

Standard Operating Procedure 5: Data Management

Version 1.0, June, 2018.

Change History

Original Version #	Date of Revision	Revised By	Changes	Justification	New Version #

Contents

Data Stewardship Roles and Responsibilities	4
Aquatic Ecologist	4
Hydrologic Technician	4
Data Manager	4
Data Flow	5
Raw Continuous Water Quality Records	7
Locations	7
Quality Assurance and Quality Control	7
Monitor Maintenance Logs	7
Cross-Section Surveys.....	7
Site Visits	8
Monitor Calibrations and Calibration Checks	8
Data Corrections and Ratings.....	8
Scanned Paper Data Sheets	8
Provisional Stage or Discharge Data (U.S. Geological Survey)	8
Continuous Water Quality Record Change List Reports.....	8
Certified Data	9
Certified Continuous Water Quality Records.....	9
Data Verification.....	10
Data Validation	10
Documentation and Certification	10
Data Archival	12
Data Versioning	13
Continuous Water Quality Data Processing with AQUARIUS.....	14
Literature Cited	16

Overview

This standard operating procedure (SOP) summarizes the data management tasks and responsibilities necessary to ensure that data are collected consistently for continuous river water-quality monitoring in the Eastern Rivers and Mountains Network. Much of this SOP is based on data management practices developed by the Southwest Alaska Network for freshwater monitoring (Shearer et al. 2015) and is intended primarily for ERMN staff use as a guide. Following the data stewardship roles and responsibilities provided in this document ensures that river water-quality monitoring data are consistently collected by the ERMN; however, this SOP is not an in-depth, step-by-step process because detailed processes are provided in other cited documents or on-line resources.

This SOP consists of three sections; the first discusses data stewardship roles and responsibilities, the second discusses data flow, and the last section describes the general procedure and resources for processing and certifying a continuous data record using NPS AQUARIUS Springboard Time Series software (hereafter, AQUARIUS; Aquatic Informatics, 2017).

Data Stewardship Roles and Responsibilities

The ERMN Aquatic Ecologist supervises data collection, provides project oversight, directs data collection and acquisition, and provides cohesive links among data collection, synthesis, interpretation, and reporting. The Aquatic Ecologist is also responsible for all other data management tasks not specifically mentioned in this SOP. While the Aquatic Ecologist must act as the primary steward for river water-quality monitoring data for the Network, other ERMN personnel are accountable for specific data management tasks. The following subsections outline data stewardship responsibilities of different ERMN staff related to river water-quality monitoring. To ensure that all data are managed properly, individuals must understand their responsibilities, communicate with one another, and assist each other as needed.

Aquatic Ecologist

The Aquatic Ecologist is responsible for and participates in all project operations. The Aquatic Ecologist ensures that data management activities are conducted according to established procedures and is responsible for data processing and certification, as well as making decisions about data sensitivity and distribution. The Aquatic Ecologist is responsible for evaluating project data at specified intervals, analyzing data for trends, and fulfilling reporting requirements.

Hydrologic Technician

The Hydrologic Technician works closely with the Aquatic Ecologist in all aspects of data management. The Hydrologic Technician's primary responsibility is to ensure that field data are collected and recorded in a manner consistent with SOPs. The Hydrologic Technician is also responsible for verifying and validating Quality Assurance and Quality Control (QA/QC) Information.

Data Manager

The Data Manager ensures that river water-quality monitoring data are organized, useful, compliant, available, and safe. The Data Manager oversees activities related to information technology acquisition and development, user support, quality assurance, documentation, backups, archiving, and data maintenance and distribution.

Data Flow

There are four data categories included in this protocol: Raw Data, Locations, Quality Assurance and Quality Control (QA/QC) Information, and Certified Data (Table 1). Continuous river water-quality records (i.e., time series) are the monitoring data collected and analyzed for this protocol; however, several types of supporting information are required to process and certify these records.

Data flow describes data handling, from collection to reporting and archiving (Figure 1). Data and metadata are collected throughout the monitoring process, not just during data collection and acquisition. For example, critical metadata that capture the spatial and temporal circumstances about where water-quality data are collected, which water-quality instruments are used, the instruments' characteristics (e.g., calibration information, serial number, model information) and deployment information, and the data collection methods employed — are all collected and managed. Capturing metadata at the appropriate time is critical to its usefulness, because obtaining metadata after the fact makes metadata generation difficult, if not impossible.

Data are collected primarily in electronic formats, and aside from continuous water-quality records, the majority of supporting data are entered into, and stored in, the ERMN River Water-Quality Monitoring database (hereafter, database). The database is used on a ruggedized laptop and includes two data entry forms (site visits and monitor calibration checks, Figure 2) that are linked when appropriate. Based on data entered into these forms, the database calculates 1) corrections for fouling and calibration drift and 2) data ratings. All additions (e.g., data entry) or changes to the database are performed on the working version, which resides on a ruggedized laptop PC — the ERMN Aquatic Ecologist routinely saves a backup copy of the database to another PC and the ERMN server. Additional details regarding data types are provided in the following subsections.

Table 1. Eastern Rivers and Mountains Network River Water-Quality Monitoring data formats.

Data Type	Format
Raw Data	
Raw Continuous Water Quality Records	*.csv
Locations	*.shp, *.shx, *.dbf
Quality Assurance and Quality Control	
Monitor Maintenance Logs	*.pdf
Cross-section Surveys	*.xls
Site Visits	*.accdb
Monitor Calibration and Calibration Checks	*.accdb
Data Correction and Rating	*.accdb
Scanned Paper Data Sheets	*.pdf
Provisional Stage or Discharge Data (U.S. Geological Survey)	*.csv
Continuous Water Quality Record Change List Reports	*.pdf
Certified Data	
Certified Continuous Water Quality Records	*.csv

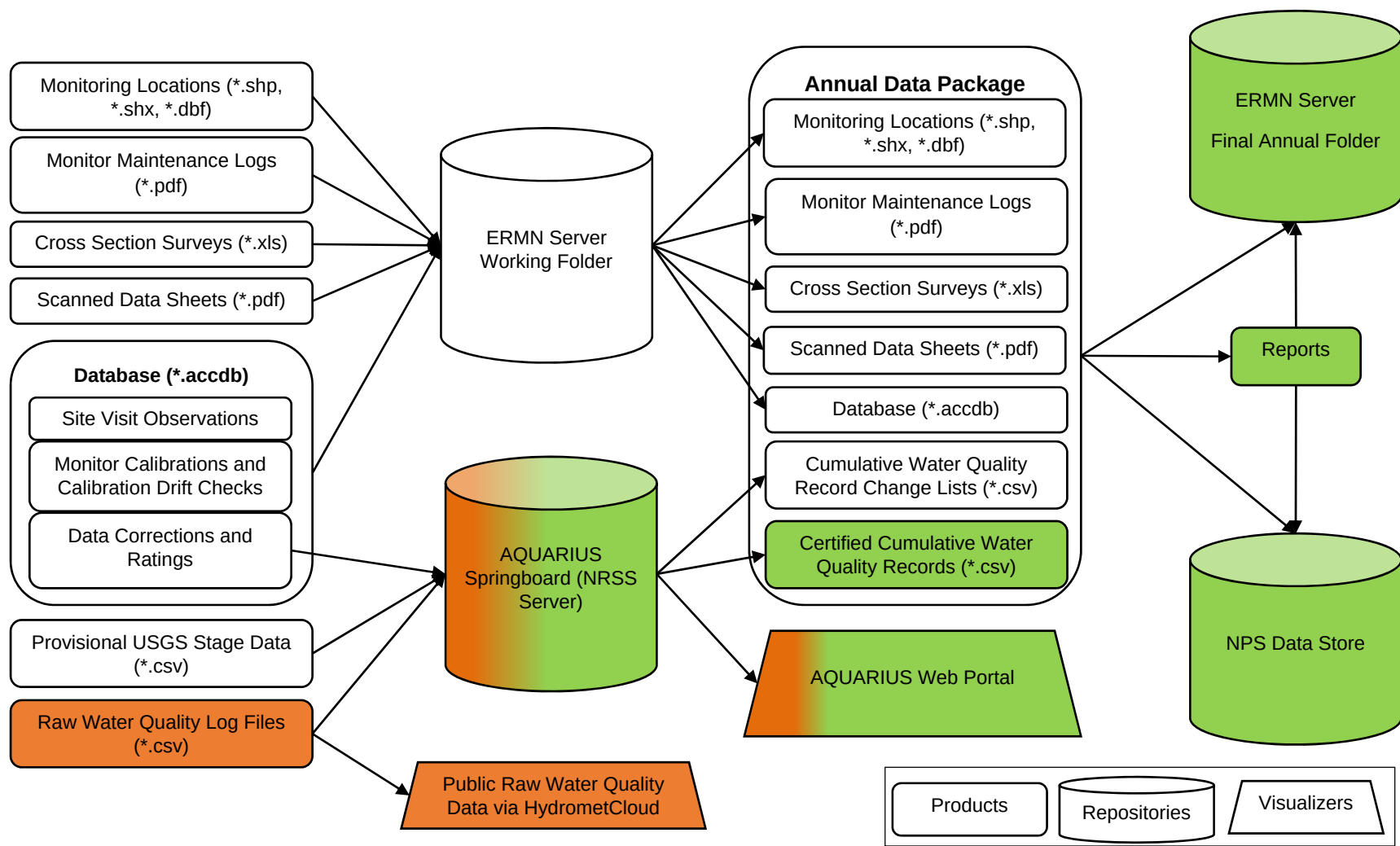


Figure 1. Eastern Rivers and Mountains Network (ERMN) continuous river water-quality data flow. Raw water-quality records (orange) are available in real-time, and after download, are certified (green) in AQUARIUS Springboard using quality assurance and quality control (QA/QC) information (white). The AQUARIUS Web Data Portal provides public access to all data. A Data Package that contains certified cumulative water-quality records and QA/QC information is archived annually on the ERMN server and NPS Data Store. Only certified data and supporting QA/QC information are used for analysis and reporting. NRSS = Natural Resource Stewardship and Science Directorate.

Continuous WQ Monitor Calibration Drift Check Sheet

Monitor: 3-MM11100550 Date: 10/2/2017 Military Time: 14:57 Cal_Technician: [Name]

Calibration Location: Site Full Check? Yes Data Verified: [Date]

Comments: Meter deployment at Montague. Need to fill in 10 sc exp date

Water Temperature Criteria: +/- 0.2 °C

Standard	Std Maker	Lot #	Exp Date	Check	Check Error	SRF	ReCal	ReCal Error
10 µS/cm	Alfa Aesar	W050034		10.7	0.7			
1000 µS/cm	Alfa Aesar	S02D070	12/1/2018	1018	18			
100 µS/cm								

pH Criteria: +/- 0.2 units

Standard	Std Maker	Lot #	Exp Date	Check	Check Error	SRF	ReCal	ReCal Error
7	BDH	2511878	11/1/2017	7.07	0.07			
10	BDH	2703F33	9/1/2018	10.09	0.09			
4								

Dissolved Oxygen (DO) Criteria: +/- 0.3 mg/L

Standard	Std Maker	Lot #	Exp Date	Check	Check Error	SRF	ReCal	ReCal Error
0 FNU	DI Water			0	0			
100 FNU	Eureka	c799577	2/1/2019	104.2	4.2			

Continuous WQ Monitor Site Visit Form

Site Name: LORDVILLE Visit Type: Routine Date: 10/12/2017 Military Time: 12:15 Tech: [Name]

Field Monitor: 1-MM11100553_10/9/2017 Last_Visit: 9/14/2017

Site Monitor: 5-MM11100555 Was Data Downloaded? Yes

Site Visit: pH electrolyte replaced and pH readings checked on monitor calibration form. Final site readings were taken at 14:00 local time, 18:00 was the Greenwich mean time.

Were Calibration Checks Done? Yes Checks: 5-MM11100555_10/12/2017 Open

Fouling Checks

Before Cleaning				After Cleaning				After Cleaning Criteria*	
Mt. Time	Site	Field	Error	Mt. Time	Site	Field	Error		
12:52	14.85	14.77	0.08	13:06	14.89	14.81	0.08	+/- 0.2°C	
	135.9	135.8	0.3		136.1	135.8	0.3	+/- 5µS/cm	
	7.66	7.76	0.07		7.73	7.66	0.07	+/- 0.25 units	
	9.88	10.04	-0.16		9.88	10.04	-0.16	+/- 0.4 mg/L	
	0.7	0.7	0.2		0.7	0.2	0.5	+/- 5.0 units	

*After Cleaning Error = (After Site) - (After Field)

Final Site Readings

Mt. Time	Is Parameter Order Correct?
18:00	Yes
Temp (°C)	15.3
SC (µS/cm)	88.2
pH (units)	7.48
DO (mg/L)	10.5
Turb (FNU)	7

Figure 2. Example data entry forms in the Eastern Rivers and Mountains Network River Water Quality Monitoring database.

Raw Continuous Water Quality Records

Continuous (i.e., 15 to 60 min. interval) water-quality records are logged by data collection platforms (DCPs) at monitoring sites and are typically downloaded in *.csv format during routine site visits. There are generally between 6 and 12 log files for each site/water year combination, depending on deployment period (i.e., growing season only versus year-round). It is important that these files are promptly imported into AQUARIUS for verification and validation. Prompt input minimizes the chances of accidental loss and allows the Aquatic Ecologist to identify and correct problems with monitoring site equipment as soon as possible.

Locations

Locations of monitoring sites and cross-section surveys are collected and archived with associated monitoring data. The coordinates of monitoring site infrastructure is corroborated with location data of co-located USGS stream gaging stations. Locations are updated when new sites are added or removed from the protocol.

Quality Assurance and Quality Control

Monitor Maintenance Logs

Handwritten, monitor-specific maintenance logs that detail non-routine maintenance (e.g., sensor replacements, 5-point thermistor checks) are scanned and archived for QA. This information provides evidence that the monitors were properly maintained but is not used for processing continuous water-quality records.

Cross-Section Surveys

Cross-section surveys are performed as frequently as possible during various flow conditions and seasonal variations throughout the year to determine how measurements from the continuous site monitor, located at a single point in the river, compare with cross-section (transect) values that are

representative of the entire river cross-section. These data are recorded and archived as *.xlsx files – there can be multiple archived files for every site, every year.

Site Visits

Site visits are categorized as monitor deployments, routine site visits, or monitor retrievals – all of these data are stored in the database. Monitor deployment data include date, time, deployed monitor serial number, and corresponding calibration check information associated with the beginning of monitor deployment periods. Routine site visit data include the same information collected during monitor deployments in addition to: 1) information derived from cleaning sensors, 2) corroborating data from a calibrated field monitor to site monitor measurements, 3) verifying sensor performance against known standards and 4) calibration verification if sensor readings exceed acceptance criteria. Monitor retrieval site visits occur at the end of monitor deployments and include 1) information derived from cleaning sensors and 2) verifying sensor performance against known standards.

Monitor Calibrations and Calibration Checks

These data are entered and stored in the database and are used to: 1) verify monitor accuracy, 2) calculate calibration drift (error) when processing continuous water-quality records, and 3) document when monitor sensors are calibrated. When appropriate, calibration check information is associated with site visits which allows for calculation of calibration drift corrections and final data ratings.

Data Corrections and Ratings

Data corrections (if applicable) and ratings are calculated and displayed by the database based on information collected during site visits. These ratings quantify the extent to which fouling and sensor “drift” affected the continuous water-quality record since the previous site visit. These data are used to process and certify data in AQUARIUS. Data corrections and ratings applied in AQUARIUS are recorded by the software and are exported and archived annually for each site-parameter combination in the AQUARIUS Change List Reports.

Scanned Paper Data Sheets

Data are almost always collected electronically (e.g., by the DCP or entered into database), but in rare instances, site visit or calibration check information may need to be recorded on paper data sheets. In these instances, data are transcribed into the database and the paper data sheets are scanned and saved in *.pdf format on the ERMN server.

Provisional Stage or Discharge Data (U.S. Geological Survey)

River stage and/or discharge data collected by the U.S. Geological Survey are periodically downloaded (*.csv files) from the World Wide Web and imported into AQUARIUS to aid visual interpretation during water-quality record processing. These data are temporarily stored in AQUARIUS but they are not archived.

Continuous Water Quality Record Change List Reports

Every certified continuous water-quality record is accompanied by a cumulative (i.e., entire record) Change List Report that details all modifications (e.g., corrections) that have been performed on the record in addition to the user and time stamp associated with the modifications. A change report is

generated for each site/parameter combination by using the AQUARIUS Reporting Toolbox and “Change List Report [System]” template.

Certified Data

Certified Continuous Water Quality Records

Certified continuous water-quality records have undergone all quality control procedures documented in this protocol and quality assurance plan. These records are stored in AQUARIUS and are annually exported to a cumulative (i.e., entire record) log file for each monitoring site and are archived in *.csv format. Each measurement in the certified record has 1) a raw value, 2) a corrected value, and 3) a corresponding data rating.

Data Verification

Data verification consists of checking forms (written and electronic) and data files for accuracy and completeness. Spot checking for consistency between the written data (if necessary) and the transcribed, electronic version of the same data (e.g., in the database) should be performed as part of the verification process. The Hydrologic Technician and Aquatic Ecologist should perform a check (that all database fields are populated and logical) as a routine step in the process of data entry and import. Verification is also performed using AQUARIUS by visually inspecting data plots and tables of water-quality records.

The working copy of the database is used by the Hydrologic Technician and Aquatic Ecologist before and during site visits; subsequently, the Aquatic Ecologist uses the database while processing water-quality records. The Aquatic Ecologist frequently reviews and verifies data entry forms and subsequently backs up the database to the ERMN server.

Data Validation

While data verification primarily involves identifying transcription or import errors, data validation involves assessing whether or not Measurement Quality Objectives are met. Data validation includes, but is not limited to: checking for out-of-bound values (domain checking), expected rates of change (e.g., data “spikes” and “dips”), missing or erroneous values, bias or drift, and outliers. Due to the large volume of data that will pass through the validation stage of the data life cycle, the primary validation methods are handled in AQUARIUS. If errors are discovered, the group of data associated with the suspect data is scrutinized and the cause is determined. Any changes to continuous water-quality records are documented in Change List Reports that are annually exported from AQUARIUS.

Documentation and Certification

Any dataset compiled by the ERMN 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/constan.html). All GIS data layers will be documented using applicable FGDC and NPS metadata standards. The Network will also generate additional metadata for non-spatial datasets to meet IMD metadata requirements.

Documentation involves the consolidation of metadata created up to this point in the data life cycle. All metadata, whether created before, during, or after data acquisition, should be captured in the appropriate location within AQUARIUS or stored on the ERMN server. The next step, certification, cannot occur until all metadata are captured and documented, and all data values have been entered, verified, and validated. In other words, certification is a confirmation by the Aquatic Ecologist that the data have passed all existing import, verification, and validation checks, and the data are

complete and well-documented (e.g., complete metadata). Data should not be released or used for ERMN analysis until documentation and certification are complete.

Data Archival

The Aquatic Ecologist archives all data in a “Working Folder” on the ERMN server as soon as possible after data collection (Figure 1). Archived digital data files should be in uniform, common, and easily accessible data formats (e.g., *.csv; Table 2) and the original data structure and supporting metadata should be updated annually during end of year closeouts. Standard file naming is essential for good data management – it provides the initial association of metadata (e.g. sampling location and date) with monitoring data. Hand-written data sheets are scanned and saved to a standard format, currently *.pdf.

After data have been certified at the end of each water year, an “Annual Data Package” that contains cumulative water-quality records and supporting QA/QC information is compiled by the Aquatic Ecologist and delivered to the Data Manager for permanent archive on the ERMN server and on the NPS Data Store. The Aquatic Ecologist and Data Manager complete a checklist (Appendix) during this process that is archived with the associated data.

Table 2. Eastern Rivers and Mountains Network River Water Quality Monitoring file naming conventions.

File Description	Naming Convention	Example
Raw Continuous Water Quality Records	Site Name + “log” + record end date (YYYYMMDD)	Lordville_log_20160516_0 ¹ .csv
Working Access Database (Site Visit and Monitor Calibration Data)	ERMN_WQ + current date (YYYYMMDD)	ERMN_WQ_20171005.accdb
Final Annual Access Database	ERMN_WQ + year certified (YYYY)	ERMN_WQ_2017. accdb
Certified Continuous Water Quality Records	Site Name + year certified (YYYY) + “Certified”	Lordville_2016_Certified.csv
Continuous Water Quality Record Change List Reports	Site Name + year certified (YYYY) + “ChangeList”	Lordville_2016_ChangeList.pdf
Cross-section Surveys	Site Name + “CrossSection” + Visit Date (YYYYMMDD) + UP or DN	Lordville_CrossSection_20160516_UP.xlsx
Monitor Maintenance Logs	Monitor SN + LabLog + Year (YYYY)	MM11100550_LabLog_2016.pdf

¹ Data collection platforms automatically append “_0” at the end of these files; subsequent files saved on the same date, which rarely occurs, would be numbered sequentially (i.e., _1, _2, _3, etc).

Data Versioning

Each time the database is altered on the field computer, it is backed up on the ERMN server and renamed on the field computer with the date it was altered using consistent naming conventions (Table 2). All field data entry and verification is conducted on the field computer.

AQUARIUS keeps an unaltered copy of all uploaded data. Subsequent changes to data are tracked in the database as a “change history”; consequently, the ERMN will not create separate archival data versions. Every year, a Certified Cumulative Water Quality Record and corresponding Change List Report are exported from AQUARIUS and are archived on the ERMN server and NPS Data Store. Change List Reports include a time- and user-stamp associated with all alterations to water-quality records.

Continuous Water Quality Data Processing with AQUARIUS

The ERMN uses the NPS service-wide AQUARIUS license to manage, certify, and analyze continuous water-quality monitoring data. This section provides a general outline of the ERMN approach to these procedures (Figure 3) but not step-by-step instructions. Training is necessary to proficiently use AQUARIUS; consequently, the Aquatic Ecologist has taken training courses provided by NPS and Aquatic Informatics. The NPS and Aquatic Informatics provide training resources at <https://nrddata.nps.gov/programs/water/aquarius/AquariusVideos.htm> and <http://aquaticinformatics.com/resources/aquarius-resources/video-library/>, respectively.

AQUARIUS includes toolboxes needed to complete each stage of managing continuous water-quality records, from importing data to saving it to the database; furthermore, AQUARIUS maintains a permanent record of all raw data and an audit log that automatically tracks corrections, processing, notes, and remarks for time series that can be exported for quality assurance purposes. AQUARIUS provides functionality for producing standard summaries and reports, and although it currently does not provide all the tools required for quantifying long-term trends, data can easily be exported for use in other software.

Users of AQUARIUS require access to the shared NPS license which is controlled by WRD. Dean Tucker (WRD) can provide “remote desktop” access and instructions (<http://nrddata.nps.gov/programs/water/aquarius/AquariusVideos.htm>) for access behind the NPS firewall or VPN connection. Users must log into the NRSS server via smartcard and then log into the AQUARIUS Springboard server using Active Directory credentials (e.g., ctzilkowski + password). Additional information on use of AQUARIUS can be found in the user manual and training videos, which can be accessed online at <http://aquaticinformatics.com/support/support-login/> or through the help menu. Access to this website requires a username and password, obtainable via email request to support@aquaticinformatics.com.

The ERMN has a “folder” in the NPS AQUARIUS server that contains all continuous water-quality monitoring sites maintained by the network. The four river monitoring sites are named consecutively from “01” through “05” with the prefix “ERMN” (Table 3) – the ERMN02 site (Delaware River near Barryville, NY), was discontinued but that site identification code is reserved for the site (Delaware River at Pond Eddy, NY) scheduled for installation in 2019. These site identifiers should not be changed.

Table 3. Site identification codes (ID) and names (Name) of Eastern Rivers and Mountains Network continuous water-quality monitoring sites.

ID	Name
ERMN01	Delaware River at Lordville, NY
ERMN03	Delaware River at Montague, NJ
ERMN04	Delaware River near Delaware Water Gap, PA
ERMN05	New River at Thurmond, WV

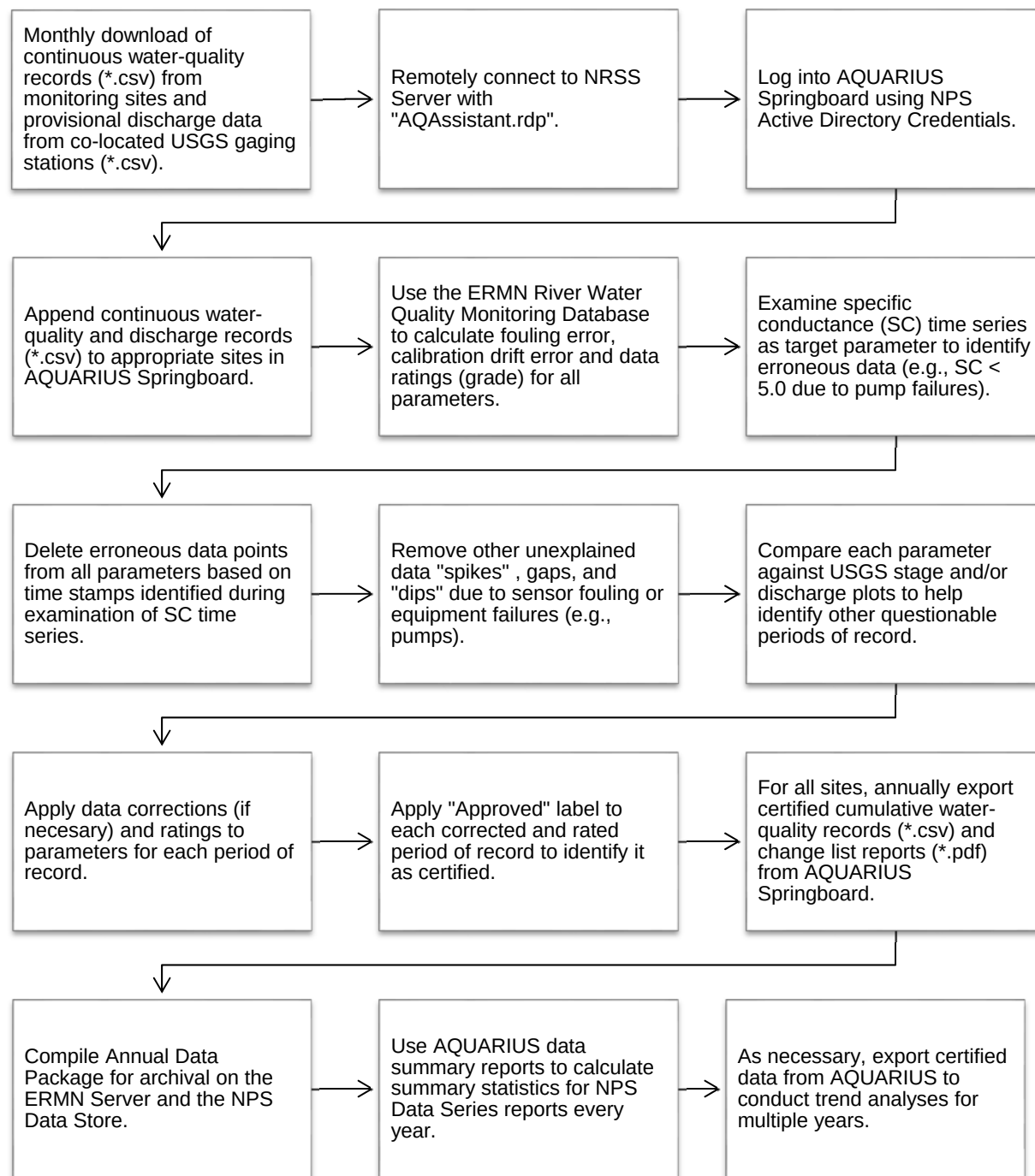


Figure 3. Generalized procedure for continuous water-quality record processing conducted by the Eastern Rivers and Mountains Network (ERMN) with AQUARIUS software (Aquatic Informatics, Vancouver, B.C., Canada). NPS = National Park Service. NRSS = Natural Resource Stewardship and Science; USGS = U.S. Geological Survey.

Literature Cited

- Aquatic Informatics. 2017. <http://aquaticinformatics.com/>. Vancouver, B. C. Canada. Accessed January 11, 2017.
- Shearer, J., C. Moore, K. K. Bartz, E. C. J. Booher, and J. Nelson. 2015. Monitoring freshwater systems in the Southwest Alaska Network: Standard operating procedures. Natural Resource Report NPS/SWAN/NRR—2015/925.1. National Park Service, Fort Collins, Colorado.

Appendix. End of Year Closeout Checklist

Water Year: _____ Aquatic Ecologist: _____ Date: _____

- ☐ Aquatic Ecologist, Compile Annual Data Package.
 - ☐ Protocol Folder
 - ☐ Archive current versions of protocol/QAP and SOPs.
 - ☐ Raw Data Folder
 - ☐ Scan all data sheets after verification.
 - ☐ Review scanned data sheets (pdfs) for completeness – are all pages included?
 - ☐ Archive and document location of all paper data sheets
_____.
 - ☐ Organize and archive raw water-quality records with appropriate naming convention.
 - ☐ QA/QC Data Folder
 - ☐ Organize and archive all QA/QC data products with appropriate naming conventions.
 - ☐ Monitor Maintenance Logs (*.pdf)
 - ☐ Cross-section Surveys (*.xlsx)
 - ☐ Scanned Data Sheets (*.pdf)
 - ☐ Database with all forms verified (*.accdb)
 - ☐ Cumulative Water Quality Record Change Lists (*.csv)
 - ☐ Certified Data Folder
 - ☐ Certify and export (*.csv) data for all sites from Aquarius.
 - ☐ Locations Folder
 - ☐ Confirm that monitoring locations have not changed. If any changes, have been made, update and republish shapefiles.
 - ☐ New shapefile name _____
- ☐ Aquatic Ecologist, deliver all folders (above) to Data Manager for archive.
- ☐ Aquatic Ecologist, Write annual summary in space below– Highlight: 1) site infrastructure changes, 2) any pertinent details for future data analysis, 3) identify any additional data that was collected during the field season (e.g., pilot or incidental data), 4) any other important details.

- ☐ Aquatic Ecologist, Schedule Operational Review, Date: _____
- ☐ Aquatic Ecologist, email (*.docx) this checklist to Data Manager
- ☐ Data Manager, Annual Data Distribution
 - ☐ Archive on NPS Data Store
 - ☐ Archive on ERMN Server