Chapter 4: Data Handling, Analysis, and Reporting

Version 1.02 (September 16, 2020)

Change History

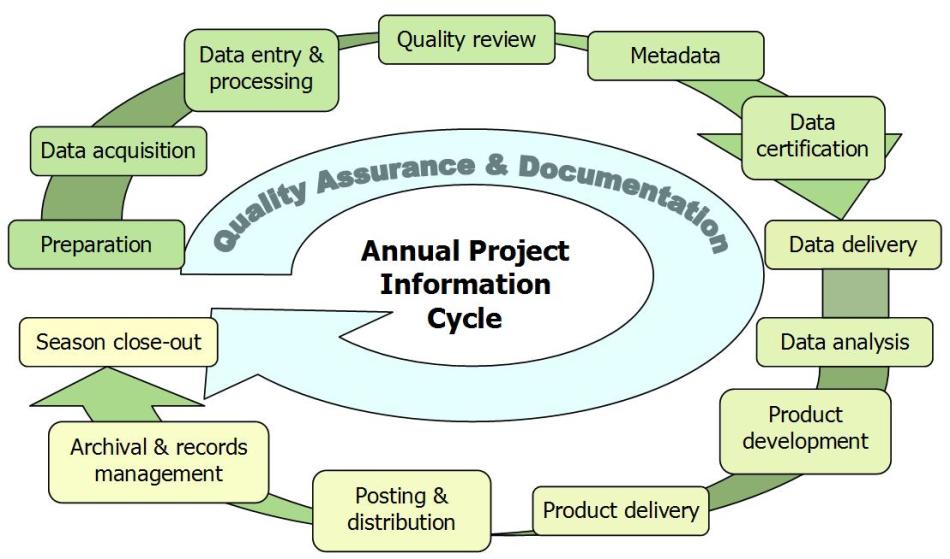
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| New Version # | Revision Date | Author | Changes Made | Reason for Change | Previous Version # |
| 1.01 | 6/27/2016 | Kim Weisenborn | Updates to general instructions. | To accurately reflect the metadata procedures. | 1.00 |
| 1.02 | 9/16/2020 | Kim Weisenborn | Minor updates to content | To update outdated SOP references and external links | 1.01 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

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, changes made, and reason for the change along with the new version number.

Data handling, analysis, and reporting are treated as three interrelated steps in managing information from the Focal Plant Communities Monitoring Protocol. Additional details and context for this chapter may be found in the PACN Data Management Plan (Dicus 2006), which describes the overall information management strategy for the network. The PACN website (NPS 2007b) also contains guidance documents on various information management topics (e.g., report development, GIS development, GPS [Global Positioning System] use, etc.).

Project Information Management Overview

Project information management may be best understood as an ongoing or cyclic process, as shown in Figure 4.1. Specific yearly information management tasks for this project and their timing are described in Appendix G: Yearly Project Task List. Readers may also refer to each respective chapter section below for additional guidance and instructions.



**Figure 4.1.** Idealized flow diagram of the cyclical stages of project information management, from pre-season preparation to season close-out. Note that quality assurance and documentation are thematic and not limited to any particular stage of the information life cycle.

The stages of this cycle are described in greater depth in later sections of this chapter, but can be briefly summarized as follows:

* *Preparation* – Training, logistics, planning, printing forms and maps
* *Data acquisition* – Field trips to acquire data
* *Data entry & processing* – Data entry and uploads into the working copy of the database, GPS data processing, etc.
* *Quality review* – Data are reviewed for quality and logical consistency
* *Metadata* – Documentation of the year’s data collection and results of the quality review
* *Data certification* – Data are certified as complete for the period of record
* *Data delivery* – Certified data and metadata are delivered for archival and uploaded to the master project database
* *Data analysis* – Data are summarized and analyzed
* *Product development* – Reports, maps, and other products are developed
* *Product delivery* – Deliver reports and other products for posting and archival
* *Posting & distribution* – Distribute products as planned and/or post to the Integrated Resource Management Applications Portal (IRMA Portal)[[1]](#footnote-1).
* *Archival & records management* – Review analog and digital files for retention (or destruction) according to NPS Director’s Order 11D (NPS 2001)[[2]](#footnote-2). Retained files are renamed and stored as needed.
* *Season close-out* – Review and document needed improvements to project procedures or infrastructure, complete administrative reports, develop work plans for the coming season.

Pre-season Preparations for Information Management

Set Up Project Workspace

A section of the networked PACN server at is reserved for this project, and access permissions are established so that project staff members have access to needed files within this workspace. Prior to each season, the Project Lead should make sure that network accounts are established for each new staff member, and that the Data Manager is notified to ensure access to the project workspace and databases. If network connections are too slow for efficient data entry and processing, individual staff members may set up a workspace on their own workstation, with periodic data transfer to the PACN server. Daily backups of the workstation to an external hard drive will ensure that no data is lost. Additional details may be found in SOP #10: Workspace Setup and Project Records Management.

GPS Loading and Preparation

The GIS specialist and field leader will work together to ensure that target coordinates and data dictionaries are loaded into the GPS units and data loggers (if used) prior to the onset of field work, and that GPS download software is available and ready for use. Additional details on GPS use and GPS data handling may be found in SOP #6a Using Garmin® GPS Units and on the PACN website (NPS 2007d)[[3]](#footnote-3).

Implement Working Database Copy

Prior to the field season, the Data Manager will implement a blank copy of the working database and ensure proper access on the part of the project staff. Refer to Overview for Database Design (below) for additional information about the database design and implementation strategy.

Overview of Database Design

PACN data management staff designed customized relational database applications to enter, store, and manipulate the data associated with this project. The design of the Established Invasive Plant Species Monitoring Database follows the hierarchical data table organization of the Natural Resource Database Template (NPS 2007b)[[4]](#footnote-4) which is the standard for the NPS I&M Program. For additional details see the table relationships diagram, data dictionary, and other documentation in Appendix F: Established Invasive Plant Monitoring Database Documentation. The PACN data management staff is responsible for development and maintenance of the database including customization of data summarization and export routines.

The database is divided into two components: (1) one for entering, editing and error-checking data for the current season (i.e., the “working database copy”), and (2) one that contains the complete set of certified data for the monitoring project (i.e., the “master project database”). A functional comparison of these two components is provided in Table 4.1.

**Table 4.1.** Functional comparison of the master project database and the working database

|  |  |  |
| --- | --- | --- |
| Project database functions and capabilities | Working database | Master database |
| Software platform for back-end data | MS Access | MS SQL Server or MS Access |
| Contains full list of sampling locations and taxa | X | X |
| Portable for remote data entry | X |  |
| Forms for entering and editing current year data | X |  |
| Quality assurance and data validation tools | X | X |
| Preliminary data summarization capabilities | X |  |
| Full analysis, summarization and export tools |  | X |
| Pre-formatted report output |  | X |
| Contains certified data for all observation years |  | X |
| Limited editing capabilities, edits are logged |  | X |
| Full automated backups and transaction logging |  | X |
| SQL = Structured Query Language, MS = Microsoft |  |  |

Each of these components is based on an identical underlying data structure (tables, fields and relationships, as documented in Appendix F: Established Invasive Plant Monitoring Database Documentation). The working database is implemented in Microsoft (MS) Access to permit greater flexibility when implementing on computers with limited or unreliable network access. Eventually, the master database may be implemented in MS Structured Query Language (SQL) Server in order to take advantage of the backup and transaction logging capabilities of this enterprise database software.

Both components have an associated front-end database application (“user interface” with forms and queries) implemented in MS Access. The working database application has separate screens for data entry, data review, and quality validation tools. The master database application contains the analysis and summarization tools, including pre-formatted report output and exports to other software (e.g., for analysis and graphics production). This front-end application arrangement allows for modification and update of the user interface with no disruption to data entry continuity. The improved front-end file can be distributed to data entry staff, who link it to the back-end file, discard the out-dated front-end file, and proceed with their data entry work. Under this arrangement, data entry staff have no need to open the back-end file, thereby reducing the risk of improper deletions or other inadvertent data loss occurring within the protocol-specific data tables. In addition, a multi-user environment can be accommodated by storing the back-end file on a server available to all users via a computer network.

During the field season, each project crew will be provided with their own copy of a working database into which they enter, process, and quality-check data for the current season (refer to the next section, and SOP #13 Data Entry and Verification). Once data for the field season have been certified they will be uploaded into the master database, which is then used to inform all reporting and analysis. This upload process is performed by the Data Manager, using a series of pre-built append queries.

Data Entry and Processing

After each field trip, technicians will examine data forms to correct obvious errors and incomplete information as soon after data collection as is practical, and enter data in order to keep current with data entry tasks. The working database application will be found in the project workspace. If the project workspace is located on the network server, it is recommended that users copy the front-end database onto their workstation hard drives and open it there. This front-end copy may be considered “disposable” because it does not contain any data but rather acts as a pointer to the data that reside in the back-end working database. Whenever updates to the front-end application are made available by the data manager, a fresh copy should be made from the project workspace to the workstation hard drive.

The functional components for data entry into the working database are described in SOP #13 Data Entry and Verification. Each data entry form is patterned after the structure of the field form, and has built-in quality assurance components such as pick lists and validation rules to test for missing data or illogical combinations. Although the database permits users to view the raw data tables and other database objects, users are strongly encouraged only to use these pre-built forms as a way of ensuring the maximum level of quality assurance.

Regular Data Backups

Upon opening the working database, the user will be prompted to make a backup of the underlying data (see SOP #13 Data Entry and Verification). It is recommended that this be done on a regular basis – at least once per day when new records are being entered – to save time in case of file corruption or mistakes. These periodic backup files should be compressed to save drive space and may be deleted once enough subsequent backups are made. All such backups may be deleted after the data have passed the quality review and been certified.

Data Verification

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, and may lead to incorrect conclusions. Quality assurance/quality control (QA/QC) procedures applied to ecological data include four procedural areas (or activities), ranging from simple to sophisticated, and inexpensive to costly:

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

To the greatest extent possible, the Established Invasive Plant Species Monitoring Database application incorporates quality assurance/quality control strategies involving the first activity (defining and enforcing standards). The database design and the allowable value ranges assigned to individual fields within the data tables help to minimize the potential for data entry errors and/or the transcription of erroneously recorded data. The other activities are integrated in the validation phase. For more details, see the section entitled [Data Quality Review](#data_quality_review), and SOP #14 Post-season Data Quality Review and Certification.

Additionally, as data are being entered, the person doing the data entry will visually review them to make sure that the data on screen match the field forms. This should be done for each record prior to moving to the next form for data entry. At regular intervals and at the end of the field season, the field leader should inspect the data being entered to check for completeness and perhaps catch avoidable errors. The field leader may also periodically run the Quality Assurance Tools that are built into the front-end working database application to check for logical inconsistencies and data outliers. This step is described in greater detail in the section [Data Quality Review](#data_quality_review), and also in SOP #16 Sensitive Information Procedures.

Field Form Handling Procedures

As the field data forms are part of the permanent record for project data, they should be handled in a way that preserves their future interpretability and information content. To minimize the possibility of data loss, hardcopy data forms and field notebooks should be stored in a well organized fashion in a secure location, with photocopies and scanned data forms stored in a separate location (e.g., on the PACN data server). Refer to SOP #11 Field Form Handling Procedures for more details.

If laptops or other digital devices are used to collect data, then the field crew should back up data in the field at least once a day on an extra memory card. Data files should then be uploaded to a computer and backed up as soon as practical after leaving the field. A hard-copy of the digital dataset should be printed and stored following the procedures developed for field data forms.

Image Handling Procedures

Photographic images should also be handled and processed with care. Refer to SOP #12 Managing Photographic Images for details on how to handle and manage these files.

GPS Data Procedures

The following general procedures should be followed for GPS data (see SOP #6a Using Garmin® GPS Units, and Appendix G: Yearly Project Task List:

1. GPS data should be downloaded by the field crew from the units at the end of each field trip and stored in the project workspace (see SOP #10 Workspace Setup and Project Records Management).
2. Raw files should be sent to the GIS specialist for processing (e.g., creating shape files for plot data points, access routes, opportunistic plant sightings) and differential correction if applicable (e.g., Trimble).
3. The GIS Specialist will process the raw GPS data and store the processed data in the project workspace.
4. The GIS Specialist will upload corrected coordinate information into the database and create any GIS data sets.

The Field Lead should periodically review the processed GPS data to make sure that any problems are identified early in the data collection process.

Data Quality Review

After the data have been entered and processed, they need to be reviewed by the Project Lead for quality, completeness, and logical consistency. The working database application facilitates this process by showing the results of pre-built queries that check for data integrity, data outliers and missing values, and illogical values. The user may then fix these problems and document the fixes. Not all errors and inconsistencies can be fixed, in which case a description of the resulting errors and why edits were not made is then documented and included in the metadata and certification report (see Metadata Procedures, Data Certification and Delivery, and SOP #14 Post-season Data Quality Review and Certification.

Data Edits After Certification

Due to the high volume of data changes and/or corrections during data entry, it is not efficient to log all changes until after data are certified and uploaded into the master database. Prior to certification, daily backups of the working database provide a crude means of restoring data to the previous day’s state. After certification, all data edits in the master database are tracked in an edit log (refer to Appendix F: Database Documentation) so that future data users will be aware of changes made after certification. In case future users need to restore data to the certified version, we also retain a separate, read-only copy of the original, certified data for each year in the PACN Digital Library (refer to SOP #17 Product Delivery Specifications).

Geospatial Data

The Project Lead and GIS Specialist may work together to review the surveyed coordinates and other geospatial data for accuracy. The purpose of this joint review is to make sure that geospatial data are complete and reasonably accurate, and also to determine which coordinates will be used for subsequent mapping and field work.

Metadata Procedures

Data documentation is a critical step toward ensuring that data sets are usable for their intended purposes well into the future. This involves the development of metadata, which can be defined as structured information about the content, quality, condition, and other characteristics of a given data set, both tabular and spatial. Additionally, metadata provide the means to catalog and search among data sets, thus making them available to a broad range of potential data users. Metadata for all PACN monitoring data will conform to Federal Geographic Data Committee (FGDC) guidelines and will contain all components of supporting information such that the data may be confidently manipulated, analyzed, and synthesized. In terms of spatial data the ArcCatalog® application along with NPS Metadata Tools & Editor should be used for metadata management.

At the conclusion of the field season (according to the schedule in Appendix G: Yearly Project Task List), the Project Lead will be responsible for providing a completed, up-to-date metadata interview form to the Data Manager. The Data Manager and GIS Specialist will facilitate metadata development by consulting on the use of the metadata interview form, by creating and parsing metadata records from the information in the interview form, and by posting such records to national clearinghouses. Refer to SOP #15 Metadata Development for specific instructions.

Data Certification and Delivery

Data certification is a benchmark in the project information management process that indicates that: 1) the data are complete for the period of record; 2) they have undergone and passed the quality assurance checks (refer to section, [Data Quality Review](#data_quality_review)); and 3) that they are appropriately documented and in a condition for archiving, posting and distribution as appropriate. Certification is not intended to imply that the data are completely free of errors or inconsistencies which may or may not have been detected during quality assurance reviews.

To ensure that only quality data are included in reports and other project deliverables, the data certification step is an annual requirement for all tabular and spatial data. The Project Lead is primarily responsible for completing a PACN Project Data Certification Form, available from the Data Manager or on the PACN website. This brief form should be submitted with the certified data according to the timeline in Appendix G: Yearly Project Task List. Refer to SOP #14 Post-season Data Quality Review and Certification, and SOP #17 Product Delivery Specifications for specific instructions.

Data Analysis

Refer to Appendix G: Yearly Project Task List for the specific analysis tasks and their timing, and to SOP #19 Statistical Data Analysis for a more detailed description of analytical procedures.

As part of the overall database design, the project lead and data manager will work together to develop a series of summary queries and tools designed to prepare the data for analysis and reporting. Some of these tools will produce summarized information that goes directly into reports, while other tools will provide data in the proper format for analysis. Database queries will handle this type of conversion and tabulation, while statistical software such as SAS or SPlus will be used for more advanced analyses.

Using the raw certified data as well as database query results, the project lead will perform two general types of analysis at the end of the field season: summary statistics and trend analysis. Chapter 5 outlines the roles of the project lead and the data manager. In short, the data manager is responsible for query design and writing database routines, while the project lead will use the extracted data to run analyses. As outlined in SOP #19 Statistical Data Analysis the project lead will compute summary statistics (means, variances, confidence intervals, etc.) for each of the vegetation attributes measured. Depending on the attribute, these statistics are aggregated across all species, grouped by life form (i.e., tree, shrub, fern, herbaceous), and/or individual species. In some instances the summary statistics are broken down by percent (e.g., proportion of plots that contain a specific species).

In terms of detecting change, two distinct types of analysis will be pursued. Changes over time for a species or group of species can be evaluated using a zero-inflated model (presence data) or a proportional odds model (cover classes) and likelihood ratio tests. For nonnative richness (i.e., number of nonnative species per segment or quadrat), changes over time can be evaluated using a paired t-test, repeated measures-analysis of variance (ANOVA), or a generalized linear model (GzLM) following the methodology described by Schneider (2007). SOP #19 Statistical Data Analysis provides further details on the analysis of these data items.

Reporting and Product Development

Refer to Appendix G: Yearly Project Task List, and SOP #17 Product Delivery Specifications for the complete schedule for project reports and other deliverables and the people responsible for them. Detailed reporting guidelines and table structures are provided in SOP #20 Reporting.

Report Content

A summary report will be produced annually for each stream/island, with a more detailed report produced every five years. Analysis is conducted at two levels: the station level and stream level, a trend assessment is conducted once adequate data is available. The annual report should:

* List project personnel and their roles.
* List sampling areas, transects and segments completed during the current year.
* Provide maps and brief discussion of the areas sampled during the current year.
* Present summary statistics for all items listed in SOP #19 Statistical Data Analysis.
* Provide trend analysis results for parameters with multiple years of data for areas sampled during the current year.
* Identify any data quality concerns and/or deviations from protocols that affect data quality and interpretability.
* Provide brief management recommendations for invasive prioritization based on status and trends data within a given park.

A more comprehensive analysis and report will be produced every five years after each round of cycling (i.e., after each park and community is sampled). In addition to the above, the five-year report should also:

* Summarize the annual results by plant community and transect, allowing for comparisons among parks and within communities (e.g., how does invasive richness vary from one park to the next, and how does invasive frequency vary between different sampling frames of the same community).
* Assess spatial patterns in the cover class and presence/absence data.
* Identify any possible distributional changes within parks, and communities.
* Place network results within the larger context of invasive plant species changes throughout the Pacific.
* Evaluate operational aspects of the monitoring program, such as whether any sampling locations need to be eliminated or moved (e.g., due to access problems), whether the sampling period remains appropriate (the optimal sampling season could conceivably change over time in response to climate change), etc.

Standard Report Format

Annual reports and trend analysis reports will use the NPS Natural Resource Publications template, a pre-formatted MS Word template document based on current NPS formatting standards. Annual reports will use the Natural Resource Report template, and trend analysis and other peer-reviewed technical reports will use the Natural Resource Technical Report template. These templates and documentation of the NPS publication standards are available at the NPS Natural Resource Publications website (NPS 2007b). Reports will include standard vegetation summary tables and figures. Tables for individual species frequency, cover, and nonnative species richness will include means and standard errors. Thematic map figures of nonnative plant species distributions will also be included.

Review Products for Sensitive Information

Certain project information related to the specific locations of rare or threatened taxa may meet criteria for protection. In this case, the data should not be shared outside NPS except where a written confidentiality agreement is in place prior to data sharing. Before preparing data in any format for sharing outside NPS—including presentations, reports, and publications—the Project Lead should refer to the guidance in SOP #16 Sensitive Information Procedures. Certain information that may convey specific locations of sensitive resources may need to be screened or redacted from public versions of products prior to release.

Product Delivery, Posting, and Distribution

Refer to SOP #17 Product Delivery Specifications for the complete schedule of project deliverables, the people responsible for them, and detailed instructions on how to deliver the final products. Upon delivery products will be posted to the Integrated Resource Management Applications (IRMA) portal as appropriate. Refer to SOP #18 Product Posting and Distributionfor more information.

Holding Period for Project Data

To permit sufficient time for priority in publication, certified project data will be held upon delivery for a period not to exceed two years after it was originally collected. After the two year period has elapsed, all certified, non-sensitive data will be posted to the IRMA Portal. Note that this hold only applies to raw data, and not to metadata, reports or other products which are posted to IRMA Portal immediately after being received and processed.

Special Procedures for Sensitive Information

Products that have been identified upon submission by the Project Lead as containing sensitive information will either be revised into a form that does not disclose the locations of sensitive resources, or withheld from posting and distribution. When requests for distribution of the unedited version of products are initiated by the NPS, by a federal agency, or by a partner organization (e.g., a research scientist at a university), the unedited product (e.g., the full data set that includes protected information) may only be shared after a confidentiality agreement is established between NPS and the other organization. Refer to SOP #16 Sensitive Information Procedures for more information.

All official Freedom of Information Act (FOIA) requests will be handled according to NPS policy. The Project Lead will work with the Data Manager and the park FOIA representative(s) of the park(s) for which the request applies.

Archival and Records Management

All project files should be reviewed, modified or revised, and organized by the Project Lead on a regular basis (e.g., annually in January). Decisions on what to retain and what to destroy should be made following guidelines stipulated in [NPS Director’s Order 11D](https://www.nps.gov/policy/DOrders/DO_11D.pdf), which provides a schedule indicating the amount of time that various types of records should be retained. Refer to SOP #10 Workspace Setup and Project Records Management.

Season Close-out

After the conclusion of the field season, the Project Lead, Data Manager, and GIS Specialist should meet to discuss the recent field season, and to document any needed changes to the field sampling protocols, the working database application, or to any of the SOPs associated with the protocol. Refer to the section on [Data Entry and Processing](#data_entry_and_processing) for additional close-out procedures not specifically related to project information management.

1. <https://irma.nps.gov/Portal/> [↑](#footnote-ref-1)
2. <https://www.nps.gov/policy/DOrders/DO_11D.pdf> [↑](#footnote-ref-2)
3. <https://www.nps.gov/im/pacn/data_management.htm> [↑](#footnote-ref-3)
4. <https://www.nps.gov/im/report-templates.htm> [↑](#footnote-ref-4)