

Standard Operating Procedure (SOP) #14

Data Entry, Verification, and Certification

Version 3.01 (August 6, 2021)

Change History

New Version #	Revision Date	Author	Changes Made	Reasons for Change	Previous Version #
2.00	6/30/2007	Eric Brown	See Track Changes	Response to reviewer's	1.00
2.01	7/20/2007	Larry Basch	See Track Changes	Corrections, additions	2.00
3.00	8/22/2019	Kelly Kozar	Substantial updates to the description of the database, data entry, verification, and certification process. Updated SOP #s.	To accurately reflect the data processing steps. Added a geospatial SOP and had to renumber subsequent SOPs.	2.01
3.01	8/6/2021	Kelly Kozar	Updates to database data entry and certification instructions.	To accurately reflect the data processing steps.	3.00

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.

Purpose

This SOP describes the general procedures for entry and verification of field data in the working project database. Refer to Chapter 4 sections “Overview of Database Design” and “Data Entry and Processing” for related guidance and a clarification of the distinction between the working database and the master database. 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 which can be found on the Integrated Resource Management Applications portal (IRMA)¹. Microsoft Access® is the primary software environment for managing benthic data. Metadata is managed using XML Notepad 2007 and/or Notepad++. ESRI ArcGIS® software will serve as a tool for exploring the spatial component of the Benthic Marine Community Monitoring data (Benthic Marine data).

¹ <https://irma.nps.gov/DataStore/DownloadFile/150631> (accessed 6 August 2021)

Benthic Marine/Marine Fish Database

Marine monitoring at the PACN presently consists of three separate protocols: Benthic Marine Community Monitoring, Marine Fish Monitoring, and Water Quality Monitoring. The two monitoring protocols, Benthic Marine Community and Marine Fish, are 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.

The Benthic Marine/Marine Fish database employs a front-end/back-end configuration in which the user interface resides in the front-end file. 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

(PACN_benthic_marine_FE_YYYYMMDD.mdb²) contains the forms, queries, modules, macros, and reports for the Benthic Marine/Marine Fish database application. The back-end file (PACN_benthic_fish_be_YYYYMMDD.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.

Database Administration

A blank working database application for each park will be made available on the network server in the Benthic Marine data project workspace at the beginning of the field season to facilitate prompt data entry and edits by the PACN Marine Biological Technician at each park. It is recommended that the PACN Marine Biological Technician stores the database files and backups in the appropriate PACN project folder on their local computer while using the database for data entry. When finish, the back-end database will be put in the benthic marine database project workspace (see SOP #12 Workspace Setup and Project Records Management for more information).

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, which prompts the user to create the backup at the onset of a data entry session. This will not require a name change or revision change. The backup files will be named by adding the current date and time to the end of the back-end file name (e.g., PACN_benthic_fish_be_certified_20210802_1130.mdb for a backup file created at 11:30 AM on August 2, 2021). Back-up copies are used for the current field season only and will not be archived.

The working Benthic Marine/Marine Fish database must be stored in the specified PACN project folder, and all data entry and edits must occur in this database file only. It is the responsibility of the PACN Marine Ecologist or designee(s) to ensure that multiple copies of the database for each park are not created, with data entry and edits erroneously occurring in multiple database files. The PACN Data Manager worked with the PACN Marine Ecologist and designated staff to design appropriate

² YYYYMMDD is the date stamp on the database, with YYYY = four digit year, MM = two digit month and DD = the two digit day.

exploratory data analysis tools (e.g., specific queries and reports tailored to the project's exploratory data analysis needs) and incorporated these tools into the Data Summary Tool in the front-end database application (see the Benthic Marine Community and Marine Fish Monitoring Database User Guide³ for more information).

Data Entry

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 let the PACN Data Manager know 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 settlement, rugosity, and coral growth data will be entered at each individual park by the PACN Marine Biological Technician. In addition, fish data from the Marine Fish Monitoring Protocol will be imported into this database as well (see Marine Fish SOP #13 Data Entry and Verification). PhotoGrid[®] comma-separated-value (CSV) files containing data on benthic cover and disease will be generated (refer to SOP #16 Benthic Image Analysis) by the Marine Biological Technician. These PhotoGrid[®] CSV files will be imported into the Benthic Marine/Marine Fish 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 settlement, rugosity, coral growth, marine fish) 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 Manager 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 PACN Data Manager will import the certified data from each park's database into the master Benthic Marine/Marine Fish database. This final certified database file, along with individual park database files for the field season, will then be archived by the PACN Data Manager in the PACN Digital Library (see SOP #12 Workspace Setup and Project Records Management for more information on the PACN Digital Library). The PACN Marine Ecologist and the Park Marine Ecologists will be notified that the master Benthic Marine/Marine Fish database has been updated with certified data from the past field season.

The Benthic Marine Community and Marine Fish Monitoring Database User Guide³ provides step-by-step database instructions accompanied by screen-capture images. Ultimately, it is the PACN Marine Ecologist's and PACN Data Manager'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

³ <https://irma.nps.gov/DataStore/DownloadFile/662972> (accessed 6 August 2021)

for becoming familiar with the field datasheets, the database software, database structure, PhotoGrid[®] export file formats, and any standard codes for data entry.

Data Maintenance

Any editing of archived certified data is accomplished jointly by the PACN Marine Ecologist 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. 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.
- Mistakes can be made during editing so updates must be compared with the original data form prior to validating the data.

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 date stamp to the file name, (e.g., PACN_benthic_fish_be_certified_20210115.mdb. Each additional version is assigned a current date stamp. 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. Any changes to the original datasheets will be documented as outlined in SOP #13 Datasheet Handling Procedures. Secure data archiving is essential for protecting data files from corruption (refer to SOP #12 Workspace Setup and Project Records Management). Once a dataset has passed the QA/QC procedures specified in the protocol, a formal metadata record is created or updated to reflect the certification of that year's dataset (refer to SOP #18 Metadata Development for more information). The certified dataset is archived on the PACN server in the PACN Digital Library. The metadata record for the certified dataset is posted to the NPS Integrated Resource Applications Portal [IRMA)⁴. A non-sensitive version of the dataset file will also be posted to IRMA.

Data Verification

Quality control of data is a critical step in data management. Verification of data (ensuring data on field datasheets match data entered into a database) is the responsibility of the PACN Marine Ecologist and designees. See recommendations outlined in Chapter 4 of the Benthic Marine protocol. Any modifications to the Benthic Marine/Marine Fish database will be described in the edit log and the functionality of the verification routines will be explained in detail in the Benthic Marine Community and Marine Fish Monitoring Database User Guide⁵.

⁴ <https://irma.nps.gov/Portal>

⁵ <https://irma.nps.gov/DataStore/DownloadFile/662972> (accessed 6 August 2021)

Data verification will occur in three steps:

Step 1: The PACN Marine Ecologist or Park Marine Ecologist, 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. Edits will be made and updated PhotoGrid[®] export files will be generated for import into the Benthic Marine/Marine Fish database (see details in SOP #16 Benthic Image Analysis). The original PhotoGrid[®] export files will be archived as documentation between originally analyzed photos and the final, certified analysis. Changes to data during this process will also be documented in the Error Checking Data Form (see Appendix SOP 14.a).

This review will serve as both a data verification process and a data certification process (See Data Certification below). 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/Marine Fish database, park staff will compare 10% of all transect site information, rugosity data, coral growth data, and coral settlement 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 PACN Marine Ecologist or Park Marine Ecologist and corrected by the Marine Biological Technician as needed.

Step 3: The geospatial component of the Benthic Marine/Marine Fish database will be verified each field season by the PACN GIS Specialist. Geospatial data will be exported from the database and imported into ArcGIS[®]. Any corrections will be tracked, and an error rate greater than 10% will require a detailed review of the coordinate fields for all records.

Data Certification

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; 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.

At the end of each field season, the PACN Marine Ecologist, designees, and the PACN Data Manager are collectively responsible for finalizing a certified dataset for that field season. The PACN Marine Ecologist and/or Park Marine Ecologist will complete all data certification.

Certification of benthic image analysis records will be accomplished using the PhotoGrid[®] software (refer to SOP #16 Benthic Image Analysis).

1. A second observer, either the PACN Marine Ecologist or Park Marine Ecologist, will compare their own benthic cover interpretations for 10% of the images against the interpretations of the PACN Marine Biological Technician (the original observer).
2. Changes to data during this process will be documented in the Error Checking Data Form (see Appendix SOP 14.a), and the error rate will be calculated by the PACN Data Manager.
3. If the error rate of the reviewed records is greater than 10%, then the two observers will review the benthic cover interpretation guidelines and the original observer will re-interpret all of the images.
 - a. In this event, the certification of benthic image analysis records would then be repeated.
4. If the error rate is less than 10%, the edits made by the PACN Marine Ecologist or Park Marine Ecologist are saved in PhotoGrid[®] and export files will be generated for import into the Benthic Marine/Marine Fish database (refer to SOP #16 Benthic Image Analysis).
5. The original PhotoGrid[®] export files, along with the Error Checking Data Form, will be archived as documentation between originally analyzed photos and the final, certified analysis.

Certification of coral settlement analysis records, if applicable, will also be accomplished the PACN Marine Ecologist and/or Park Marine Ecologist. As with the benthic image analysis records, the PACN Data Manager will randomly select 10% of the coral settlement 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 juveniles counted, then the two observers will review the guidelines for identifying and counting coral juveniles, and the original observer will re-analyze all of the plates. In this event, the certification of 10% of the plates by a second observer would be repeated.

The PACN Marine Ecologist, in conjunction with the Park Marine Ecologist and PACN Marine Biological Technician, will be responsible for evaluating all of the database records for a given field season in order to provide final approval for a certified dataset for that field season. The PACN Data Manager will work closely with the PACN Marine Ecologist to provide any needed database queries, reports, graphs, or export file formats to assist with this overall certification. Once a given field season's dataset has been certified, the PACN Data Manager will be responsible for archiving the certified dataset, posting the metadata (and dataset, if appropriate) to the NPS Data Store⁶, and letting the Park Marine Ecologist know of the certified dataset locations (see SOP #21 Product Posting and Distribution for more information).

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 PACN Marine Ecologist is primarily responsible for completing a PACN Project Data Certification Form available

⁶ <https://irma.nps.gov/DataStore>

in Appendix SOP 20.b in SOP #20 Product Delivery Specifications. This brief form should be sent to the PACN Data Manager once the data is certified.

Appendix SOP 14.a. Error Checking Data Form

Data form for error checking of point identifications in photoquadrat images.

Year: _____

Park: _____

Marine Ecologist: _____

Photgrid® IDer: _____

Total # of Images Checked: _____

Total # of Points Checked: _____

Total # of Errors: _____

Events	Transect	Image	Point #	Incorrect Substrate	Correct Substrate
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					