by selecting the **File** button and **Open** from the drop down menu that appears. It may also be done by selecting the **Open** project button  from the main menu.

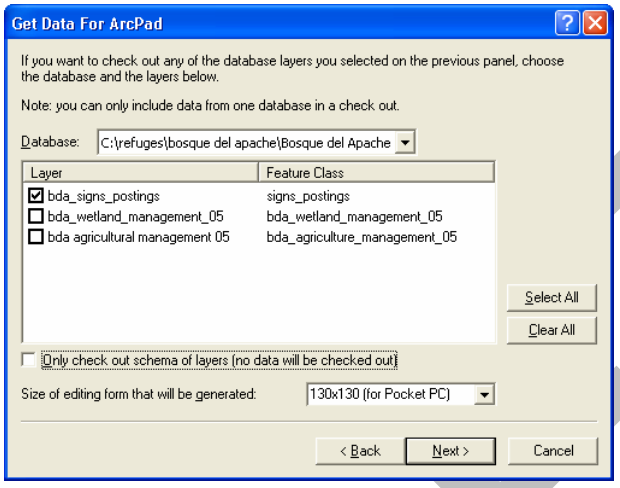
2. Activate the ArcPad menu if is not already. To do this select **View** from the main menu and **Toolbars** from the dropdown menu that appears. Place a check next to **ArcPad**. This loads the ArcPad Tools for ArcMap toolbar  . Dock or move the toolbar if desired.

3. Select the get data for ArcPad button form the ArcPad menu  . This will open the **Get Data For ArcPad** dialog. On the first page of the dialog select all the layers in the project by placing a check in the box to their left. Once complete select **Next >**.

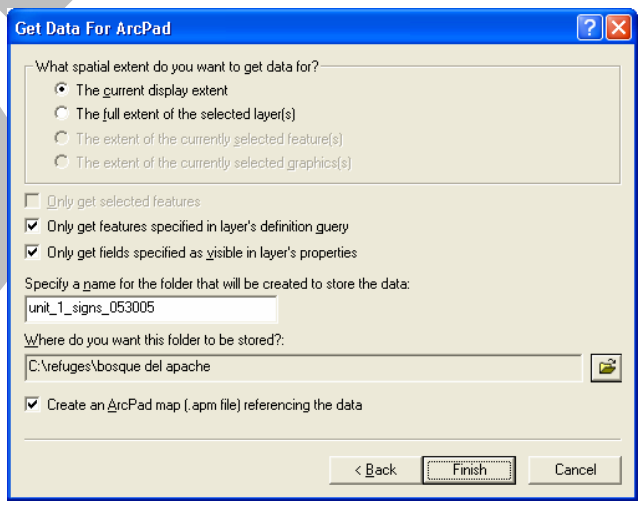


4. Select the databases or feature classes you intend to edit in the field by placing a check in the box to their left. If data already exist in the feature class you wish to edit but you do not wish to check it out check the **Only check out schema of layers** box. If this box is selected you will only check out the schema databases

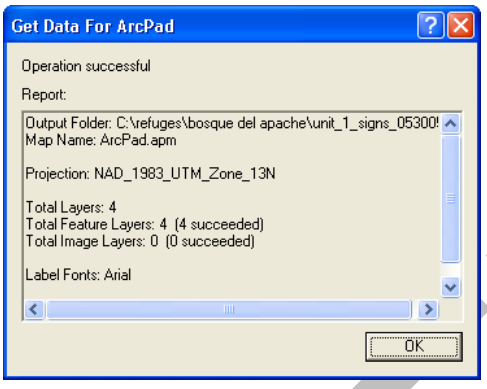
of feature classes checked above and none of the geometry (points, lines or polygons). In the **Size of editing form that will be generated** box take the default **130x130 (for Pocket PC).** Once complete select **Next >**.



5. Select **The Current Display Extent** button. This will set the extent of your project to the current display of the ArcMap project. You may zoom in or out to change the display extent as necessary. Take the checked defaults for **Only get features specified in layer's definition query** and **Only get fields specified as visible in layer's properties**. Next, name the folder the data will be checked out to. This name should include the GPS data logger name it is being checked out to or the name of the individual collecting the data in addition to a description of the data and date it was checked out. Specify the appropriate location to store this file and check the box next to **Create an ArcPad map (.apm file) referencing the data**. Once complete select **Finish**.



6. After a couple of seconds you will be notified if the operation was successful. Select **OK** to complete the check out process. Minimize or close ArcMap.



Considerations and Errors when checking out data to ArcPad

- Checking out the same data to multiple GPS data loggers**:**
  If multiple data loggers are to be used to collect the same data an individual dataset must be checked out for each unit. Each dataset checked out must have a different name following the suggested naming convention outlined above. This will allow the tracking of edits more efficiently and prevent data loss.

- Data projection errors when checking out data for ArcPad**:**
  ArcPad does not support reproject on the fly. Because of this limitation all data being checked out to ArcPad must be in the same projection. If they are not you will receive an error message during the check out process.

- Raster errors when checking out data for ArcPad**:**
  ArcPad has file size limitation for encoding raster data. All users of ArcMap can only encode individual, uncompressed raster smaller than 50 MB. The 50 MB size rule will be evaluated as follows: Image Width ***** Image Height ***** Number of Bands **cannot be greater than 50 million**. If you purchase the MrSid extension from Lizard Tech, then your limit is 500 MB with the added ability to mosaic images.

Transferring checked out data layers to and from the GeoXT

The software used to transfer data between your GPS data logger device and PC is called ActiveSync. This software is free and can be downloaded from: http://www.microsoft.com/windowsmobile/resources/downloads/pocketpc/default.mspx

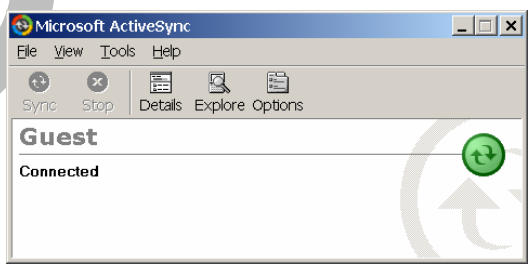
These instructions were developed using ActiveSync version 3.8 and assume that you have not established a partnership between the GPS data logger device and your PC. The GPS data logger used in this example is a Trimble **GeoXT**, however these steps should work with most GPS data loggers capable of running ArcPad.

1. Ensure the cradle of your GeoXT data logger is connected to the USB port of the computer containing the ArcPad data.

2. Place your GeoXT onto the cradle. ActiveSync software should detect your device. If it doesn't, check cabling or try another USB port and then lift your device off the cradle, turn it off and then place back on the cradle. ActiveSync can sometimes be finicky software.

3. Check **NO**, when asked to set up a partnership and connect as a guest. Select **Next >**.

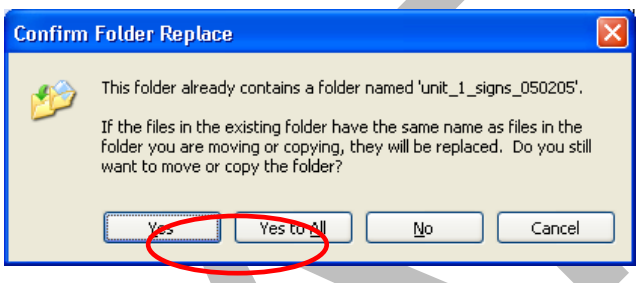
4. The Microsoft ActiveSync dialog appears and your connection status is displayed in the window below **Guest**. Ensure you have a connection (Connected appears below Guest in the dialog window).



5. Select the **Explore** button 

form the Active Sync dialog. This will launch the windows explorer for the GeoXT. Once open copy and past the folder containing the ArcPad .apm files (unit_1_signs_053005). Accept any data formatting dialogs that may appear during the transfer by selecting **OK**. The folder and files containing the checked out data should now be transferred to the GeoXT.

6. Use the same process to move the data back your PC when field work is complete. ***Be sure to place the checked out folder from the GeoXT containing the edits in the same location on the PC as it was copied from.*** When this is done you should be asked if you would like to overwrite the existing folder and all of its contents. Select **Yes to All**.

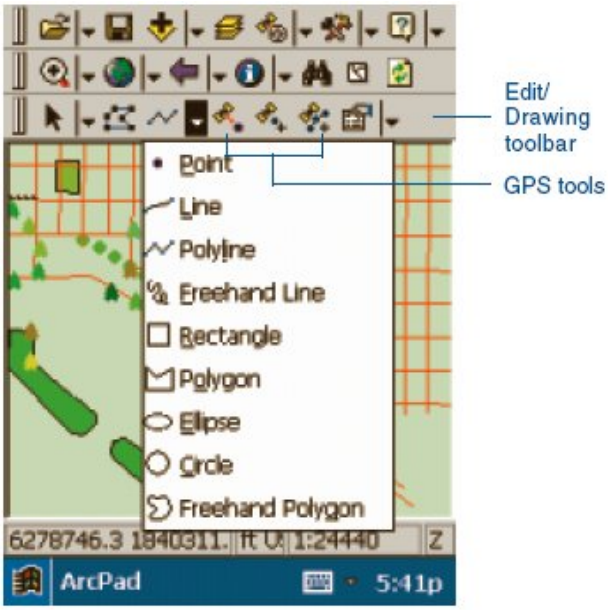


GPS settings.

1. In **logging** settings, set to **supercorrect** for optimal post-processing error Correction.
2. Leave **log GPS to SSF** on.
3. Adjusting slider for **productivity vs precision** will change the max PDOP threshold. This can be adjusted any time during data collection. Trimble recommends a PDOP of 5 or lower.

Field Data Collection

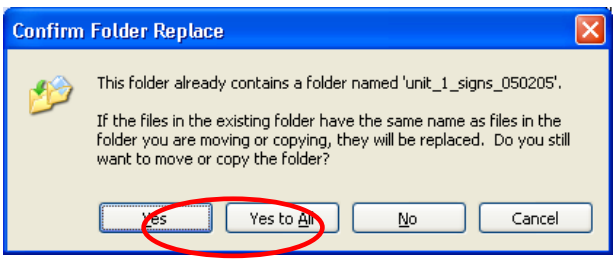
1. Start **Arcpad** by double tapping **icon**.and open .apm project created during the data checkout procedure described above.
2. If you are creating new shp's, create a new layer using dropdown from **Open Map** and pick **New Layer**. Pick appropriate **Type** and add necessary fields
3. To set up for editing, tap **Layers**. When the boxes (under the pencil icon) are checked, those layers become available for editing
4. Then enable satellites by tapping **arrow** to the right of the **GPS Position Window button** and tap **GPS Active** from the dropdown list. Tapping **F1** on the right side of the screen opens the satellite information window.
5. If **edit** toolbar is not visible, go to **toolba**r menu and enable the **edit toolbar**



6. Choose **feature type**(usually point or line) from dropdown menu.
7. To collect point data, click first (**Capture GPS Point**) button (**GPS tools**) and wait until averaging reaches 100%. Make sure in **Options-> Capture that settings** are correct. Averaging should be 10 – 20 points.
8. To collect polyline data, click second (**Add GPS Vertex**)or third (**Add GPS Vertices Continuously**) satellite icon. (see ArcPad documentation).
9. When entering field data, hit the **enter** or **tab** key to lock in the information, otherwise, it will be lost.

Checking in field data from ArcPad

1. Follow the steps outlined in **Transferring checked out data layers to and from the GeoXT.** ***When transferring files from the GeoXT to your PC be sure to place the checked out folder from the GeoXT containing the edits in the same location on the PC as it was copied from.*** When this is done you should be asked if you would like to overwrite the existing folder and all of its contents. Select **Yes to All**.



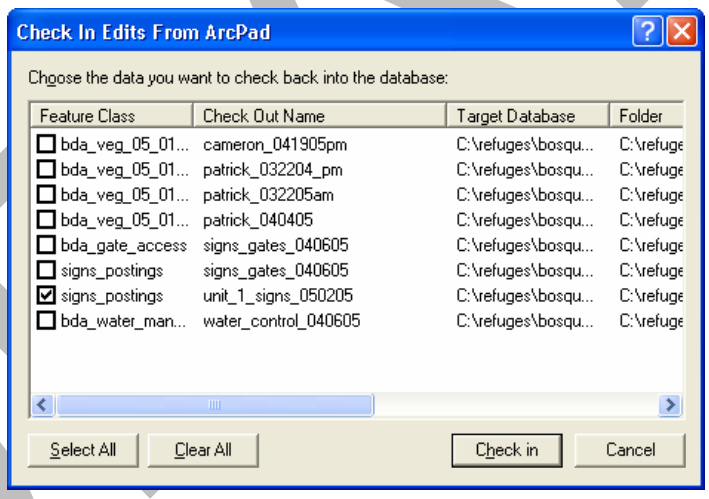
2. Launch an ArcMap session. Open up the .mxd project file which contains the data you originally checked out. You can also add the data layers that were checked out if you did not save them as a .mxd project.

3. Open your **Editor Toolbar** if is not already open. You can do this going to **View** button on the main menu, selecting **Toolbars** and placing a check next to **Editor**. This will open the **Editor Toolbar**.

4. Start an edit session by selecting the **Editor** button on the Editor Toolbar and select **Start Editing** from the drop down menu that appears. Select the data layer that you intend to check in data to in the **Target** box.

5. Select the Check In Edits From ArcPad  button on the ArcPad Toolbar. This button is activated only when you are in an edit session.

6. The Check In Edits From ArcPad Dialog should now appear. Select the edits you would like to check in form ArcPad and select **Check in**.



7. After the check in, your data should appear in your project. From the Edit Toolbar select the **Editor** button and **Save Edits** from the drop down menu that appears. You have completed the check in check out process.

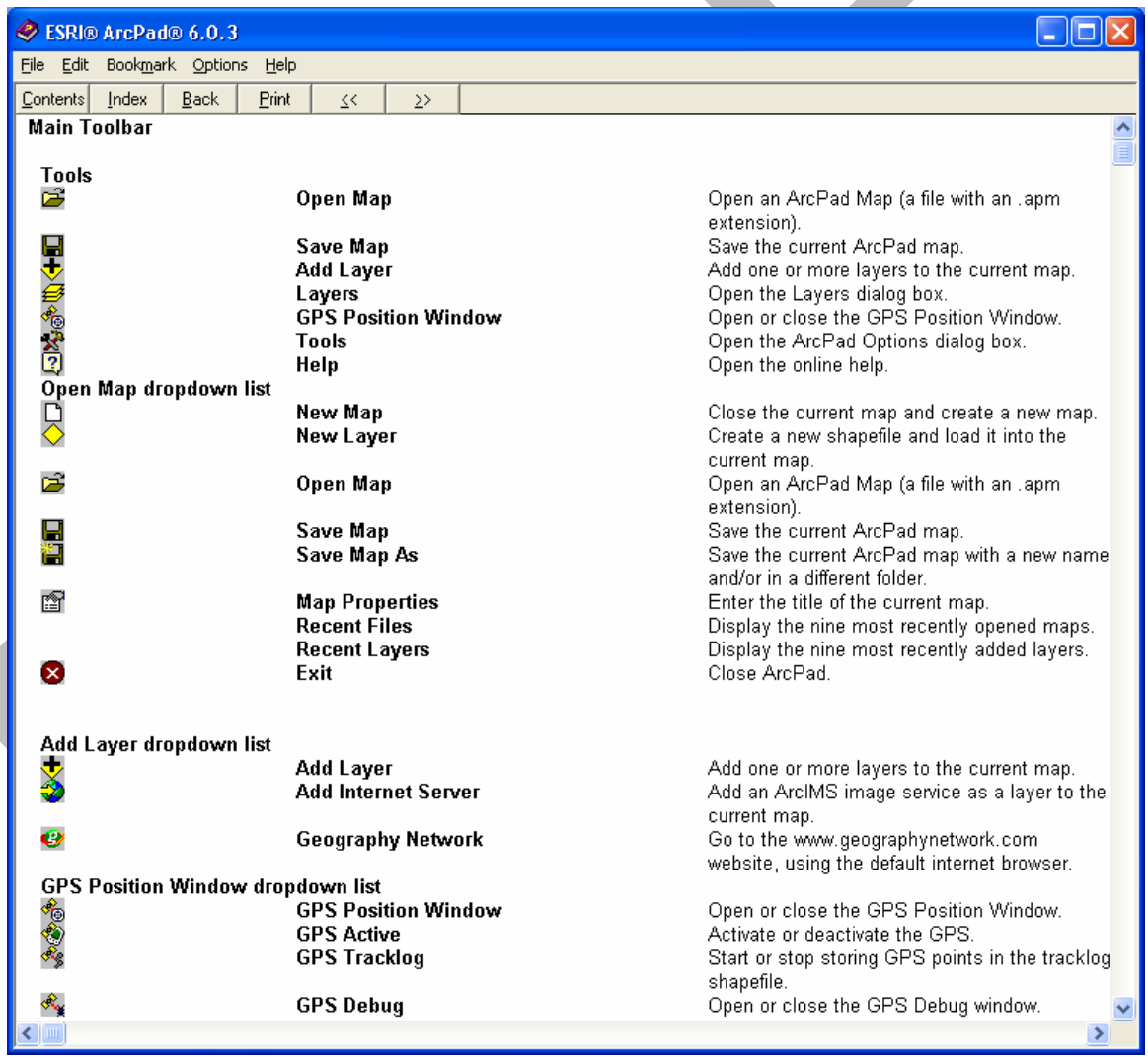
Post-processing

- Open Pathfinder Office and create a new project in the 'pfdata' folder. 'Backup', 'Base', and Export' folders will automatically be created.
- Get base files and store in 'Base' folder. Source: ftp2.pacificgps.com username and

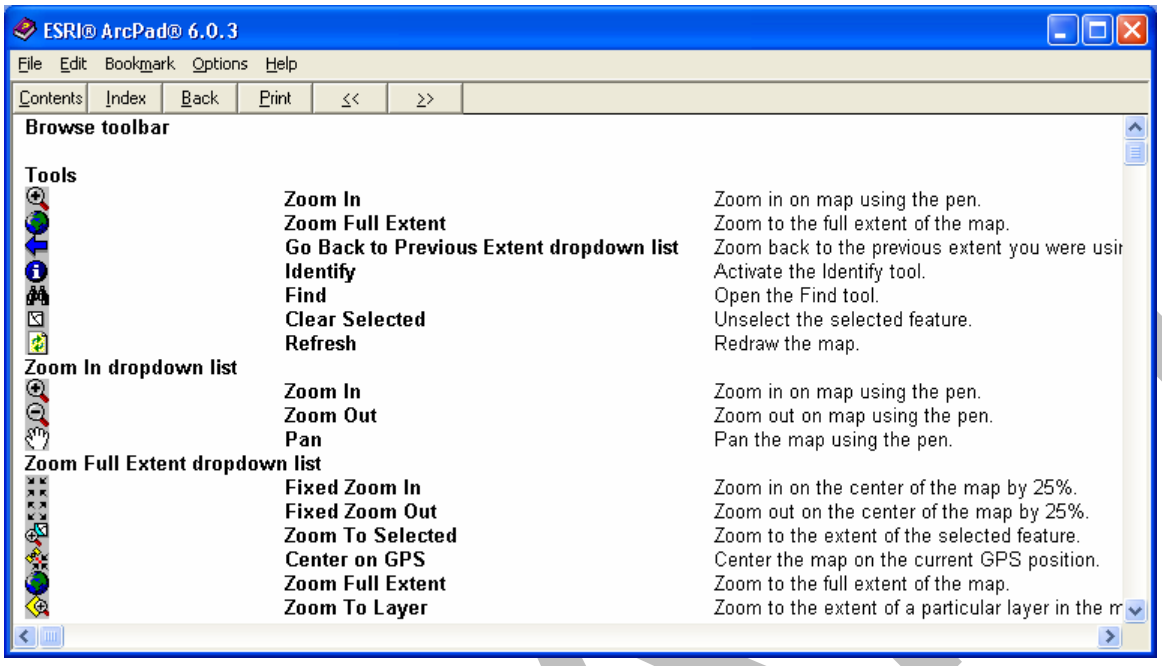
password are the same: pgpspublic or closest Continuously Operating Reference Station (CORS) site at http://www.ngs.noaa.gov/CORS/download2/.

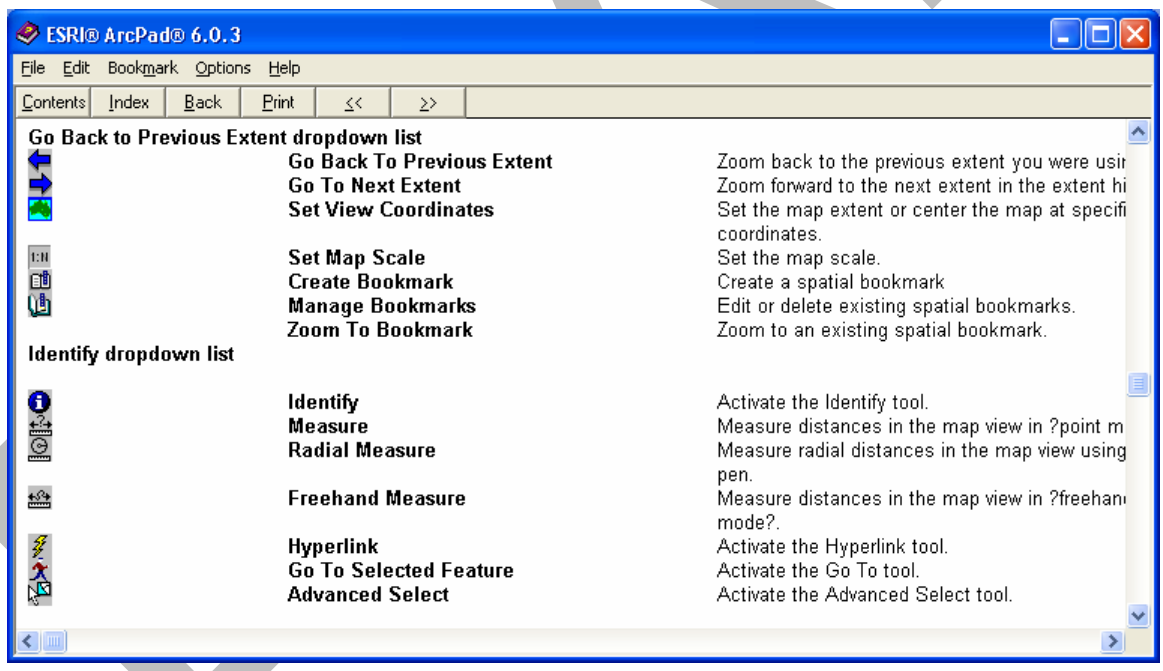
- Add 10 hours to obtain correct base files. They are in UTC and we are -10 hours
- File format is ymmddtt. Time is military.
- Run differential correct utility in PF Office to create cor file
- Run **shapecorrect.e**xe located in the Pathfinder Office directory. This will autocorrect the shapefile
- The shp fss and cor files must be in the same directory to successfully run 'shapecorrect'
- Must have Pathfinder Office v2.9 or higher

ArcPad Main Toolbar Icon Definitions

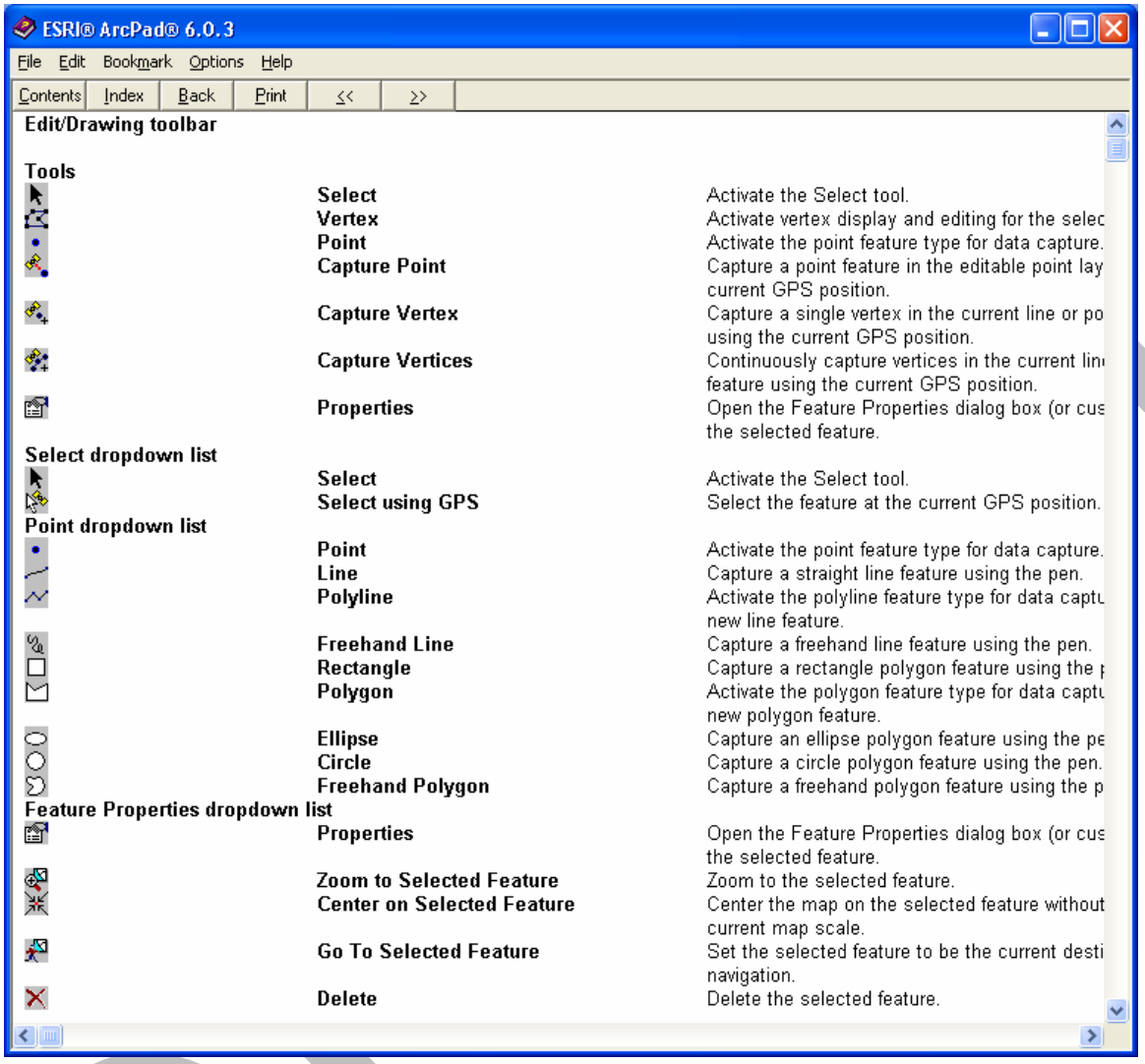


ArcPad Brows Toolbar Icon Definitions





ArcPad Edit/Drawing Toolbar Icon Definitions



APPENDIX I. PACN SOP EXAMPLE – RELATING GIS TO MS ACCESS

Relating GIS to Microsoft Access SOP

DRAFT March 2006

Developed by Allison Snyder, Database Programmer, PACN

Overview

There are 3 ways to work with Access and GIS files:

1) Arc2Ax
Benefits: Creates a link between a shapefile and complex database- can look at queries/report/forms in database. Allows for easy spatial and tabular queries between both. Editing is easier in Access than in the ArcGIS attribute table.
Drawbacks: Still in Beta version- all bugs are not worked out. Currently need to be an administrator on your own computer to be able to use this. Still has some problems if using GUIDs as autonumbers.

2) Import an Access table directly into the map document
Benefits: Very easy
Drawbacks: Can only import tables- no queries. Only creates a temporary file until you make it a permanent file.

***3)* Create an ODBC Connection**
Benefits: Can import queries.
Drawbacks: After displaying XY data, it is only a temporary file until you make it a permanent file.

Instructions

1) Arc2Ax
The Alaska Region GIS team developed an ArcGIS to Access link tool- also known as Arc2Ax. It can be downloaded from http://www.nps.gov/akso/gis/**.** The ArcGIS to Access Link provides a connection between spatial data in an ArcGIS map and tabular data in an Access database. This link allows you to shift between these two products to view related data in both spatial and tabular formats. This tool was developed by Angie Southwould of the Alaska Support Office of the National Park Service. This product is available to the public and may be freely distributed. It is provided “AS IS”, without warranty of any kind. NPS shall not be liable for any damages related to your decision to install and/or use this product.

Requirements: Arc2Ax was developed and tested under Windows 2000 and Windows XP.

1. ArcGIS 9.0 with Service Pack 2 (or later)
2. *Access 2002 (XP) or 2003*

You must save your map project before this will work. It should be saved when changes are made because changes won't be recognized until saved.

Must be a shapefile, etc. Will not recognize if you "Add XY", get an event, and then don't convert to shapefile or other type.

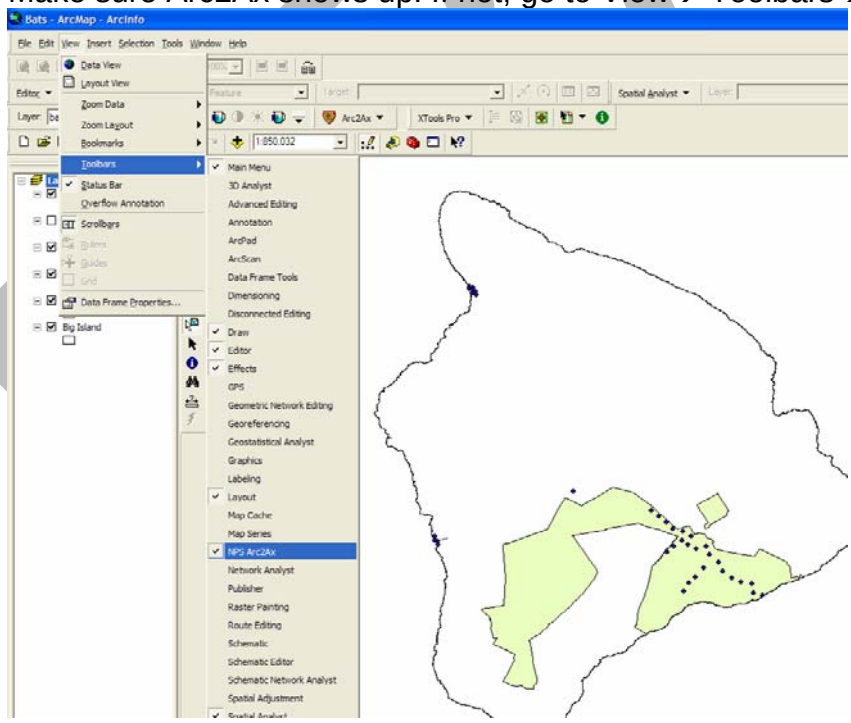
If you delete a point/line/polygon in a shapefile that is linked to an Access database, that record will be deleted in the Access database. So BE CAREFUL! However, if you add new data to the Access table and then run active links, there will be no updates in ArcGIS. Similarly, if you digitize a point in ArcGIS, no data will be added in the Access database. Basically, there needs to be records in the shapefile that relate to records in the Access table. This tool is great for editing where records already exist, but adding data to one (ArcGIS or Access) will not automatically add data to the other.

Using Arx2Ax

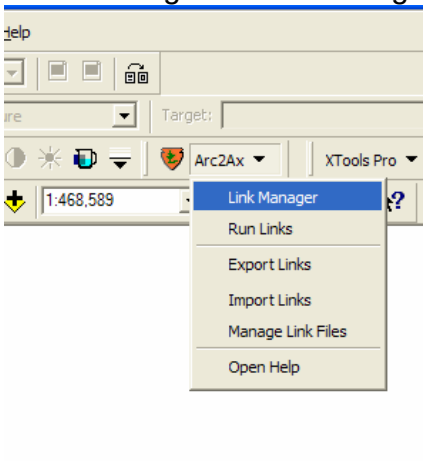
Open ArcMap.

Add a shapefile. You will be linking between an Access object and a shapefile. Therefore, you need a matching field that can link the two. This could be a name, number, or combination of fields. For example, when collecting points on a GPS unit, you might name the points with a location ID or some other unique identifier. When entering the data into a database, you would also need to enter the location ID for each record that will be linked to a feature in a shapefile.

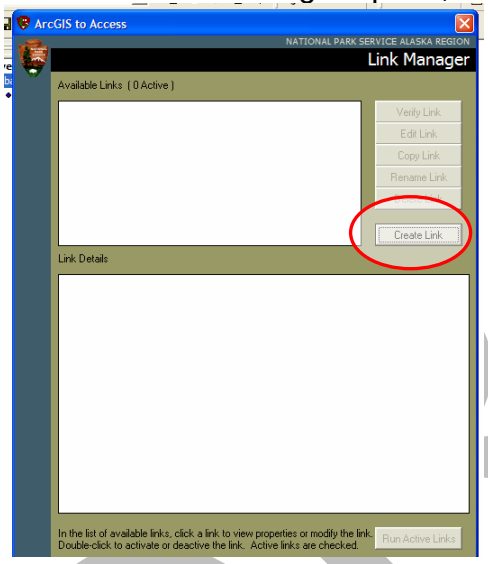
Make sure Arc2Ax shows up. If not, go to View→ Toolbars→NPS Arc2Ax.



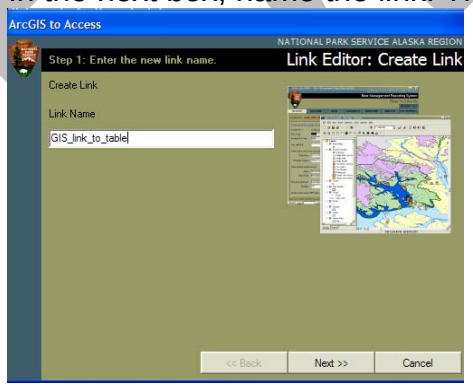
From the Arc2Ax toolbar, click on Link Manager. Remember to save your map project before using the Link Manager.



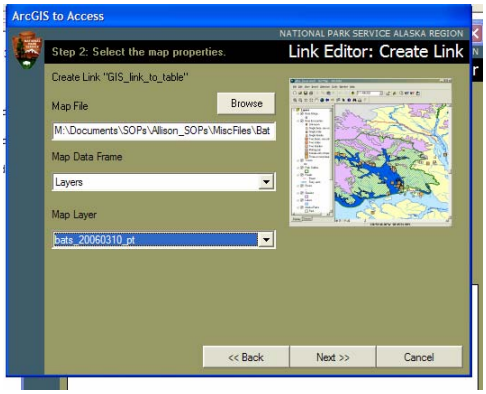
When the Link Manager opens, click on Create Link



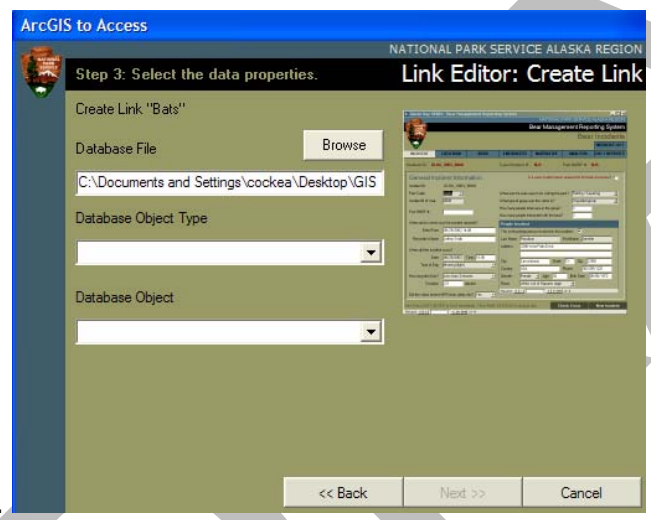
In the next box, name the link. Then click on Next.



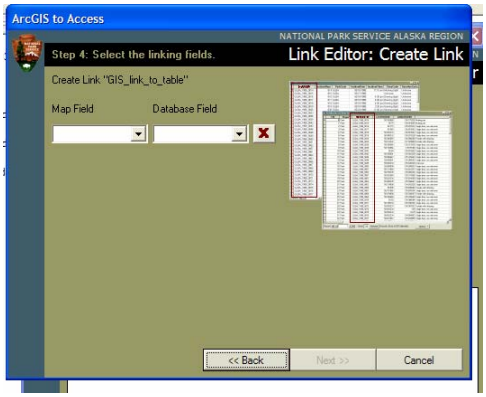
In the next window, you will enter information about your map. The open map document will automatically be entered as the Map File. Select the Map Data Frame, and the Map Layer you want to link to Access. Click on Next.



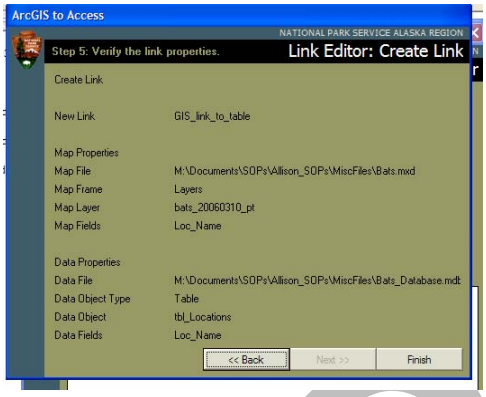
Next you will find the database you are linking to. Click on Browse and find your database. Select your Database Object Type and Database Object. Click Next.



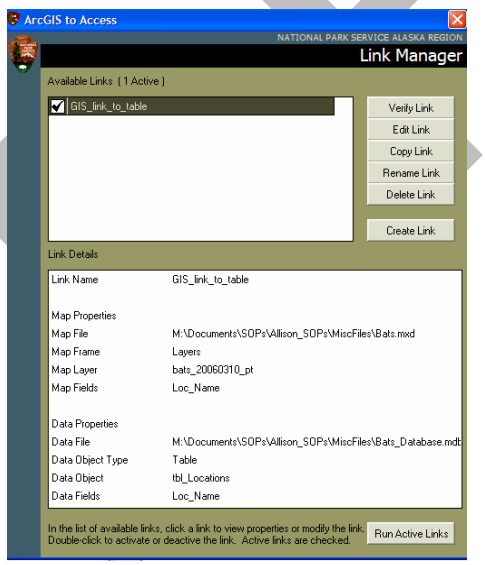
Next you will create the link between the database and the map document. Select the common field- they do not need to have the same name but must contain data that will link the two. Click Next.



On the final page, verify that all information is correct. Click Finish.



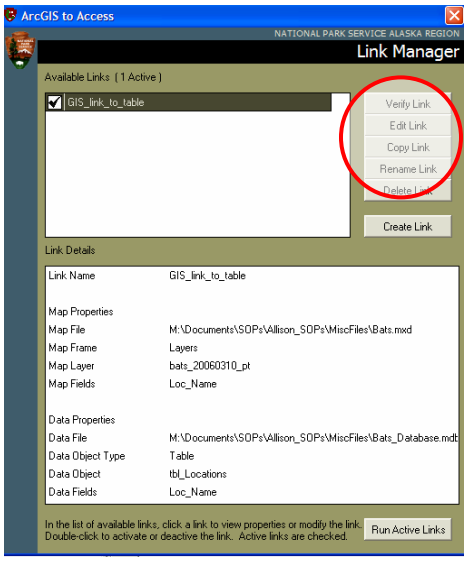
Click close on the dialog that shows you have successfully created a link. The Link Manager should now show your link under “Available Links”. Check the box next to it to make it active.



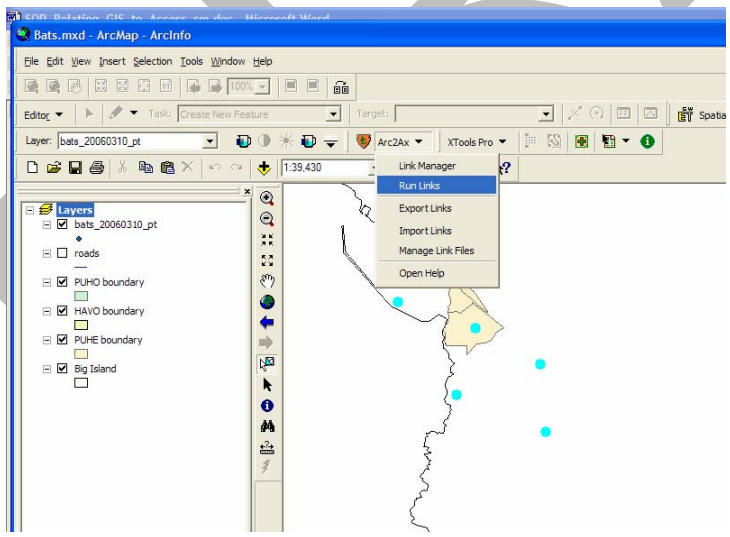
Close the Link Manager.

Modify Links

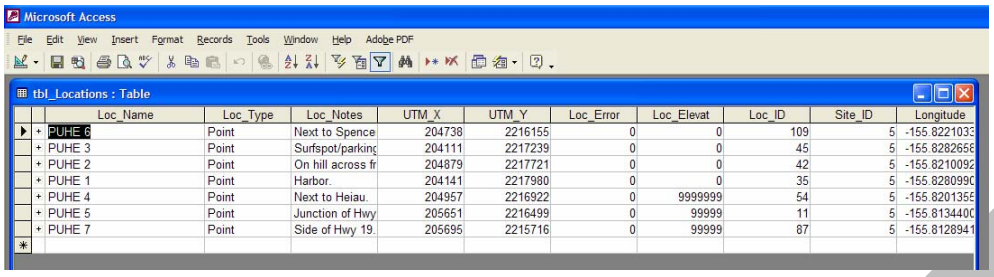
If any links need to be modified, go back to the Link Manager. Here you can edit, copy, rename, and delete links.



In your map document, select part of your shapefile using 
Once you have selected some points/lines/polygons, go to Arc2Ax and click on Run Links.



This will open the related Access database and query for those fields you selected in the map documents.

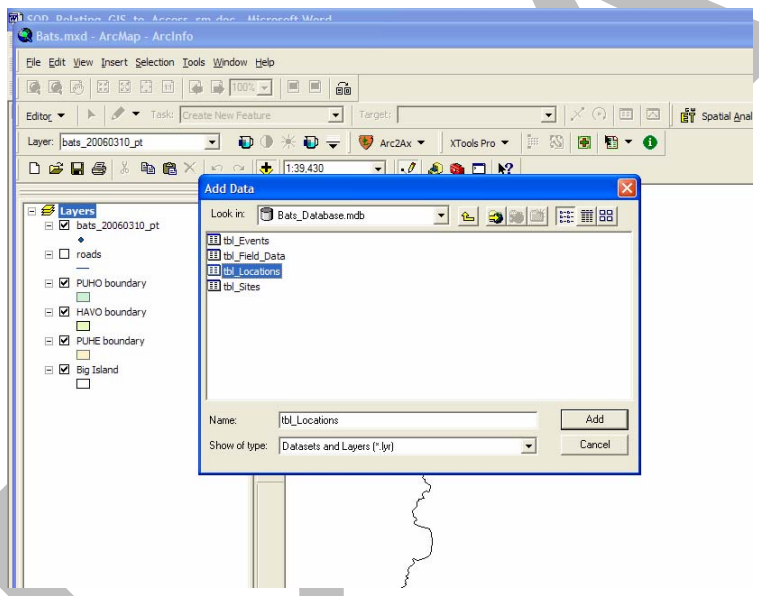


Loc_Name	Loc_Type	Loc_Notes	UTM_X	UTM_Y	Loc_Error	Loc_Elevat	Loc_ID	Site_ID	Longitude
PUHE 6	Point	Next to Spence	204738	2216155	0	0	109	5	-155.8221033
PUHE 3	Point	Surfspot/parking	204111	2217239	0	0	45	5	-155.8282658
PUHE 2	Point	On hill across fr	204879	2217721	0	0	42	5	-155.8210092
PUHE 1	Point	Harbor.	204141	2217980	0	0	35	5	-155.8280990
PUHE 4	Point	Next to Heiau.	204957	2216922	0	9999999	54	5	-155.8201355
PUHE 5	Point	Junction of Hwy	205651	2216499	0	99999	11	5	-155.8134400
PUHE 7	Point	Side of Hwy 19.	205695	2215716	0	99999	87	5	-155.8128941

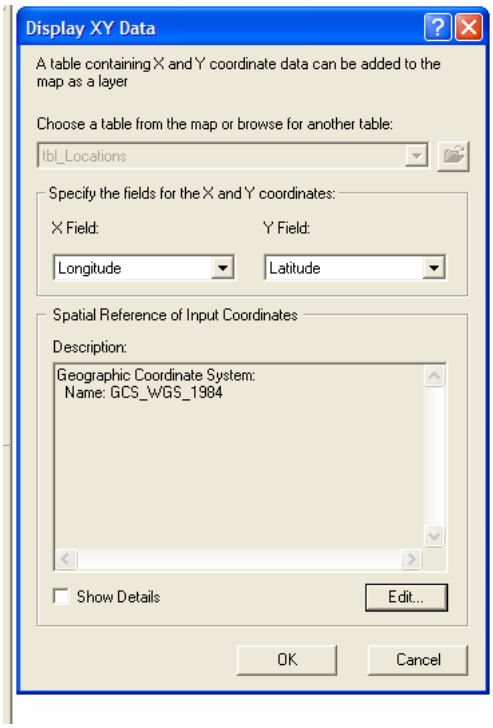
This process can also be done from MS Access. Once you select records in a table/query, you click on Run Links and the map document opens, highlighting the related points.

2) Import an Access table directly into the map document

Open ArcMap. Click on  (Add data). Find an Access database and select any table that includes location information (UTMs, Lat/Long, etc). Click Add.



Choose your X and Y Fields. It is good to define your spatial reference now also. See SOP_AddXY_Data_to_GIS for exact details on defining spatial reference.

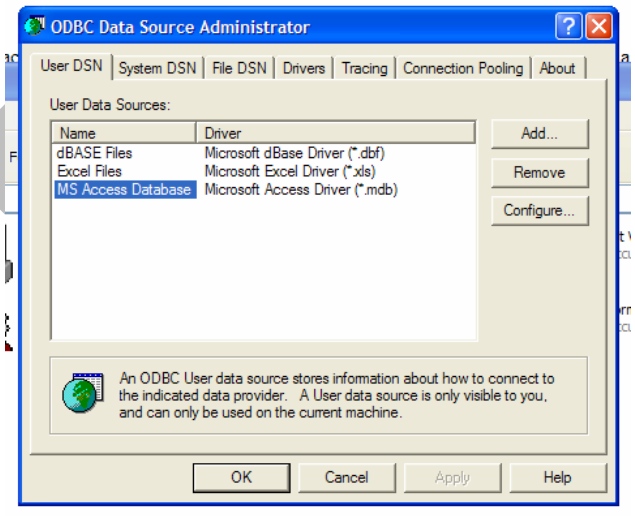


Your points will now show up on the map.

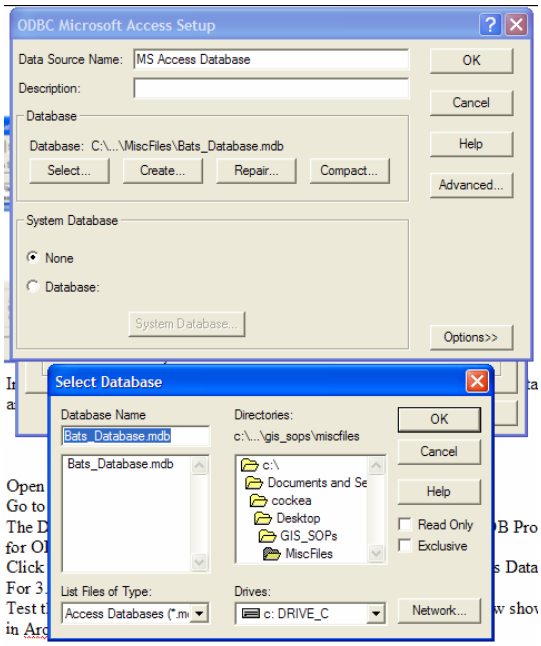
3) Create an ODBC Connection

1) In your control panel, go to Administrative Tools, then Data Sources (ODBC). ODBC = Open DataBase Community

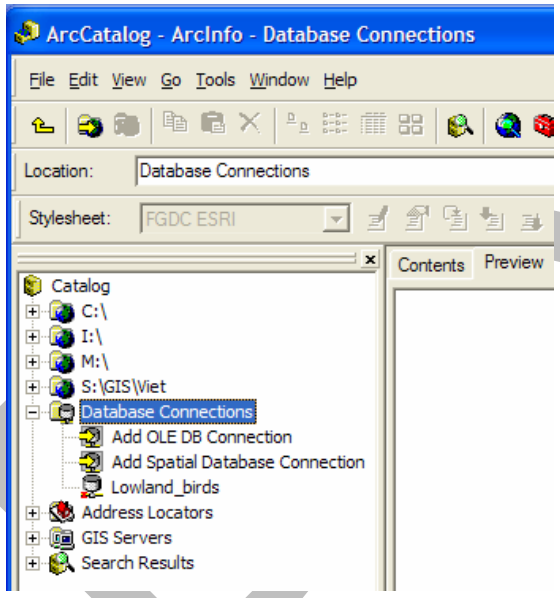
2) On the first tab “User DSN” click on MS Access Databases and then click Configure.



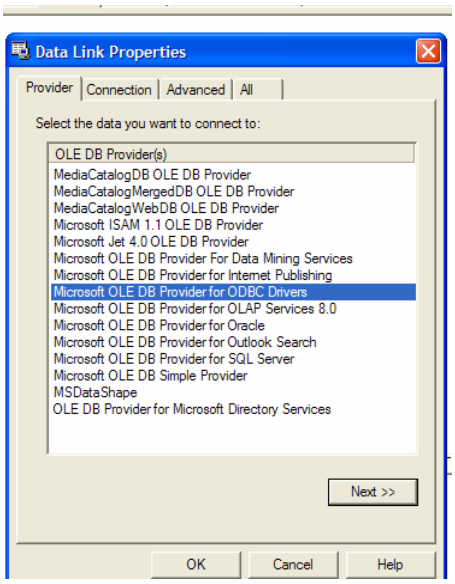
3) In the dialog, click on Select and navigate to your database. Select the database and click OK to all open dialogs.



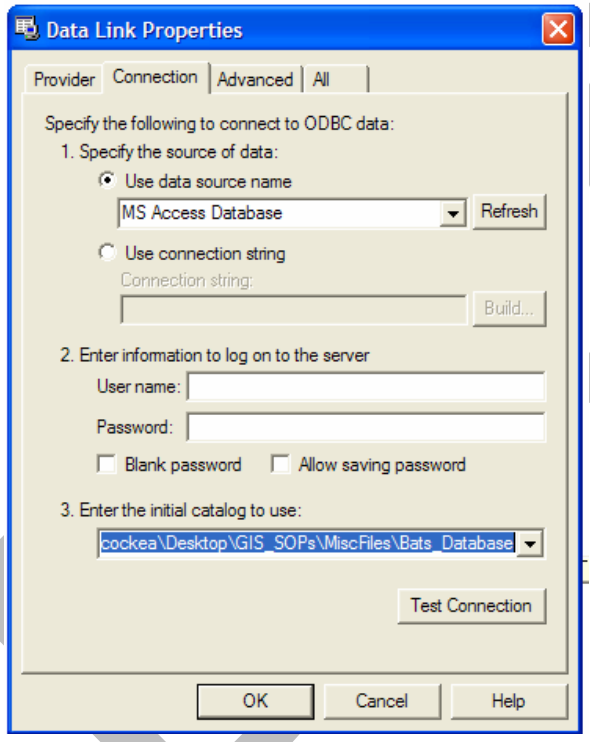
4) Open ArcCatalog. In the Table of Contents go to “Database Connection” and double click “AddOLEDB Connection”.



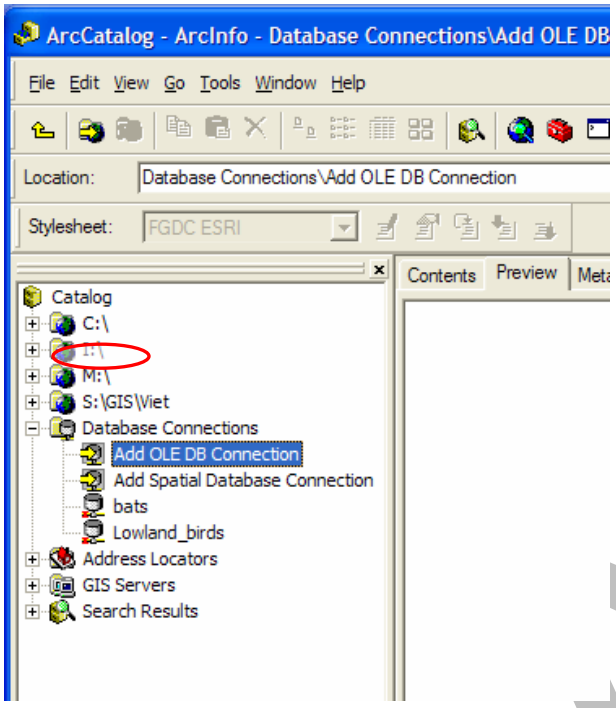
5) The Data Link Properties Dialog will open. Select “Microsoft OLEDB Provider for ODBC drivers”.



6) Click on “next”. For 1, specify the source of data; select “MS Access Database”. For 3, enter the initial catalog to use; select your databases.



7) Test the connection. If good, click OK, then OK again to exit the Data Link Properties Dialog. A highlighted connection now shows up in ArcCatalog. Rename to reflect the connection.



APPENDIX J. CURATORIAL RESPONSIBILITIES GUIDANCE FROM PACN PARKS

CURATORIAL RESPONSIBILITIES OF COLLECTORS
HAWAII VOLCANOES NATIONAL PARK

Collecting and doing research in the National Parks entails certain responsibilities and obligations regarding the curation of specimens and eventual public access to data. This is based on the Code of Federal Regulation which requires registration of specimens with the NPS Catalog. The CFR 36 2.5 (g) Research Specimens states, "(1) specimens placed in displays or collections will bear official National Park Service museum labels and their catalog numbers will be registered in the National Park Service Catalog. (2) specimens and data derived from consumed specimens will be made available to the public and reports and publications resulting from a research specimen collection permit shall be filed with the superintendent." If you collect specimens that are to be permanently retained—regardless of where they are kept—those specimens must be accessioned and cataloged into the National Park Service's Automated National Catalog System, and must bear National Park Service (NPS) labels containing NPS accession and catalog numbers.

Before you begin collecting:
Call the Hawaii Volcanoes National Park Curators office (phone numbers appear at the end of this document) to obtain an accession number for your collection. The accession number is a unique park specific number that helps the curatorial staff track all acquisitions.
THIS ACCESSION NUMBER MUST APPEAR ON ALL REPORTS, CORRESPONDENCE AND ORIGINAL FIELD RECORDS PERTAINING TO YOUR COLLECTION.
Specimens may not leave the park until they are given an accession number.

When you have determined which specimens are to be retained and will not be used in destructive analyses:

1. Call the Curator's office and obtain a block of catalog numbers for the specimens that will be permanently retained. The catalog number is a unique identifying number use by the curatorial staff to track collections. When you call, please have your permit at hand so you can provide the following information:
    - Your permit number;
    - Dates collecting began and ended;
    - Number of specimens collected (estimates are acceptable for large collections);
    - The name of the institution, in which the specimens will be curated, and the name and title of the individual who will be responsible for the specimens.
2. The Curator's Office will send you (or the individual responsible for cataloging your collection):
    - Blank catalog worksheets for you to fill out and send back so that we can enter all catalog records in the Automated National Catalog System. Or we will send you a template in Excel for you to enter all your data. When

we receive all your data we will check it and import it into our cataloging program.
    - Official NPS labels will be generated and sent to you to be place on or with all specimens.
    - An NPS Outgoing Loan Agreement form, to be signed by the individual responsible for the loan. (All specimens, as well as their derivatives and byproducts, remain the property of the United States). NPS policy requires that long-term loans be renewed every 10 years. All collections are subject to a random annual inventory. If part of your collection is randomly selected we will be contacting you to report on the status of the specimen.
3. You must submit:
    - Any specimens that are to be permanently retained, along with their labels and associated documentation, to the repository in which they are to be curated. **Specimens that are to be consumed in the course of research need not be cataloged.** The data derived from consumed specimens will be made available to the public. Reports and publications resulting from a research specimen collection permit shall be filed with the Superintendent.
    - Copies of all of your field records (notes, maps, recordings, etc.) must be sent to the Hawaii Volcanoes National Park Curator's office at the address below, within a year of the final date of collection. A five year time lock will be placed on these records which will prohibit access. Please copy maps and other written or printed matter onto acid-free paper as it will become permanent records.
4. Within a year of the final date of collecting, the following must be submitted to the Hawaii Volcanoes Curator's office:
    - Completed NPS catalog worksheets or Excel database containing all cataloging information.

If you find that you will have trouble meeting this deadline, call or write the curator at the address below to make other arrangements.
PLEASE FEEL FREE TO CONTACT THE HAWAII VOLCANOES CURATOR'S OFFICE AT ANY TIME IN THIS PROCESS WITH ANY QUESTIONS. WE ARE HERE TO HELP YOU COMPLY WITH THE TERMS OF YOUR PERMIT AND MAKE COMPLIANCE AS EASY AS POSSIBLE, UNDER CURRENT REGULATIONS.

Keola Awong, Curator
Tracy Laqua, Museum Technician
Hawaii Volcanoes National Park
P.O. Box 52
Hawaii National Park, HI 96718

Telephone: (808)985-6141 or (808)985-6142
Fax: (808)985-6029
e-mail: keola_awong@contractor.nps.gov
tracy_laqua@nps.gov

CURATORIAL RESPONSIBILITIES OF COLLECTORS
HALEAKALA NATIONAL PARK

Scientific research and collecting within national parks entail certain obligations regarding the curation of specimens and public access to data. The "General Conditions for Scientific Research and Collecting Permit" (Attachment 1), "Haleakala National Park's Park-Specific Permit Conditions" (Attachment 2), and permit specific stipulations/conditions outline your responsibilities as a collector within Haleakala National Park (HALE). These conditions are attached to the permit you will be receiving from the park. Please read them carefully before you sign your permit to ensure you are willing to meet these conditions.

With regard to specimen collection permits, the Code of Federal Regulations (CFR) requires the following conditions: "(1) specimens placed in displays or collections will bear official National Park Service museum labels and their catalog numbers will be registered in the National Park Service National Catalog; and (2) specimens and data derived from consumed specimens will be made available to the public and reports and publications resulting from a research specimen collection permit shall be filed with the superintendent" (ref. 36 CFR 2.5(g)).

Therefore, **if you collect specimens that will be permanently retained—regardless of where they are kept—those specimens must be accessioned and cataloged into the NPS Automated National Catalog System (ANCS+), and must bear labels with the HALE accession and catalog numbers on them.** Permanently retained specimens are defined as specimens not consumed in, or discarded after, scientific analysis.

NOTE: If you intend to deposit permanently retained specimens in a non-NPS repository an NPS Outgoing Loan Agreement must be signed by the official who will be responsible for these specimens at the custodial institution. NPS policy requires that long-term loans be renewed every 10 years. Review and discuss the conditions for outgoing loans (Attachment 3) with the responsible curation official. If there are any questions/concerns regarding these conditions, please contact the Cultural Resources Program Manager as soon as possible.

When you are finished collecting and identifying the specimens that are to be permanently retained:

1. Write or email the Cultural Resources Program Manager to obtain an accession number and cataloging information.
2. You will be sent a Microsoft Excel or Access template that will need to be filled out to complete the cataloging process.

3. Send the completed template to the Cultural Resources Management (CRM) office, along with a copy of your field records (notes, maps, etc.) on acid-free paper.
4. Submit the permanently retained specimens and associated documentation to the repository in which they are to be curated.
5. The CRM Office will send the NPS Outgoing Loan Agreement to the responsible curation official for signature and upon request, labels for the specimens. An institution's own labels may be used so long as they are of archival quality (such as acid-free paper or Tyvek) and contain the HALE accession and catalog numbers in permanent ink, in this format—

   HALE-9999
   [For the accession number; note the hyphen is necessary to distinguish the accession number from the catalog number]

   HALE 9999

   [For the catalog number]

Please feel free to contact the Cultural Resources Program Manager, Elizabeth Gordon, at any time during this process at:

Haleakala National Park
P.O. Box 369
Makawao, HI 96768
Telephone: (808) 572-4424
Email: elizabeth_gordon@nps.gov