

# **Standard Operating Procedure: Data Management**

*Arctic Inventory and Monitoring Network*

*Stream Communities and Ecosystems Monitoring Protocol*

*Lake Communities and Ecosystems Monitoring Protocol*

Jonathan A. O'Donnell

National Park Service

240 W. 5<sup>th</sup> Avenue

Anchorage, Alaska 99501

September 2018

U.S. Department of the Interior

National Park Service

Natural Resource Stewardship and Science

Fort Collins, Colorado

Please cite this Standard Operating Procedure (SOP) as:

O'Donnell, J. A. 2018. Standard Operating Procedure: Data Management, Version 1.0. Stream Communities and Ecosystems Monitoring Protocols, Arctic Inventory and Monitoring Network. National Park Service, Fairbanks, Alaska. Available online at <https://irma.nps.gov/DataStore/Reference/Profile/2254732>

The main Narrative for the Protocol for monitoring Stream Communities and Ecosystems for the National Park Service's (NPS) Arctic Inventory and Monitoring Network (ARCN) is available from the NPS ARCN website (<http://science.nature.nps.gov/im/units/arcn>).

# SOP: Data Management

Version 1.0 (September 2018)

| Version No. | Revision Date | Author | Changes Made | Reason for Change |
|-------------|---------------|--------|--------------|-------------------|
|             |               |        |              |                   |
|             |               |        |              |                   |
|             |               |        |              |                   |
|             |               |        |              |                   |
|             |               |        |              |                   |

## Contents

|                                                  | Page                                |
|--------------------------------------------------|-------------------------------------|
| Figures.....                                     | iv                                  |
| Tables.....                                      | <b>Error! Bookmark not defined.</b> |
| Appendices or Appendixes .....                   | <b>Error! Bookmark not defined.</b> |
| Introduction.....                                | 5                                   |
| Data Stewardship Roles and Responsibilities..... | 5                                   |
| NPSTORET Database.....                           | 6                                   |
| Database Design .....                            | 6                                   |
| Field observations and measurements .....        | 8                                   |
| Data Validation and Certification.....           | 10                                  |

## Figures

|                                                                                                                                                                                                                                                                   | Page |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| <b>Figure 1.</b> Schematic of data flow to the STORET database: water quality 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. .... | 6    |
| <b>Figure 2.</b> Schematic of data flow to the Aquarius database: water quality data from data loggers are transferred from the Network to the WRD-hosted version of Aquarius. Public access to this data is provided through the Aquarius WebPortal. ....        | 7    |
| <b>Figure 3.</b> Arctic Network folder structure for information related to vital sign monitoring in general (left panel) and to the specifics of Stream Communities and Ecosystems vital sign (right panel). ....                                                | 12   |

## Introduction

Water quality data associated with the Stream Communities and Ecosystems monitoring protocol (O'Donnell and Miller 2018a) and the Lake Communities and Ecosystems monitoring protocol (O'Donnell and Miller 2018b) for the Arctic Inventory and Monitoring Network (ARCN) must be entered, checked for quality, documented, managed, and made available to others. 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 discrete water quality data. NPSTORET facilitates the importing of water quality data into a copy of Earthsoft's EQuIS database maintained by WRD. From EQuIS, water quality data is imported into the National STORET (STORage and RETrieval) Data Warehouse, which is administered by 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 water quality data that comes from discrete sampling during individual site visits.

To better accommodate continuous or time-series data that are generated from automated data loggers, WRD adopted Aquatic Informatics' Aquarius system for storage and visualization of continuously sampled water quality data. While the use of Aquarius is not currently mandated, ARCN personnel are encouraged to use Aquarius for data entry, analysis and visualization of continuous monitoring data.

This SOP outlines data stewardship roles and responsibilities, and provides specific instructions and references for entering, processing, and managing water quality data generated by both the Stream and Lake Communities and Ecosystems vital signs.

## Data Stewardship Roles and Responsibilities

Data stewardship is for the process of managing data and information resources that are organized, useful, compliant, available, and safe. The demand for detailed, high quality data and information about water quality 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 ARCN aquatic ecologist serves as project manager for both the Stream and Lake Communities and Ecosystems monitoring. The project manager will supervise data collection, provide project oversight, and provide cohesive links among data collection, synthesis, interpretation, and reporting. The network data manager, together with the project manager, ensures that water quality monitoring data are organized, useful, compliant, available, and secure. The network data manager works with project personnel to ensure Network water quality data are received on schedule by WRD. The network data manager oversees activities related to training, user support, quality assurance, documentation, backups, archiving, and data maintenance and distribution.

## NPSTORET Database

Water quality data are 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 discrete water quality data. The ARCN oversees the use of NPSTORET per the Network's water quality monitoring protocols and ensures that data are transferred at least annually to the NPS Water Resource Division for upload to the EPA STORET database/Water Quality Exchange (WQX) (Figure 1).

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 aquatic monitoring data collected in and around national 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.

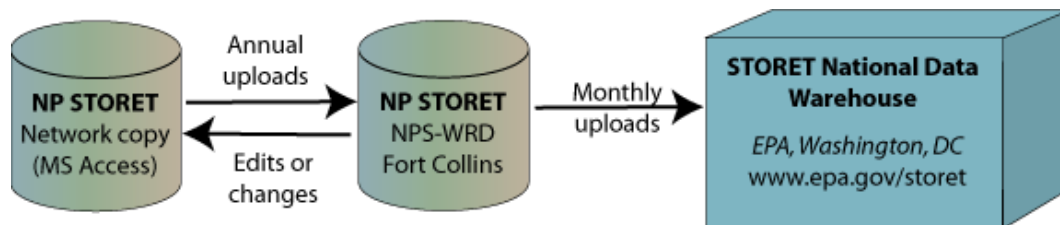


**Figure 1.** Schematic of data flow to the STORET database: water quality data are transferred from the Network to the national master database. Public access to STORET is available via an online clearinghouse.

### Database Design

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 MS Access implementation. The latest version of NPSTORET can be found at: <http://nrdata.nps.gov/programs/water/npstoret/>.

The project manager for the ARCN Streams and Lake Communities and Ecosystems vital signs will maintain one master copy of NPSTORET at the Alaska Regional Office in Anchorage. This copy of NPSTORET that can be used to export data to other locations (i.e., WRD and ARCN's vital sign databases). For analysis, the data from the master copy of NPSTORET, that has passed all QA/QC procedures, must be used.

## Aquarius Database

Guidelines for managing time-series data are being developed by the NPS Water Resources and Inventory and Monitoring (IMD) divisions. For the ARCN project manager, data collection, management, and stewardship responsibilities for time-series data are the same as those for discrete data. Aquarius is a sophisticated tool that allows the project manager to load data, inspect data, and correct data using a suite of automated processes. At the time of this latest SOP revision (Version 1.0), WRD and IMD are in the process of creating SOPs to use Aquarius tools to import, process, validate, and certify continuous data. A general data flow schematic for continuously collected data is shown in Figure 2.

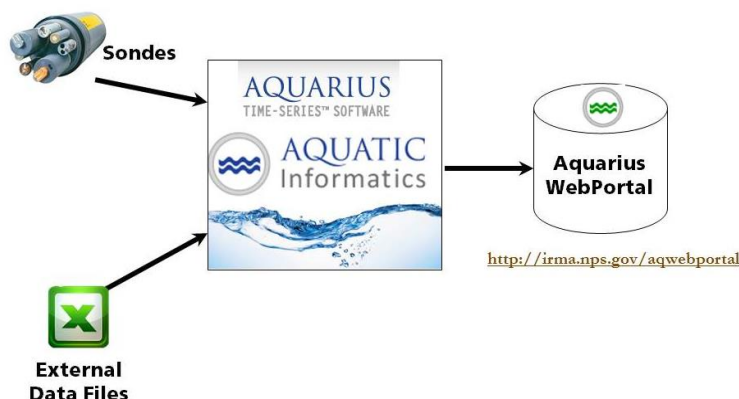


**Figure 2.** Schematic of data flow to the Aquarius database: water quality data from data loggers are transferred from the Network to the WRD-hosted version of Aquarius. Public access to this data is provided through the Aquarius WebPortal.

## Data Management Procedures

### Data Collection

Data values are generated from *in situ* field observations and measurements, laboratory analyses of discrete water samples, and continuous monitoring sensors following ARCN Stream Communities and Ecosystems monitoring protocol (O'Donnell and Miller 2017a (in review) and the ARCN Lake Communities and Ecosystems monitoring protocol (O'Donnell and Miller 2017b (in review)). Field measurements and observations are recorded on a field sheet and field notebooks (following *SOP: Field Measurements*). ARCN personnel are responsible for legible, accurate entries on field forms and in log books, including the calibration log. As a first step to verify data, crew members will check and double-check the recorded values on the day of data collection.

Discrete data collected with a multi-parameter sonde are stored directly on a datalogger attached to the sonde 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.

For the Stream Communities and Ecosystems vital sign, continuous time-series data are generated by multi-parameter sondes that are deployed seasonally in headwater streams. For the Lake

Communities and Ecosystems vital sign, continuous temperature profiles are generated by temperature sensor array mounted to customized lake buoys. Continuous data collected by multi-parameter sonde are stored directly on the sensors, and are downloaded and transferred to a computer following sonde retrieval prior to river freeze-up. For the lakes, continuous temperature data are transferred in real-time to a webserver using Iridium satellite technology. Hard copies of continuous temperature data do not exist, although electronic data are stored in multiple locations and backed-up daily.

Digital photographs of sample sites are acquired during site establishment and periodically as sites change (*SOP: Field Measurements*). 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.

GPS coordinates are stored as waypoints and recorded on the field sheets. When possible, the GPS data will be differentially corrected to improve the accuracy of location coordinates.

Water samples are collected, labeled, and packaged for laboratory analysis according to *SOP: Chemical Characterization of Surface Water*. Identification numbers on sample containers, chain of custody forms, laboratory reports, and on the field data collection form facilitate management of laboratory results.

### **Data Entry, Verification, and Documentation**

Requirements for data entry into NPSTORET are detailed in the documentation for users (<https://www.nature.nps.gov/water/vitalsigns/vitalsignsmgt.cfm>). We present a summary of data entry below, and refer staff to the most recent NPSTORET documentation for the specifics.

#### ***Field observations and measurements***

On a regular schedule, the project manager gathers the field sheet and notebooks where field observations and measurements were recorded. The project manager verifies the completeness, accuracy, and legibility of each form. Following each round of sampling, the crew leader will make a digital scanned copy of each field sheet and notebook page and ensure each scan is legible. The copy will be saved on the Stream Communities and Ecosystems vital sign's folder on the Arctic Network's O: drive. The original field sheets and field notebooks will be archived by the project manager. Additional photocopies will be made as needed. Upon receiving the original data sheets, the project manager will proofread the datasheets, making sure that they have been filled out completely. All data sheets should have been reviewed for completeness while in the field, however some deficiencies in data recording may not be identified until all data sheets have been reviewed as a group.

Project staff enters 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 100kb and less than 300kb. Project staff will reduce the size of image files larger than 300kb and copy images to

a file folder named “images” for the appropriate park. Original ‘raw’ images can be stored in the Pictures folder on the “O Drive” (see Data Archival and Distribution Procedures below) if project staff determine it is important to keep higher resolution images.

Discrete electronic data from the multiparameter sondes are downloaded from the datalogger to a computer in an MS Excel spreadsheet or text file at the end of each sampling day or as soon as possible thereafter, 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. Any notes taken in the field regarding collection of data with the sonde are transcribed to NPSTORET in a field for comments and notes.

Results of laboratory analyses from CCAL are typically sent to the project manager in a MS Excel spreadsheet via email. The project manager will verify datasets and follow-up with the CCAL laboratory manager if any discrepancies are noted or any questionable results are reported (see *SOP: Data Quality Assurance and Quality Control*). Project staff will import the verified data into NPSTORET.

Non-electronic field data recorded on hard copy data sheets are manually-entered into a MS Excel spreadsheet. At the end of the season when all data are entered, the project leader will perform or coordinate a check these manually-entered data prior to import into NPSTORET during the Data Verification process.

Electronically transferred data (GPS, multiparameter sonde, and laboratory data) will also receive random check after it is imported into NPSTORET, to ensure no problems occurred during data transfer. For electronically transferred or manually-entered data, the project leader resolves errors according to established procedures.

Data verification starts with the QA/QC steps that are detailed in *SOP: Data Quality Assurance and Quality Control*. As data are entered into NPSTORET, either by keyboard or using an 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. The Network's water quality data records are regarded as being in provisional status until they are returned to ARCN from WRD, or are accepted by WRD without changes after the final QA/QC steps. Each individual record in the water quality 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 datasheets, 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.

The Aquarius database software used for continuous water quality data has built-in QA/QC tools specifically tailored for time series data. These feature tools are particularly useful for dealing with data gaps, overlapped and out of range values, and will be utilized as part of the QA/QC process used for continuous data imported into Aquarius. At the time of this SOP development WRD and IMD are working cooperatively to create SOPs for QA/QC and data certification (see next section) for the Aquarius tool. ARCN will adopt the SOP's for these tasks when they are completed.

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. Water quality monitoring meets the definition of geospatial data and, thus, ARCN is responsible for documenting all public datasets using metadata that meets that standard. NPSTORET and Aquarius require and store a great deal of metadata at the record and project level. Metadata will be developed, using NPSTORET and Aquarius as primary sources that meet the federal standard. Consult the ARCN Data Management plan (NPS 2008) 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 water quality data file according to procedures in *SOP: Data Quality Control and Quality Assurance*. The integrity of each master data file must be preserved during the validation process in cases where the person performing data validation does not have direct access to the master water quality data file. Validation is performed at least once each year after data collection and entry are complete and before data are submitted to the NPS Water Resource Division, usually 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 needs, or reports can include a statement explaining that results are based on data that have not

been validated. The statement should include an explanation of what significance this has for using preliminary data.

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

Data certification will be performed annually according to policies in the ARCN Data Management Plan (NPS 2008) and ARCN/CAKN [DMSOP-1](#) (Miller 2017). The certification process will take place before the finalization of the annual NPSTORET submission to WRD, and the release of data in Aquarius for public viewing.

### **Data Analysis and Reporting**

For discharge data, analysis and reporting are described in *SOP: Discharge Analysis*. For field measurements and water chemistry, we will examine temporal and spatial patterns for different parameters and chemical constituents. Summary statistics will be calculated, including but not limited to mean, median, standard deviation, and range. Water-quality data will be analyzed over space according to spatial variations across watersheds that vary with underlying parent material and ground ice content. Water-quality data will also be measured over time, at both annual and seasonal (continuous time-series data only) time scales.

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 in Microsoft Excel.

Aquarius has primarily been a tool used by hydrologists to compile stage and discharge data for flowing waters. It can generate the same data summary statistics as NPSTORET as well as a few others, such as reports by location showing minimum, maximum, mean, and count number of data points per time series. Aquarius can also be used to calculate means over several time scales (monthly, daily, or hourly) and to generate time series graphs. In addition to these and other analysis procedures, data can also be exported from Aquarius to flat file formats for use by other graphing or statistical software, such as R or SigmaPlot.

### **Data Archival and Distribution Procedures**

All data from the Streams and Lake Communities and Ecosystems protocol will be archived at the earliest possibility on the ARCN "O Drive". A diagram showing the folder structure is shown in Figure 3. Data archiving serves two primary functions: it provides a source to retrieve a copy of any data set 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 Network 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.

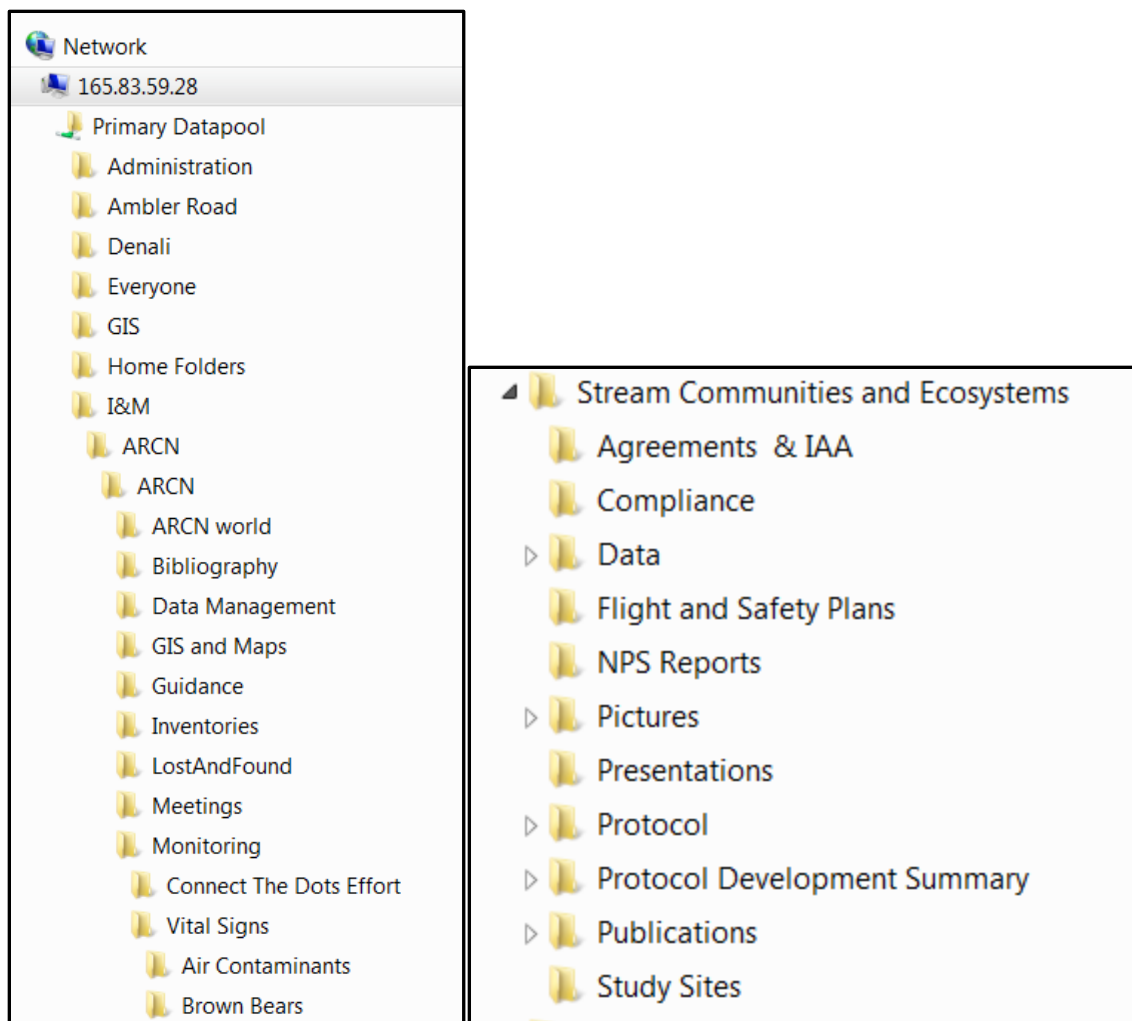


**Figure 3.** Example Arctic Network folder structure for information related to vital sign monitoring in general (left panel) and to the specifics of Stream Communities and Ecosystems vital sign (right panel).

All digital data have a duplicate file created at the earliest opportunity. At least two complete copies of any water quality dataset are required by WRD, including digital replicas (scanned versions), if they are created, of hard copy datasheets. 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 possibility. 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.

Raw time series data are stored on the ARCN “O Drive”, which is backed up daily, and backed up off-site weekly. The version of Aquarius that ARCN uses is housed on a server at WRD in Ft. Collins, Colorado. For the time series data to be stored and used in Aquarius, the flat file versions are uploaded to the SQL server using the Aquarius Springboard. All Aquarius data are uploaded to this

server at least annually by ARCN staff or cooperators. WRD is currently working on a public-facing web application where this time series data will be available, once the data is officially certified.

ARCEN has a number of different avenues for data distribution depending on the audience served and the degree of analysis and customization needed by the end-. The NPS IMD Integrated Resource Management Applications (IRMA) web portal (<https://irma.nps.gov/DataStore/>) is the primary tool that provides interactive access to ARCN data. Some ARCN water quality datasets are available through links to the IRMA portal on the ARCN webpages (e.g., <https://science.nature.nps.gov/im/units/arcn/vitalsign.cfm?vsid=13>). 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. ARCN versions of NPSTORET are used to update a master version of STORET maintained by NPS WRD. The WRD master copy of STORET data is used to regularly update the EPA database that provides on-line access to ARCN water quality data through the EPA WQX and forthcoming Water Quality Portal. NPS IMD is also developing a web map portal to monitoring data tentatively called the NPS Data Visualizer. This will allow a user to spatial query data, including monitoring data, and retrieve the scope of interest. The pathway for this tool is via a Common View Data Template that will allow data in NPStoret to be read by the database server supplying the Data Visualizer.

Data distribution will primarily occur through the EPA STORET portal, the NPS Data Store, Aquarius WebPortal 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](http://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 ARCN surface-water quality data.

In cases where specific data are requested that are unavailable through the public distribution methods outlined above, the project manager or appropriate water resource specialist 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

- Miller S.D. 2017. Product Archival, Seasonal Closeout and Dataset Certification Procedures for the Arctic and Central Alaska Inventory and Monitoring Networks. Standard Operating Procedure. National Park Service Arctic and Central Alaska Inventory and Monitoring Networks. Fort Collins, CO.
- NPS (National Park Service). 1999. Natural Resource Challenge. Available at [www.nature.nps.gov/challenge/index.cfm](http://www.nature.nps.gov/challenge/index.cfm) (accessed 26 January 2007).
- O'Donnell, J. A., and S. D. Miller. 2018a. Stream communities and ecosystems monitoring protocol for the Arctic Network, Alaska. Natural Resource Report NPS/ARC/NRR-2018/1725. National Park Service. Fort Collins, Colorado.
- O'Donnell, J. A., and S. D. Miller. 2018b. Lake communities and ecosystems monitoring protocol for the Arctic Network, Alaska. Natural Resource Report NPS/ARC/NRR-2018/1708. National Park Service. Fort Collins, Colorado.

The Department of the Interior protects and manages the nation's natural resources and cultural heritage; provides scientific and other information about those resources; and honors its special responsibilities to American Indians, Alaska Natives, and affiliated Island Communities.

NPS XXXXXX, Month Year (The Fort Collins Support Office will fill out this line for you)

**National Park Service**  
**U.S. Department of the Interior**



---

**Natural Resource Stewardship and Science**

1201 Oakridge Drive, Suite 150  
Fort Collins, CO 80525

[www.nature.nps.gov](http://www.nature.nps.gov)