



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

Executive Summary	7
1. Introduction	13
1.1. The Inventory and Monitoring Program	13
1.2. Scope of the Data Management Plan	13
1.3. Data Management Goals	14
1.4. Audience	15
1.5. Revisions to the Data Management Plan	15
1.6. Types of Data Covered by this Plan	15
1.7. Sources and Priorities of Natural Resource Data	16
1.8. Pacific Island Network Organization and Structure	17
1.9. Organization of this Data Management Plan	19
2. Data Management Roles and Responsibilities	20
2.1. Data Stewardship Roles and Responsibilities	20
2.2. The Hub of Data Stewardship	22
2.3. Data Management Coordination	25
3. Project and Data Management Overview	27
3.1. Project Management versus Data Management	27
3.2. Project Workflow	28
3.3. Data Life Cycle	32
4. Infrastructure and System Architecture	36
4.1. Computer Resources Infrastructure	36
4.2. PACN System Architecture	39
4.3. LAN and WAN Infrastructure	41
4.4. IT Network Security	41
5. Database Design Overview	42
5.1. NPS and Program Standards	42
5.2. Introduction to Data Models	42
5.3. Fundamental Database Structure	47
5.4. Integrating Databases	47
6. Data Acquisition and Processing	48
6.1. National Standards	48
6.2. PACN Network Standards	48
6.3. Data Sources	48
6.4. Data Discovery	49
6.5. Data Harvesting	50
6.6. Data Collection for Projects	50
6.7. Data Collection for Remote Sensing	53
6.8. Changes to Data Collection Procedures	54
6.9. Data Compilation, Processing, and Integration	54

7. Quality Assurance and Quality Control	55
7.1. NPS Standards	55
7.2. Quality Assurance and Quality Control Mechanisms	56
7.3. Roles and Responsibilities	57
7.4. Quality Assurance/Quality Control Goals and Objectives	58
7.5. Data Collection Quality Assurance/Quality Control	58
7.6. Data Entry	60
7.7. Verification and Validation Procedures	62
7.8. Version Control	65
7.9. Data Quality Review and Communication	66
8. Data Documentation	68
8.1. NPS Standards	68
8.2. PACN Network Standards	68
8.3. Project Documentation	68
8.4. Spatial Data Documentation	69
8.5. Tabular Database Documentation	69
8.6. Vital Signs Protocol Documentation	69
9. Data Analysis and Reporting	71
9.1. Periodic and Annual Reporting	71
9.2. Long-term Trends and Analyses	71
10. Data Dissemination	72
10.1. Mechanisms for Distribution	72
10.2. Ownership, FOIA, and Sensitive Data	73
11. Digital Data Maintenance, Storage, and Archiving	77
11.1. NPS and PACN Network Standards	77
11.2. Electronic Archiving and Storage	77
11.3. Electronic Archiving Process and Workflow	78
11.4. Quality Control for Converted Data	79
11.5. Digital Data Backups	79
12. Non-Digital Data Archiving and Records Management	81
12.1. NPS Standards	81
12.2. Park Unit and PACN Network Standards	81
12.3. Role of Curators in Curation and Museum Collection Storage	82
12.4. Workflow	82
12.5. PACN Network Archiving Process	83
12.6. Archive and Storage Locations	84
12.7. NPS Automated National Catalog System (ANCS+)	85
12.8. Photographs	85
12.9. Specimens	86
12.10. Other Materials	86
13. Implementation of the data management plan	87
13.1. Common Poor Practices in Data Management	87
13.2. Years 1-5 of Implementation	88
13.3. Years 5+ of Implementation	88

Acknowledgments	89
References Cited	91

LIST OF TABLES

Table 1.1. Categories and examples of data products covered by the Data Management Plan.	16
Table 1.2. Summary of land and water spatial characteristics for each park in the PACN.	18
Table 2.1. Categories of data stewardship involving all Network personnel and cooperators.	20
Table 2.2. Programmatic roles and associated data stewardship responsibilities.	21
Table 4.1. General directory structure and their roles in maintaining master data libraries. Exact directory structure is subject to change.	40
Table 6.1. Satellite resolution, swath width, area, cost, and history of various remote-sensing imagery sources.	54
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).	70
Table 10.1. Primary repositories for PACN information and associated specimens.	73

LIST OF FIGURES

Figure 1.1. Map of the Pacific Island Network.	17
Figure 2.1. Core data management responsibilities for a project leader and the network data manager.	22
Figure 3.1. Project and information workflow model depicting the five primary project management stages.	28
Figure 3.2. Data life cycle workflow model depicting sequence of associated events and tasks.	33
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.	37

Figure 4.2. Schematic representing the park, regional, and service-wide server infrastructure.	38
Figure 5.1. Schematic of Conceptual Data Model.	45
Figure 5.2. Schematic of Logical Data Model.	46
Figure 5.3. Schematic of Physical Data Model.	46
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.	56
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.	73
Figure 11.1. Schematic depicting workflow for storage and archiving of electronic and hardcopy products.	79
Figure 12.1. Schematic of natural history curation workflow for both hard copy documents and specimens.	83

LIST OF APPENDICES

Appendix A. PACN Project Deliverables Guidance Document	92
Appendix B. Project Tracking Database Application	100
Appendix C. PACN Vital Sign Protocol SOP Example – Data Management SOP From Benthic Marine Community Protocol	107
Appendix D. PACN Vital Sign Protocol SOP Example – Coral Recruitment Analysis SOP From Benthic Marine Community Protocol	128
Appendix E. Table of PACN GIS Coverages	134
Appendix F. PACN SOP Example – Creating Metadata	137
Appendix G. PACN SOP Example – Uploading GPS Data To GIS	150
Appendix H. PACN SOP Example – Using ArcPad with Trimble GeoXT GPS Unit	160
Appendix I. PACN SOP Example – Relating GIS to MS Access	171
Appendix J. Curatorial Responsibilities Guidance from PACN Parks	183

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.