

Standard Operating Procedure 9: Data Management (Version 1.0)

Summary

This Standard Operating Procedure describes the conceptual framework and key components of ERMN Benthic Macroinvertebrate Monitoring data management and is modeled after the Northeast Temperate Network approach (Gawley et al. 2014).

Revision Log

Original Version #	Date of Revision	Revised By	Changes	Justification	New Version #
NA	4/28/2016	C. Tzilkowski	NA	Protocol Implementation Plan publication	1.0

Background and Scope

All data collected during ERMN Benthic Macroinvertebrate (BMI) Monitoring go through a series of steps (Figure 1) to ensure comparable data of consistently high quality. After data have been recorded in the field or laboratory on paper data sheets (i.e., data acquisition), the data sheets are scanned (.pdf format) as soon as possible and stored on the ERMN server. Subsequently, data are manually entered into the ERMN Benthic Macroinvertebrate Database from paper data sheets, typically by the Hydrologic Technician. Digital photographs and some quality assurance records (e.g., equipment maintenance logbooks) are not stored in the ERMN BMI Database but are archived on the ERMN server.

There are quality assurance and quality control processes, outlined in the Quality Assurance Plan (in development), that ensure that certified data exported from the ERMN BMI Database meet the highest possible standard and can be distributed to cooperators or included in analyses and reports. Data that have not been certified are not released to cooperators or the public.

The ERMN Server is the primary repository for certified ERMN BMI data whereas certified core water quality parameter data will be uploaded to EQuIS in addition to being stored within the ERMN BMI Database.

Database Design

The ERMN BMI Monitoring Database is a custom relational database developed in Microsoft Access that consists of tables, queries, data-entry forms, and built-in reporting tools. The database builds on the Natural Resource Database Template which is a set of core database tables and the standard for the NPS I&M Program.

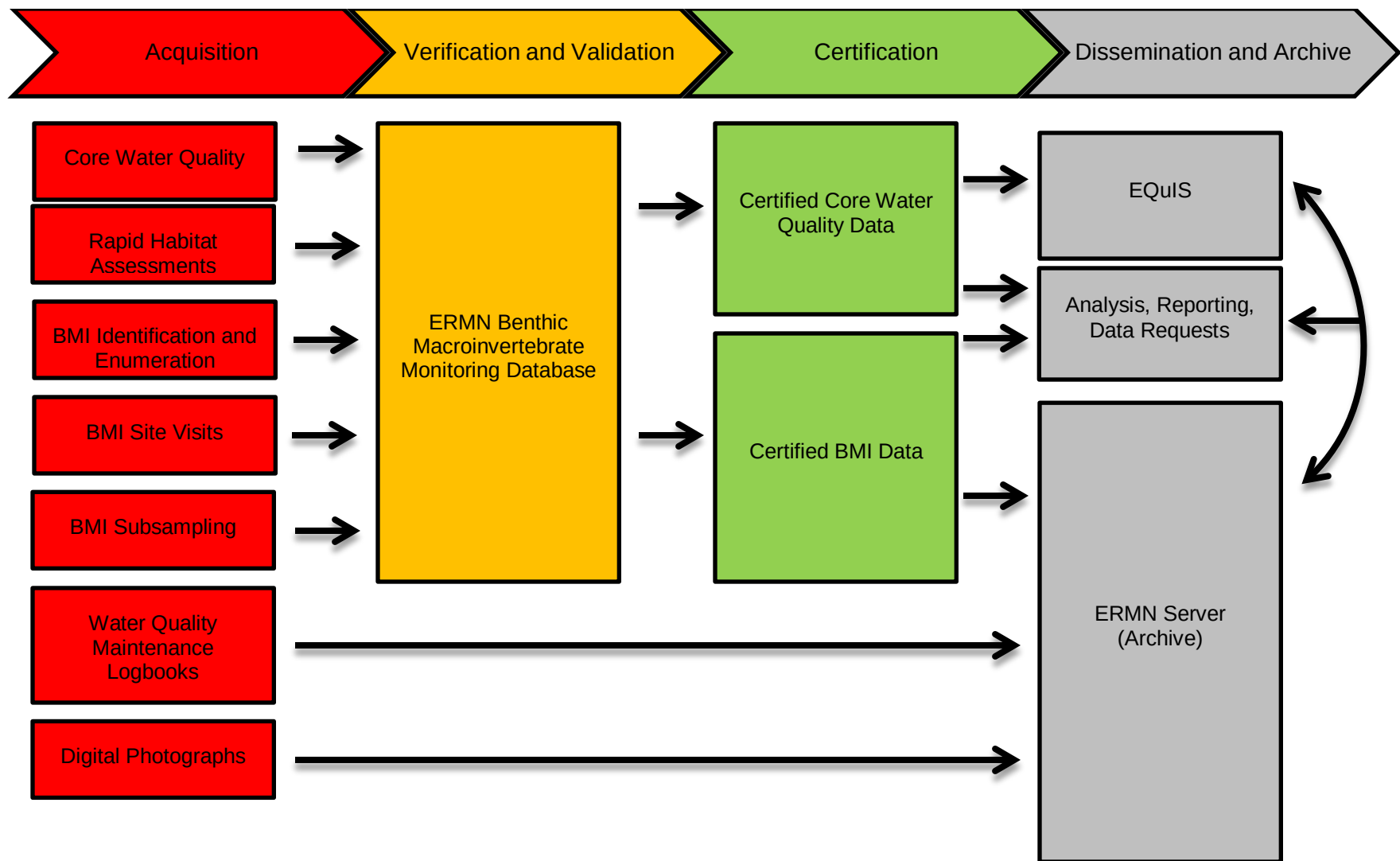


Figure 1. Data flow of the Eastern Rivers and Mountains Network (ERMN) benthic macroinvertebrate (BMI) monitoring program.

Data Entry

Data entry ideally takes place as soon as data collection is complete and is best performed by a person who is familiar with the data, oftentimes the Hydrologic Technician. Because field seasons can be particularly busy, data entry into the BMI Monitoring Database can be delayed if necessary, as long as hard copies have been scanned, backed-up and archived.

General procedures for data entry are as follows:

1. *Become familiar with data sheets.* Some errors or omissions are detectable at this point and can necessitate setting aside some forms for clarification or correction before entering potentially erroneous data into the database. The best stopping point is at the completion of the entry of a completed data sheet rather than in the middle of a logically single operation.
2. *Enter data.* Data are entered one logical set at a time – usually one complete set of field or laboratory sheets. Errors or questions about data can be recorded in the database as notes that can be useful during data verification. Interrupt data entry only at logical stopping points.
 - a. There are two “modes” in the ERMN BMI Monitoring Database – EDIT Mode and BROWSE Mode. Typically the Database is in BROWSE mode which minimizes accidental deletion or manipulation of previously entered data by locking the data entry fields.
 - b. To enter data, click on the “Enter new data” box on the main screen of the ERMN Benthic Macroinvertebrate Monitoring Database (Figure 2) which will open the “Data Viewing Form” (Figure 3).
 - c. For the data fields to be editable (i.e., entering new data) the database must be switched into EDIT mode by clicking on the box in the top right corner of the Data Viewing Form. The form background will change to a light green color to notify the user that the data entry fields are now editable.
 - d. There are three tabs in the Data Viewing Form – (1) Field Work, (2) Lab Work and (3) Vertebrates. Typically, data are entered in the same sequence of the tabs – vertebrate observations are usually recorded in the Field Work phase of data entry.
 - e. All sections of the “Field Work” data entry form should be populated whereas the number of entries in the “Lab Work” form can differ depending on the diversity of the BMI sample that is being entered. Each data entry form is patterned after the structure of the corresponding paper data sheet and has built-in quality assurance components such as pick lists and validation rules to test for missing data or illogical combinations.

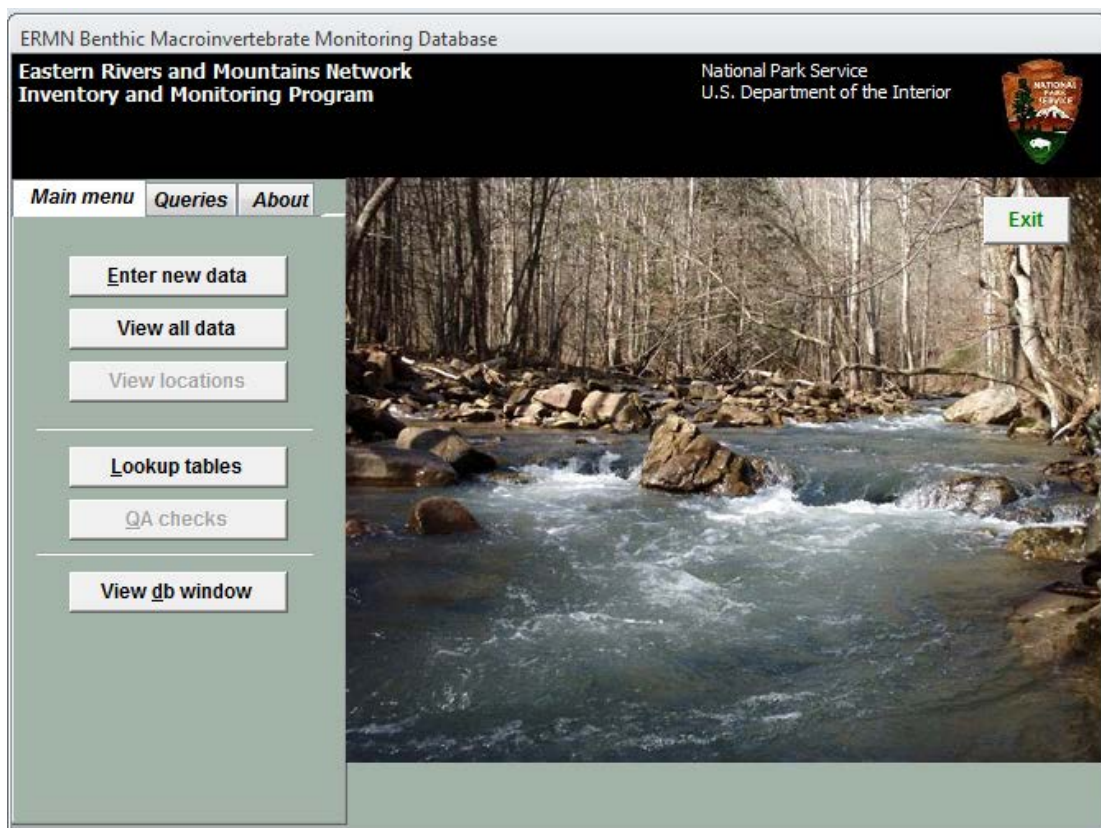


Figure 2. Main menu of the Eastern Rivers and Mountains Network Benthic Macroinvertebrate Monitoring Database.

Data Viewing Form - Filter by sampling event

Show All Records Find a Record: NERI.3036 Switch to EDIT Mode

Location: UNT Fall Branch 2 Start Date: 3/29/2014 Start Time: 9:00 Event Notes: Verified Date: 6/12/2015 Verified By: Weber Observers: Manning, Doug Weber, Andy Close

Field Work Lab Work Vertebrates

Related Sampling Activities: Check all that apply and/or fill in Other

☒ Water Chemistry ☐ Discharge ☒ Habitat Survey Other:

Physical Site Conditions

Cloud Cover (%) 100 Wind Light Precipitation Intensity Light

Riparian Shading Shaded Water Clarity Clear Water Color Clear

Beginning water measurements

Water temp (C) 5.52 Specific Cond (uS/cm) @ 25C 66.8 DO (%) 103.5 DO (mg/L) 11.87 pH 7.52

Physical site condition or other comments:

Bird midpoint is also the DS BMI point. Moved due to waterfall near DS end. Paced 100m and marked US end with GPS. Did not mark DS end since it is already on GPS. 2 unidentified larval sal. And one Desmognathus sp.? released from sample. Most (4) samples taken in lower 50m of reach, one taken near US end.

Sampling Information

Is Sample Split? No

Sample number	Water depth (m)	Substrate embeddedness (%)	Velocity (m/s)	Substrate Type Dominant	CoDominant
1	0.13	0	0.2	7	4
2	0.1	0	0.49	7	4
3	0.08	0	0.49	8	4
4	0.05	0	0.47	4	7
5	0.05	0	0.66	4	7

1. Epifaunal Substrate / Available Cover 18
2. Embeddedness 19
3. Velocity / Depth Regime 18
4. Sediment Deposition 18
5. Channel Flow Status 19
6. Channel Alteration 17
7. Frequency of Riffles (or bends) 19
8. Bank Stability Left 4 Right 8
9. Vegetative Protection Left 7 Right 8
10. Riparian Vegetative Zone Width Left 4 Right 9
Habitat Assessment Score: 168

Records: 1 of 1 Filtered Search

Data Viewing Form - Filter by sampling event

Show All Records Find a Record: DEWA.3007 Switch to EDIT Mode

Location: Spackmans Creek 8 Start Date: 4/12/2014 Start Time: 8:40 Event Notes: Verified Date: 6/12/2015 Verified By: Weber Observers: Manning, Doug Weber, Andy Close

Field Work Lab Work Vertebrates

Method 300 Correction Factor 7

Sort By Weber Sort Date 6/24/2014

Non-Chironomid Id By Tzilkowski ID Date Lab Sample Type: I&M

Taxon	Life Stage	Id Notes	Notes	Count	Large Rare	Corrected Count
Acronuria sp.	L			1		7
Baetidae	L			28		196
Cheumatopsyche sp.	L			21		147
Chironomidae	L			87		609
Dicranota sp.	L			2		14
Diplectrona sp.	L			33		231
Dolophlodes sp.	L			2		14
Ectopria sp.	L			1		7
Epeorus sp.	L			9		63
Ephemerella sp.	L			9		63
Ephemerellidae	L	imm.		9		63
Heptageniidae	L			8		56
Hydropsyche sp.	L			4		28
Leptophlebiidae	L	imm.		8		56
Leucrocuta sp.	L			14		98
Leuctridae	L	imm.		4		28
Maccaffertium sp.	L			10		70
Nemouridae	L			4		28
Neophylax sp.	L			8		56
Oligochaeta	NA			1		7
Oulimnius sp.	L			5		35
Paraleptophlebia sp.	L			3		21

Records: 1 of 29 No Filter Search

Records: 1 of 1 Filtered Search

Figure 3. "Field Work" form (top) and "Lab Work" form (bottom) for data entry into the Eastern Rivers and Mountains Network Benthic Macroinvertebrate Monitoring Database.

Verification

Data verification ("proofing") follows data entry and involves checking the accuracy of computerized records against the original paper data sheets. Although the goal of data entry is to achieve 100 percent correct entries, this is rarely accomplished. To minimize transcription errors, all records are verified to their original source during the first review (i.e., immediately after data entry). The records are reviewed a second time by a different person (typically the Aquatic Ecologist) and the results of that comparison reported with the data. If errors are found in the second review, then the entire data set is verified again.

Verification is typically done by one person but could be made more efficient if two people are available. To verify data:

Compare entered data to hard copies immediately after data entry (1st Verification). Three common types of error are: duplicated records (entered twice), missing records (inadvertently skipped during entry), and misspellings.

1. When an error in the database is found during the 1st Verification, immediately correct it in the database.
2. After verifying data from each field and laboratory sheet, the initial verifier (typically ERMN Hydrologic Technician) dates and initials the original data sheet, indicating that verification was done.
3. A secondary verifier (typically ERMN Aquatic Ecologist) repeats the verification at a later date and initials the data sheet.
4. Run simple summary statistics of the entered data. This is important because even when care is taken up to this point, a duplicate or omitted entry could have been overlooked. For example, the number of known constant elements, such as the number of sampling sites, samples per site, or sites per date can be viewed. Ask the same question in different ways to find differences in the answer that will provide clues to errors. The more variety of checks to test the completeness of the data, the greater is the confidence that the data are completely verified.

Data Validation and Quality Control

The ERMN Aquatic Ecologist validates data after verification is complete. Validation is the process of reviewing data for range and logic errors. Although data may be correctly transcribed from the original data sheets, the results may not be accurate or logical. For example, a stream pH of 25.0 or a temperature of 95°C is illogical and almost certainly incorrect, whether or not it was properly transcribed from data sheets.

Step-by-step instructions are not practical for data validation because each data set has unique measurement ranges, sampling precision, and accuracy requirements. Nonetheless, validation is an important step in the certification of the data. Certain components of data validation are built into data entry forms (such as range limits). Data can be compared to previous years to identify

gross differences. Validation procedures will identify generic errors such as missing, mismatched, or duplicate records.

One of the most important tasks of rigorous validation is when the checker returns to hard copies of data sheets to make corrections and notations about the errors that were found and fixed in digital files. Without annotating the original data sheets, the digital and paper records do not match. If this mismatch is discovered without adequate documentation, *all* data are rendered suspect.

The following generic procedures can be used to develop a validation strategy for most data sets. Examples of validation strategies (and strange errors) are also provided.

- *Catalog the error types found in each data set.* When particular validation errors are found, it is important to catalog them in an error log for that data set. Notes on the error(s) include a description, how they were detected, and how they were corrected.
- *Perform exploratory data analysis to look for outliers.* Database, graphic, and statistical tools can be used for ad-hoc queries and displays of the data. Histograms, line plots, and basic statistics reveal possible logic and range errors. Such exploratory methods identify obvious outliers.
- *Modify field data forms to avoid common mistakes.* With a list of validation errors and exploratory data results in hand, the field-data forms as the source of the logic errors can be reevaluated. Minor changes, small annotations, or adding check boxes to a data sheet remove ambiguity about what to enter on the form. In fact, any time the same type of validation error occurs repeatedly in different data sets, the data sheet – not the field crew – is usually at fault. Repeated validation errors can also mean that protocol(s) or field training is faulty, which must be recognized and corrected.

Metadata Procedures

Data documentation is a critical step toward ensuring that data sets are usable for their intended purposes well into the future. This involves the development of metadata, which can be defined as structured information about the content, quality, condition, and other characteristics of a data set. Additionally, metadata provide the means to catalog and search among data sets, thus making them available to a broad range of potential data users. The Aquatic Ecologist and Data Manager will be jointly responsible for completed metadata at the end of every calendar year.

Data Backup

As monitoring data are entered, verified, and validated, the ERMN BMI Monitoring database is backed up to the ERMN server, which is regularly backed up by the ERMN Data Manager. The paper field forms are clearly identified, organized, and filed in the ERMN office.

File Naming Conventions

Standard naming is essential for good data management – it provides the initial association of metadata (e.g. sampling location and date) with field data. Digital photographs, field data sheets, and laboratory data sheets are named and saved on the ERMN Server using standardized conventions (Table 1) in annual monitoring archive “folders” (e.g., P:\Monitoring\Benthic_Macroinvertebrates\2014_BMI_Archive\Data_Sheets).

There are typically four pages (two, 2-sided datasheets) of field data sheets per sampling event that are scanned and saved as one .pdf file; similarly, three pages of laboratory data per sampling event (one, 2-sided datasheet and a 1-sided datasheet) are combined and archived in .pdf format.

At a minimum, there are typically two digital photo points captured per sampling event (see SOP 4). Digital photographs are downloaded from cameras and renamed appropriately so that the file name contains the year, site, and photo point location (i.e., downstream or upstream end of sampling reach). Because permanent markers are inconspicuous, these photos are critical for identification of reach bounds in future years – these photos are included on future datasheets and field crews use them to identify where to begin and terminate sampling. If extraordinary observations (e.g., fossils, vandalism) are made while sampling, digital photos are named similarly but with an appended descriptor.

Table 1. Eastern Rivers and Mountains Network Benthic Macroinvertebrate Monitoring File Naming Conventions.

File Type	Naming Convention	Example
Field Data Sheets	Site Code + Site Name + 4-Digit Year + F	NERI.3021.FireCreek17.2014.F.pdf
Laboratory Data Sheets	Site Code + Site Name + 4-Digit Year + L	DEWA.3020.MillCreek12.2014.L.pdf
Digital Photographs	Site Code + Site Name + 4-Digit Year + “U” or “D”	BLUE.1004.JarrellBranch.2014.U.jpg

Data Certification and Archiving

Certification is a benchmark that indicates that data can be distributed or analyzed because: (1) data are complete for the period of record and (2) they have undergone and passed the quality assurance and quality control measures as described in this SOP and in the ERMN BMI Monitoring Quality Assurance Plan (in development). Certification does not imply that data are completely free of errors; however, it does mean that they are of consistently high quality and can be included in reports, presentations, or other publications. The Aquatic Ecologist is responsible for certifying data. After data are labeled as “certified” in the ERMN BMI Monitoring Database, they are either (1) archived in EQuIS (water quality data) or (2) in the appropriate annual archive on the ERMN server (all other BMI sampling data) as .csv files. Data are not distributed or used for analysis unless they are certified.

EQuIS

The National Park Service is transitioning to the database management system “EQuIS” for storage of discrete water quality data (Dean Tucker, NPS Water Resources Division, Fort Collins, CO, Person. Comm.) – ERMN will annually upload all certified water quality data associated with BMI monitoring to EQuIS. WRD is preparing an EQuIS Electronic Data Deliverable (EDD) specification that can be used by parks and networks to upload their data into EQuIS. Once data are uploaded into EQuIS they will automatically be archived in the EPA STORET Data Warehouse.

Data Dissemination

The majority of data analysis and reporting (SOP 10) will be conducted by the Aquatic Ecologist. All completed reports and articles will be publicly available from multiple sources including the ERMN website (<http://science.nature.nps.gov/im/units/ermn/>) and the NPS integrated resource management information portal (<https://irma.nps.gov/App/Portal/Home>).

All data requests should be made by contacting the ERMN Aquatic Ecologist.

Literature Cited

Gawley, W. G., B. R. Mitchell, and E. A. Arsenault. 2014. Northeast Temperate Network lakes, ponds, and streams monitoring protocol: 2014 revision. Natural Resource Report NPS/NETN/NRR—2014/770. National Park Service, Fort Collins, Colorado.