

Standard Operating Procedure 18: Data Management and Archiving, version 1.0

This Standard Operating Procedure (SOP) gives step-by-step instructions for managing the data collected in the Aquatic Macroinvertebrate Monitoring Program at Shenandoah National Park. This SOP describes the procedure for field checking data, data entry (in conjunction with SOP 17), data verification and correction (in conjunction with SOP 17), and database organization and documentation. Accurate collection and stewardship of data are critical to the success of the program. Ensuring data quality must be a priority during crew training, and throughout the data collection, entry, verification, and validation process.

Responsibilities for quality control

The biologist and lead biological technician are responsible for ensuring that all crew members:

- 1) understand all field protocols,
- 2) understand data fields and values on paper data sheets and in the database.

Crew members are responsible for learning field protocols, and all data fields and values on paper data sheets and in the database. In addition, each crew member is responsible to do the following whenever they are unsure or unclear about the protocol:

- 1) Do not guess.
- 2) **Ask** the crew leader or other crew members while in the field.
- 3) Ask the lead biologist.
- 4) Take copious and clear notes including photographs about the uncertainty.
- 5) Follow through to securing guidance and correcting the uncertainty.
- 6) Err on the side of collecting too much data.

Data collection

Field crews record data on paper data sheets (SOP 16). Carry datasheets within an enclosed clipboard. **Print legibly** all information on datasheets using pencil. Waterproof/all-weather copies of blank datasheets should be carried into the field at all times. All corrections to the datasheet should be made by crossing through the error. Information should never be erased and incorrect or older information should never be overwritten.

Regardless of the data collection method, **all data values shall be called out and repeated** when one crew member is measuring and one crew member is recording. The measurer shall clearly call out the value to be recorded. The recorder shall clearly repeat the value aloud as it is recorded, to ensure that the value was heard and recorded accurately. If the measurer does not hear a response from the recorder, the measurer shall repeat calling out the measurement until a response is obtained. Some larger streams are extremely noisy and the data taker will have to move with the measurer or another crew member will have to verbally relay the data across the stream.

Field checking data

Field checking the data is the process of reviewing all paper data sheets before leaving the site to ensure all fields are complete and no obvious errors exist. A single person is needed to complete this step. Usually this task is completed by the field crew leader, but any field member who is

knowledgeable of species in the park and the monitoring protocol and has an eye for detail can also complete this task. Datasheets should be field checked by someone other than the data recorder.

On each datasheet, the person conducting the field check should ask the following:

- a. Is the header information complete – crew, date, plot number, transect identifier?
- b. Does the plot name used on the datasheet match the one used on the tree tag?
- c. Are multiple pages of the same datasheet numbered?
- d. Are there any notes in the margins of the datasheets that need to be addressed, moved, or crossed out to eliminate future confusion?

Once it has been determined that the datasheet is complete, the initials of the crew member who field checked the datasheet should be added to the datasheet in the lower right hand corner of each page.

Post Data collection Office tasks

After returning from the field, the datasheets go into the appropriate file cabinet drawer in the office. If questions exist for the site the datasheets should be flagged and every effort should be made to answer questions or fully document problems before filing the datasheets.

Electronic Files

The electronic files for this program can be found on the O: Drive in the working directory under “O:\WORKING\AQUATIC”. There should be several directories with files for the program. Data management files are located in the “O:\WORKING\AQUATIC\Database\” directory (Figure 2). Electronic data files including the databases, scanned datasheets and photo monitoring images are stored in a directory structure on a file server with appropriate access levels granted to users. Prior to a data entry session an automated backup of the back-end data file is made and stored on the server to allow “roll-back” if database corruption is encountered. A scheduled back-up utility makes copies of changed files to a server in a different building at the SHEN HQ campus nightly. Annually these back-up directories are evaluated and archived or replaced by a “fresh” copy of the directory.

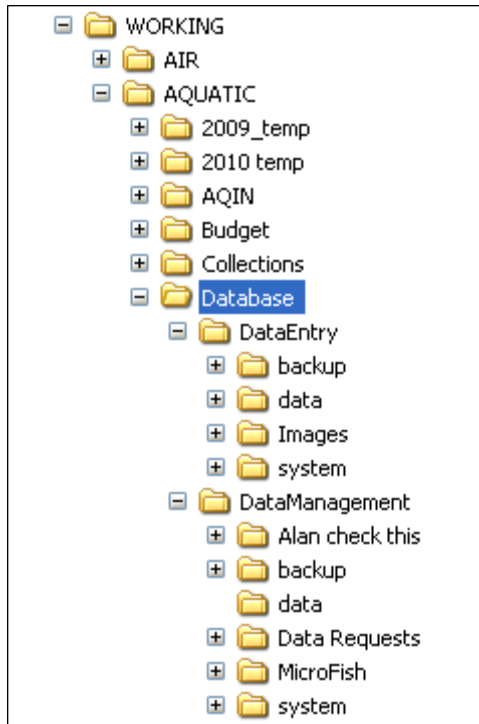


Figure 2. O Drive directories.

Database System

The Shenandoah National Park aquatic macroinvertebrate data system consists of three main components, the data entry application, the temporary storage databases, and the master database. Field data is entered into the first temporary database using the data entry application (see SOP 15 for details on data entry). Then to help assure accurate data the same data is entered a second time into a second temporary database. The two temporary databases are compared and any discrepancies can be reviewed and fixed. At the end of the field season the data in the temporary databases are reviewed by the program staff and then added to the master database.

An ESRI shapefile is maintained to house the point locations of the transects/sites. The shapefile can be joined to data exports in ArcGIS to display attributes of the plots spatially.

The current Shenandoah National Park aquatic macroinvertebrate monitoring database was developed in Microsoft Access 95 but has been converted and updated to the current version of MS Access. This database predates the NPS I&M Natural Resources Database Template (NRDT) but many of the design principles are present.

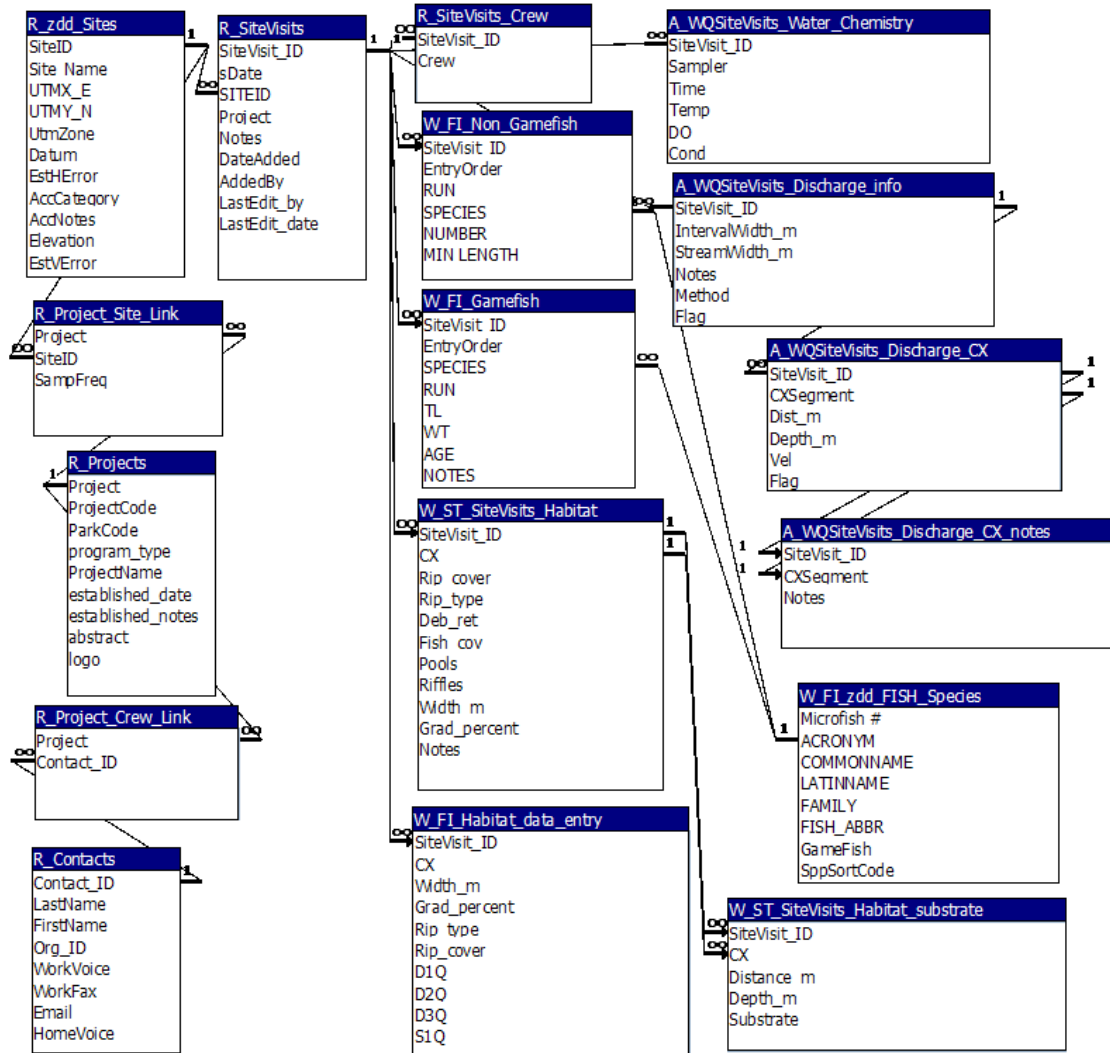


Figure 3. Back-End table design and relationships for the fisheries database.

Data Entry and Verification

Data entry is typically accomplished by teams of two people working together. One person calls out the data and the other types it in. The database can be found on the O: Drive in the working directory for the program in the “O:\Working\AQUATIC_New_Data\Data_Entry” directory. There should be several copies of the database available here so several teams of two can perform data entry at the same time.

Data is entered into the first temporary database using the data entry application. Then to help verify that the data entered matches accurately what has been written on the datasheet, data the data is entered a second time into a second temporary database by a different data entry team. The two temporary databases are compared by the application and any discrepancies are reviewed and fixed (see SOP 17 for a detailed guide to data entry).

Scanning Field datasheets

All of the field datasheets will be scanned to provide a permanent back up of the raw data. Use the copier/printer/scanner in the hallway of the Lower NCR building for scanning data sheets. Press the “scan/mail” button on the far left. Using the touch-pad, press “T Drive > NCR”. Quality Adjust > Scanner settings: Resolution= 600, Contrast=Middle, Type=Text/Photo. Load the stack of pages to be scanned face up with the top of the documents to the left side of the scanner (or going into the document feeder head first) making sure corners are not folded and sheets are flat. Press the green button and wait for the prompt that the scan is complete then proceed to the next set of documents to scan.

After documents have been scanned, log into a computer and go to “T Drive” (\\Inpshenhq2\ftp\Scanjobs\NCR). Search for new scanned documents with the scanner generated names like, “0800230291F0-170503112037-0035.pdf”. Open these, confirm the site and date information for the file, close the file then rename. Rename each of the files using the format, “SiteID_YYYY-MM-DD_Date Sampled.pdf” (example, 1F114_2018-07-02.pdf). Double check that all sites that were scanned were saved and re-named then move the pdf files into the O:\Working\AQUATIC_New_Data\Scanned_Datasheets\. After all this is completed, file the site folders after the “Scanning Complete – To be Filed” section of the macroinvertebrate data file drawer.

Digital Photos

Standard Operating Procedure 10 includes information on management of digital photos.

End of Season - Database Tasks and Data Quality Control (QC)

Following data entry, 100% verification, scanning of field datasheets, and renaming of plot photos for the current season a QC data review will be conducted to identify additional potential mistakes. This QC step should include, running a set of predefined QC summaries to catch common errors, omissions, and problems. The summaries will include:

- List of plots entered in the database checked against the crew leader’s notes and the schedule for the year.
- List of renamed digital photos checked against photo logs or visited site list
- List of scanned datasheets checked against folders of entered data and database
- Identify out-of-range values
- Search for missing values
- Evaluate outliers

After identified issues are addressed or corrected then the data records can be Labeled “Certified” the records can be locked by the program lead. Future edits to the data will need to “un-lock” the record and describe why in the notes then relock the record after edits are made. Then data in the temporary / seasonal databases are added to the master database using pre-programmed macros (Figure 4). The same interface is also used to integrate the identified sample data that is double entered by the contracted lab then transferred electronically to the park. The macros import the taxonomic information for each sample and pair it up with previously entered habitat data. The master database is used for all data summaries and analysis.

Scanned datasheets and renamed site photos should then be moved from the “O:\Working\AQUATIC_New_Data” folder into the master directories for each site such as “O:\Working\AQUATIC\Database\Scanned_Datasheets\AQIN\Site_Folders” and “O:\Working\AQUATIC\Aquatic_Images\Site_Folders”

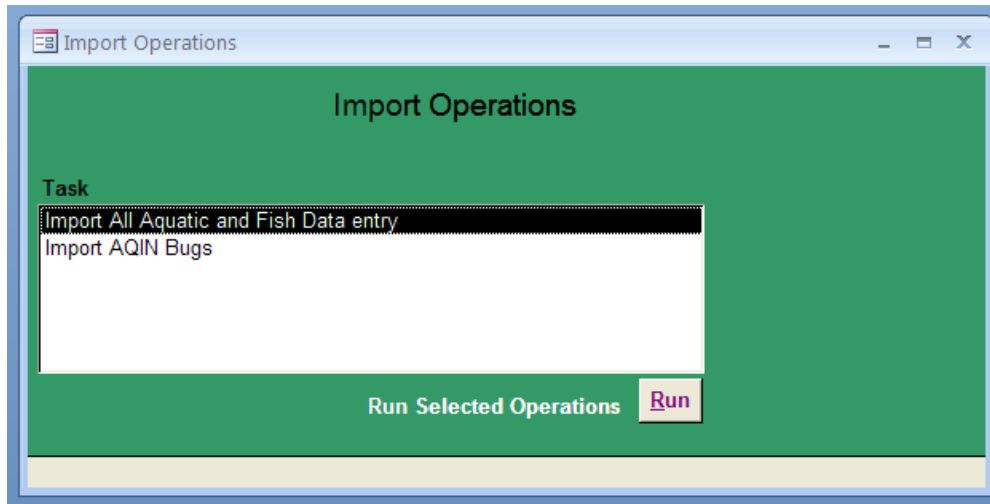


Figure 4. Import data operations form.

Metadata Procedures

Any dataset compiled by Shenandoah National Park must be accompanied by metadata. This includes both spatial and non-spatial datasets. For metadata associated with geospatial data, we abide by Executive Order 12906, which mandates that every federal agency document all new geospatial data it collects or produces using the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM; www.fgdc.gov/metadata/). All GIS data layers will be documented with applicable FGDC and NPS metadata standards. Shenandoah National Park will also generate FGDC-style metadata for non-spatial datasets that meet this standard, without the geospatial-specific elements.

SOP Revision History

All protocols, regardless of how sound they are originally, require modification as new and different information becomes available. SOP version numbers increase incrementally by tenths (e.g. version 1.1, version 1.2...etc.) for minor changes. Major revisions should be designated with the next whole number (e.g., version 2.0, 3.0, 4.0 ...). Small changes or additions to existing methods will be reviewed in-house by Shenandoah or Mid-Atlantic Network staff. However, if a complete change in methods is sought, an outside review is required.

Revision History Log:

Prev. Version #	Revision Date	Author	Changes Made	Reason for Change	New Version #