

Standard Operating Procedure 4: Data Entry and Management

Version 1.0

In Protocol for Monitoring Mercury in Resident Lake Fish

Prepared by:

David D. VanderMeulen¹, Richard Damstra¹, Mark Hart¹, and Krista K. Bartz^{2,3}

¹ National Park Service
Great Lakes Network
2800 Lake Shore Drive East
Ashland, Wisconsin 54806

² National Park Service
Southwest Alaska Network
240 West 5th Avenue
Anchorage, Alaska 99501

³ Contact author: krista_bartz@nps.gov

September 2018

Suggested citation:

VanderMeulen, D. D., R. Damstra, M. Hart, and K. K. Bartz. 2018. Standard operating procedure 4: Data entry and management – Version 1.0. *In* K. K. Bartz, K. M. Junghans, and D. D.

VanderMeulen. 2018. Protocol for monitoring mercury in resident lake fish – Version 1.0: Southwest Alaska Inventory and Monitoring Network. Natural Resource Report. NPS/SWAN/NRR—2018/1734. National Park Service, Fort Collins, Colorado.

Contents

	Page
Figures.....	iii
Tables.....	iii
Revision History Log.....	1
Acknowledgements.....	1
Introduction.....	2
Data Stewardship Roles and Responsibilities.....	2
NPSTORET Database.....	2
Data Management Procedures	4
Data Collection.....	4
Data Entry, Verification, and Documentation	4
Data Validation and Certification.....	6
Data Analysis and Reporting.....	6
Data Folder and File Organization	7
Data Archival and Distribution	7
Literature Cited	8

Figures

	Page
Figure 1. Schematic of data flow to the STORET database.	3

Tables

	Page
Table 1. Data stewardship responsibilities related to monitoring Hg in fish.	3

Revision History Log

The following table lists all edits and amendments to this document since the original publication date. Information entered in the log must be complete and concise. Users of this Standard Operating Procedure (SOP) will promptly notify the Southwest Alaska Network (SWAN) project manager and/or data manager about recommended and required changes. The project manager must review and incorporate all changes, complete the revision history log, and change the date and version number on the title page and in the header of the document file. For complete instructions, please refer to SOP 8 (Protocol Revision).

Revision History Log:

Previous Version #	Revision Date	Revision Author	Revision Description and Location in SOP	Justification for Revision	New Version #
—	—	Bartz	Initial creation	—	1.0

Add rows as needed for each change or set of changes tied to an updated version number

Acknowledgements

This SOP was adapted from a similar SOP in the Great Lakes Network (GLKN) protocol for monitoring mercury in dragonflies and fish (VanderMeulen et al. 2018). The GLKN SOP was based on the GLKN Data Management Plan (Hart and Gafvert 2006), as well as SOPs from other networks.

Introduction

Data collected under the mercury (Hg) in resident lake fish protocol must be entered, quality-checked, documented, managed, and made available to others for a variety of purposes, such as decision-making, education, and research. The National Park Service (NPS) Water Resources Division (WRD) developed a database application called the Water Quality Database Template, or NPSTORET, to facilitate management of environmental data. NPSTORET facilitates the importing of data into a copy of Earthsoft's EQuIS database maintained by WRD. From the EQuIS database, water quality data are imported into the National STORET (STOrage and RETrieval) Data Warehouse administered by U.S. Environmental Protection Agency (EPA), as required by the water quality component of the Natural Resource Challenge (NPS 1999), NPS-75, and NPS Reference Manual 77. NPSTORET is best suited for environmental data from discrete sampling during individual site visits.

This Standard Operating Procedure (SOP) outlines data stewardship roles and responsibilities, and provides specific instructions and references for entering, quality checking, and managing data collected through the Hg in fish protocol.

Data Stewardship Roles and Responsibilities

Data stewardship is for the process of managing data and metadata so they are organized, useful, compliant, available, and safe. The demand for detailed, high quality data and metadata requires a group of people working together to ensure that data are collected using appropriate methods, and that resulting datasets, reports, maps, and other derived products are well managed.

The Southwest Alaska Network (SWAN) Aquatic Ecologist serves as project manager for monitoring Hg in fish. The project manager will provide project oversight, supervise data collection, and create cohesive links among data collection, synthesis, interpretation, and reporting.

While the project manager must act as the steward for data collected through this protocol, other SWAN personnel are also accountable for specific data management tasks. Table 1 lists stewardship responsibilities of personnel involved in the management of Hg data. To ensure that data are managed properly, individuals must understand their responsibilities, communicate with one another, and assist one another as needed.

NPSTORET Database

Mercury data (and associated metadata) collected through this protocol will be managed according to guidelines from the NPS Water Resources Division. In accordance with these guidelines, the desktop database application NPSTORET will be used to enter, store, document, and transfer data. The SWAN will oversee the use of NPSTORET and ensure that data are transferred after each sampling cycle to the NPS Water Resource Division for upload to the EPA STORET database/Water Quality Exchange (WQX) (Figure 1).

Table 1. Data stewardship responsibilities related to monitoring Hg in fish.

Personnel Role	Data Stewardship Responsibilities
Project crew member	Crew members collect and manage data with direction and guidance from the crew leader and/or project manager. Data collection includes operating sampling equipment, collecting samples, and recording measurements and observations. Crew members are responsible for quality control by following data collection and recording instructions and by promptly verifying recorded data. Crew members may also perform data entry and verification.
Project crew leader	The crew leader normally performs the same duties as the other crew members and ensures adherence to data collection and processing protocols, including data verification and documentation. The crew leader also works with the project manager and data manager on data management in the office.
Project manager (Network Aquatic Ecologist)	The project manager is responsible for all project operations and results, and may also participate in field operations. The project manager ensures that data management activities are conducted according to established procedures. He/she is responsible for validating and certifying data, approving the data content, quality, and documentation, as well as making decisions about data sensitivity and distribution. The project manager is responsible for evaluating project data at specified intervals, analyzing data for trends, and following reporting requirements.
Network Data Manager	The network data manager, together with the project manager, ensures that data are organized, useful, compliant, available, and safe. The network data manager provides the most current version of NPSTORET and works with project personnel to ensure Network data are received by WRD. The network data manager oversees activities related to training, user support, quality assurance, documentation, backups, archiving, and data maintenance and distribution.

STORET is an interagency water quality database developed and supported by the EPA to house local, state, and federal water quality data collected in support of managing the nation's water resources under the Clean Water Act. NPS Director's Order 77 indicates that the NPS should archive water quality data in STORET, and the NPS WRD mandates that any data collected as part of a WRD-funded project get archived in STORET. The NPS uses STORET as a repository of physical, chemical, biological, and other monitoring data collected in and around park units by park staff, contractors, and cooperators. The NPS operates its own Service-wide copy of STORET and makes periodic uploads to the EPA STORET National Data Warehouse so that data collected by and for parks will be accessible to the public.

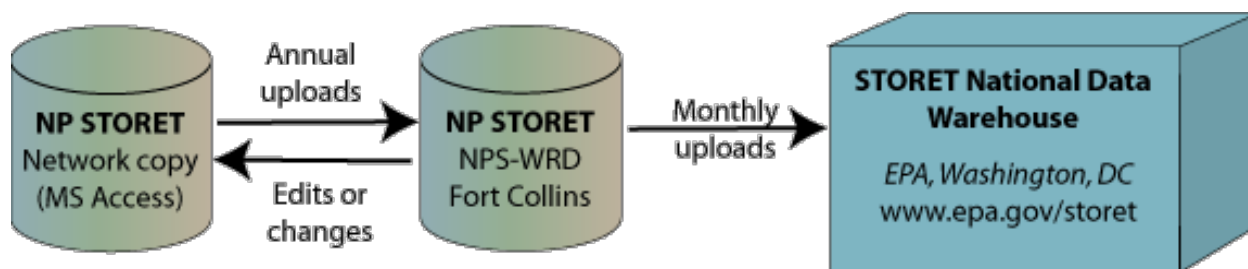


Figure 1. Schematic of data flow to the STORET database: physical, chemical, and biological data are transferred from the Network to the national master database and ultimately to the master database. Public access to STORET is available via an online clearinghouse.

The database design for NPSTORET is described in its associated documentation. Maintenance and documentation of NPSTORET is performed by NPS WRD. Because it must be compatible with the Oracle-based EPA STORET, NPSTORET is a complex Microsoft (MS) Access implementation. The latest version of NPSTORET can be found at: <http://nrdata.nps.gov/programs/water/npstoret/>.

The SWAN will maintain one master copy of NPSTORET at the Alaska Regional Office (AKRO) on a central server. This is the only copy of NPSTORET that can be used to export data to other locations (i.e., WRD). Additional copies of NPSTORET can be used by SWAN personnel stationed at parks, but they can only be used as a conduit for data entry and the importation of data to SWAN's master version of NPSTORET. For analysis, the data from the master copy of NPSTORET, that has passed all quality assurance/quality control (QA/QC) procedures, must be used.

Data Management Procedures

Data Collection

Data values are measured, observed, or estimated according the SWAN protocol for monitoring Hg in fish at various monitoring locations (sample sites) and recorded on field forms (see SOP 3, Field Work). Crew members are responsible for legible, accurate entries on field sheets, specimen labels, shipping labels, and chain-of-custody forms.

GPS coordinates are stored as waypoints if using a recreational GPS unit, or as features if using a mapping-grade unit with a data dictionary, and recorded on the field sheets at the time of sampling. The hard copy of the data serves as a back-up should something happen to the electronic data.

Digital images of sample sites are acquired during site establishment and periodically as sites change (SOP 3). Digital images of individual fish are also taken during processing (SOP 3). Crew members are responsible for proper settings and use of digital camera equipment and should refer to the user manual for details specific to the camera.

Fish are collected, processed, packaged, and shipped for laboratory analysis according to the steps outlined in SOP 3. Sample identification numbers on field forms, specimen labels, chain-of-custody forms, and laboratory reports facilitate the management of laboratory results.

Data Entry, Verification, and Documentation

Requirements for data entry into NPSTORET are detailed in the documentation for users. We present a summary of data entry below, and refer staff to the most recent NPSTORET documentation for the specifics.

On a regular schedule approved by the SWAN project manager and data manager, the crew leader will gather the field forms and verify their completeness, accuracy, and legibility. The crew leader will then make a digital scan of each field form, ensuring scans are legible. Scans will be printed and placed in the office binder kept at the park, and originals will be sent to the project manager for archival in the SWAN office at the AKRO. Additional prints will be made as needed. Upon receiving the original field forms, the project manager will proofread them to ensure that they are complete. All field forms should have been reviewed for completeness while in the field, however some

deficiencies in data recording may not be identified until all field forms have been reviewed as a group.

Project personnel will enter site establishment data in the NPSTORET Station Entry Template as soon as possible following the initial site visit to each sample location. This prompt action is a good data management practice and an NPSTORET requirement that enables parameter data entry. Linking digital images of sample sites to stations in NPSTORET is required as part of this data entry process. File size for digital images linked to stations in NPSTORET should normally be at least 100 kb and less than 300 kb.

Any electronic data (e.g., waypoints on a GPS device) will be downloaded to a computer in an MS Excel spreadsheet or text file at the end of each sampling day or as soon as feasible after sampling, and verified for accuracy. Likewise, at the end of each sampling day or as soon as feasible after sampling, all digital images will be downloaded to a computer and labeled with date, location, and subject matter. Import data from the MS Excel spreadsheet into NPSTORET following the guidelines.

Results of laboratory analyses typically are sent to the project manager in a MS Excel spreadsheet and on hard copy forms. The project manager will verify both sets of data and follow up with the contract laboratory if any discrepancies are noted or any questionable results are reported, in accordance with QA/QC procedures listed in the Data Quality Standards (DQSs; Bartz et al. 2018). Project staff will import the verified data into NPSTORET.

A random check of 3% of all electronically-transferred data (GPS or laboratory data) will be performed after import into NPSTORET, to ensure no problems occurred during data transfer. For electronically transferred or manually-entered data, the project leader will resolve errors according to established procedures. See the DQSs for additional QA/QC details.

Data verification starts with the QA/QC steps that are detailed in the DQSs. As data are entered into NPSTORET, either by keyboard or import routine, a suite of QA/QC procedures exist that compare the entered data with expected formats and accepted data value ranges or domains. If the entered data do not pass a form-based QA/QC test, NPSTORET prompts the user to make corrections and re-enter the data. For most of the form-based tests, the NPSTORET database will not accept out-of-bounds data, and correction is mandatory; for some range tests, out-of-bounds data are accepted after a user prompt, but are flagged accordingly. The QA/QC procedures on imported data perform similar tests, but exceptions that do not pass QA/QC are presented to the user and must be reconciled before the record or dataset can be accepted by NPSTORET.

Additional comparison tests will be run on data in NPSTORET to flag records that are outside the expected rate of change for a parameter based on previous records for that parameter. These suspect data points are reviewed by an expert user and can be corrected, flagged, or excluded from the dataset.

As mentioned above, NPSTORET performs numerous QA/QC checks on the data as they are entered and stored. Additionally, once the data are transferred to WRD, more QA/QC procedures are

performed. Each individual record in the dataset always has one, and only one, status flag indicating its status as provisional or final. Finalized data can still be edited and changed if errors are discovered after review in reporting and analysis, but the status of those records reverts to provisional and they must be resubmitted to the master STORET version at WRD and returned before they can be reassigned as finalized data records.

A user's identification is assigned to each aspect of data handling, from collection through the final steps of QA/QC. On data sheets, the identity of the data collector(s) is recorded and is transcribed into the data entry forms in NPSTORET. Additionally, NPSTORET records the user name (login) for every table entry or modification. Only qualified users who have been trained and given edit permissions are allowed to edit data in NPSTORET. These procedures protect the integrity of the data and allow the history of each data record to be traced.

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 Content Standard for Digital Geospatial Metadata. Monitoring Hg in fish meets the definition of geospatial data and, thus, the SWAN is responsible for documenting all public datasets using metadata that meet that standard. NPSTORET requires and stores a great deal of metadata at the record and project level. Metadata will be developed, using NPSTORET as the primary source that meets the federal standard. Consult the SWAN Data Management Plan (Mortenson 2006) for additional details about metadata procedures and requirements.

Data Validation and Certification

Prior to distributing the data for any type of use, the project manager validates the NPSTORET database content in the master data file, according to procedures in the DQs. Validation is performed at least once each year after data collection and entry are complete, and before data are submitted to the WRD by the end of each calendar year. If this annual validation does not meet scheduled reporting requirements of the Network, then the project manager may coordinate more frequent validation to meet those requirements.

Laboratory results must be imported and validated in NPSTORET as soon as possible following receipt of the results. The project manager will coordinate receiving and entering data with project staff and will then perform or coordinate the validation of lab results.

Data certification will be performed annually according to policies in the SWAN Data Management Plan. The certification process will take place before the finalization of the annual NPSTORET submission to WRD.

Data Analysis and Reporting

Project staff will follow the procedures for data analysis in SOP 5 and data reporting in SOP 6. Data summary statistics will typically include, but not be limited to:

- mean, median, mode;
- minimum, maximum, range;

- count, sum, percentiles (10, 25, 75, and 90th percentiles);
- standard error, standard deviation, variance; and
- kurtosis and skewness.

Some of these statistics are available through NPSTORET's query function and others can be derived from NPSTORET outputs. NPSTORET provides an export for any portion of the data in tabular format for use with the Data Analysis Toolpak.

All data reports disseminated for general (public) use will be based on certified data. In general, data will be certified prior to internal use for supporting park-, regional-, or national-level resource management decisions. However, in some cases the use of provisional data (data that have not undergone complete QA/QC as defined in the DQSSs) may be warranted to base decisions on best-available science. In addition, provisional data may be provided to internal and/or external science partners who are involved in data analysis and reporting, as a part of the data QA process. Biostatisticians and partnering scientists must examine data closely during the analytical phase, as they may find data that need to be excluded from analysis or qualified as to their validity.

Data Folder and File Organization

All data from the Hg monitoring protocol should be stored, as soon as possible, on the SWAN central server at T:\Vital_Signs\Water_Fish\Data\Fish_Contaminants. Files should be named in accordance with the conventions in the SWAN Data Management Plan. File names should include underscores in place of spaces and end with a date formatted as YYYYMMDD (e.g., 20180120).

Data Archival and Distribution

Data archiving serves two primary functions: it provides a source to retrieve a copy of any dataset when the primary dataset is lost or destroyed, and it provides a data record that is an essential part of the QA/QC process. Original data will be archived at the SWAN office. Original data for printed forms are either the physical datasheets or exact and complete digital copies of the forms that capture all entries and notations. The unedited files are the original data for digital data.

All digital data will have a duplicate file created at the earliest opportunity. At least two complete copies of any dataset are required by WRD, including digital replicas (scanned versions) of hard copy field sheets. Digital field data that are entered directly into a field computer or collected from a data logger must be backed up to a second medium at the earliest possible opportunity. The data files on field computers and loggers must not be erased until the integrity of these data files are verified on the duplicate storage medium. The removal of original data files from a field computer or logger must be a balance of keeping memory available for new data collection and a need to keep data in their most original form for as long as possible. Field files should only be deleted when memory space is needed for new data collection.

The SWAN's master version of NPSTORET will be maintained on a central server at the AKRO that is backed up daily, and backed up off-site weekly. Complete details of the SWAN Server archiving procedure are found in the Infrastructure chapter of the SWAN Data Management Plan; the general strategy for data archiving is also described in this plan and its appendices.

The SWAN has a various avenues for data distribution depending on the audience served and the degree of analysis and customization needed by the end-user (see Mortenson 2006 for details). The NPS IMD Integrated Resource Management Applications (IRMA) web portal is the primary tool that provides interactive access to SWAN data. Some SWAN datasets are available through links to the IRMA portal on the SWAN webpages. However, due to the complexity of the NPSTORET database, providing the data in that format will not meet the needs of those users unfamiliar with that application. Network (SWAN) versions of NPSTORET are used to update a master version of STORET maintained by WRD. The WRD master copy of STORET data is used to regularly update the EPA database that provides online access to SWAN water quality data through the EPA WQX and forthcoming Water Quality Portal.

Data distribution will primarily occur through the EPA STORET portal, the NPS Data Store, and the NPS Data Visualizer. These tools are described previously in this document. The USGS and EPA are developing the Water Quality Portal (www.waterqualitydata.us). This will probably become the primary tool for reaching data stored in EPA STORET/WQX. For many users, the EPA tools will be the easiest way to retrieve subsets of SWAN Hg data.

In cases where specific data are requested that are unavailable through the public distribution methods outlined above, the project manager will respond based on the nature of each request and the state of the data, for example, whether or not the data are fully qualified and documented to meet the needs of the requestor.

Literature Cited

- Bartz, K. K., C. J. Sergeant, and D. D. VanderMeulen. 2018. Data quality standards for monitoring mercury in resident lake fish – Version 1.0: Southwest Alaska Inventory and Monitoring Network. Natural Resource Report NPS/SWAN/NRR—2018/1731. National Park Service, Fort Collins, Colorado.
- Hart, M., and U. Gafvert (editors). 2006. Data management plan: Great Lakes Inventory and Monitoring Network. Great Lakes Inventory and Monitoring Network Report. GLKN/2006/20. National Park Service, Ashland, Wisconsin.
- Mortenson, D. 2006. Data management plan: Southwest Alaska Network, Inventory and Monitoring Program. National Park Service. Anchorage, Alaska.
- NPS (National Park Service). 1999. Natural Resource Challenge. Available at <https://www.nps.gov/nature/pressrelease.htm> (accessed 9 October 2018).
- VanderMeulen, D. D., B. Route, J. Wiener, R. Haro, K. Rolffhus, M. Sandheinrich, S. J. Nelson, A. Klemmer, C. Eagles-Smith, and J. Willacker. 2018. Protocol for monitoring mercury in dragonfly larvae and fish (version 1.0): Great Lakes Inventory and Monitoring Network. Natural Resource Report NPS/GLKN/NRR—2018/1726. National Park Service, Fort Collins, Colorado.