Best Practices for Submitting Benthic Macroinvertebrate Data to the Water Quality eXchange (WQX)

Introduction

The Water Quality eXchange (WQX) Benthic Macroinvertebrate Best Practices Guide was created to guide organizations in submitting benthic macroinvertebrate data to WQX. It was also created to make submitting benthic macroinvertebrate datasets more user-friendly, promote consistency when submitting data, and to maximize the extent to which data can be reused. The Best Practices Guide provides recommended approaches to addressing eight areas of concern with respect to documenting benthic macroinvertebrate data. These include recommendations on (1) assemblage sampled name, (2) biological intent, (3) habitat selection method, (4) collection effort, (5) sample collection equipment name, (6) net mesh size, (7) characteristic name, (8) target count, (9) proportion sample processed, (10) target taxonomic levels, (11) subject taxonomic name, (12) subject taxonomic name user supplied, and (13) uploading QA/QC documents. Each metadata element is important to ensure the reusability of the data for secondary data users and allows data users to make informed comparisons between disparate data sources.

This Guide is designed to be used with the 3.0 version of WQX. Some data elements mentioned in this Guide are new in WQX 3.0 and are not compatible with earlier versions of WQX, such as habitat selection method, target count, proportion sample processed, and subject taxonomic name user supplied. Some of these practices are required elements for benthic assemblages. We encourage data submitters to review this guide before transitioning to WQX 3.0 and to capture all these data elements going forward in future monitoring. WQX Web users are required to use WQX 3.0.

1. Assemblage Sampled Name

Assemblages are the group of organisms belonging to a number of different taxa that co-occur in the same area and interact through trophic and spatial relationships. In other words, these are the species occurring together in space and time. Benthic macroinvertebrates are an example of an assemblage as they are different taxa found in the same physical and trophic community. Assemblage sampled indicates what type of biological monitoring was conducted and is the element used to derive all benthic metadata requirements. *Assemblage sampled name* is a **required** field and is populated from a domain list of allowable values.

2. Biological Intent

Biological intent is the goal of the assessment such as to calculate species density or to conduct a population census. Knowing the purpose of the sampling helps data users understand data to expect given the goal of the sampling. *Biological intent* is a **required** field and is populated from a domain list of allowable values.

3. Habitat Selection Method

When sampling for benthic macroinvertebrates there are two common ways to select the habitat to sample; targeted and non-targeted. Each method is utilized to answer specific research questions. A targeted habitat selection method is a sampling approach that samples specifically in the most productive habitat such as riffles or root mases and will commonly yield the greatest quality and diversity at the site. These methods typically rely on the best professional judgement of the person sampling. Alternatively, non-targeted methods use systematic sampling approaches regardless of habitat, such as sampling regularly spaced transects. These methods typically rely on measurement or random elements, mostly free of human judgement. Oftentimes a non-targeted approach is more quantitatively representative of the entire site but may not capture all taxa. An exception is for sites composed of mostly productive habitat, which will yield similar results using either targeted or non-targeted methods. To facilitate the greatest reuse of the benthic data, *habitat selection method* is captured in a **conditionally required** field for benthic assemblages at the Activity level. Although this field is, if the habitat selection method is unknown for older data, a value of “Unknown” can be entered.

4. Collection Effort

Most macroinvertebrate sampling protocols constrain the amount of effort put into collecting a sample to conserve resources for additional monitoring and ensure samples from different sites are comparable in nature, this is known as collection effort. Protocols provide constraints on collection effort such as (1) time-constrained: where sampling only occurs for a certain amount of time, (2) area-constrained: where only a certain spatial extent is sampled and/or (3) constrained by staff and equipment used (gear procedure), such as the number of net sweeps or jabs. Including collection effort allows those reusing the data to better understand sample results can be expected given the amount of effort put in. For example, sampling 1 square meter will result in a smaller sample than sampling 3 square meters. In addition, area-constrained collection effort is needed in conjunction with the proportion sample processed to calculate total abundance for the area sampled. Collection effort is captured in **optional,** but **recommended**, data elements for collection duration, collection area, and gear procedure unit. These data elements should be populated whenever the information is known. These data elements are captured at the activity level, meaning they represent the collection effort for conducting the sampling, not the collection effort for the individual result. If an effort type is both time and area-constrained, information can be entered for both collection duration and collection area.

Collection Duration Measure and Units

Collection Duration Measure is the numeric value of the time-constrained effort type, for example, “30” could be entered to denote “30 Seconds”. *Collection Duration Units* is the data element with the units for the time-constrained effort type, such as “seconds”.

Collection Area Measure and Units

Collection Area Measure is the numeric value of the area-constrained effort type, for example, “3” could be entered to denote “3 Square Meters”. *Collection Area Units* is the data element with the units for the area-constrained effort type such as “m2”.

Gear Procedure Measure and Units

Gear Procedure Measure is the numeric value of the gear procedure unit, for example “3” could be entered to denote “3 net sweeps”. *Gear Procedure Unit Code* can be entered in the Gear Procedure Unit Code field as text from a domain list such as “# of net sweeps”.

5. Sample Collection Equipment Name

Macroinvertebrate sampling protocols should identify the equipment used for meeting the sampling objectives, such as a D-frame net. The equipment can impact what taxa or life stages are captured and give data users an idea of what taxa or lifecycles to expect in the results. *Sample Collection Equipment Name* is an **optional** but **recommended** field where data are available. It is populated from a domain list of allowable values.

6. Net Mesh Size

Similar to the sample collection equipment name, mesh size of the net used in sampling can vary, which has a direct effect on what taxa or life stages are captured. Therefore, mesh size for any nets used in sampling benthic macroinvertebrates should be documented in the *net mesh size measure* and *net mesh size unit* data elements. These elements are **optional** but highly **recommended** where the data are available.

7. Characteristic Name

Characteristic name is the name of the parameter being monitored. For benthic macroinvertebrate monitoring the characteristic name is frequently a count with the value of that count entered as a result and the subject taxonomic name indicating the species represented by the count. Another example of a characteristic is relative abundance. *Characteristic name* is a **required** field and is populated from a domain list of allowable values.

8. Target Count

During sorting, the sorter is typically aiming to count a certain number of individual macroinvertebrate specimens, this is called the target count. Once the sorter has met the target count, they will usually finish sorting the current grid cell, and the remainder of the grid cells will be left unidentified. Depending on the protocol used, the target is either the full sample or often a multiple of 100, such as 100, 200, 300, or 500 individuals. The target count is useful in determining data comparability for research objectives and whether a user wishes to randomly subset the data before using it. For example: if a data user is combining a target count of 300 with a target count of 500, they may wish to randomly subsample the higher count data to match the size of the lower count data set. Often the actual count of a sample will not be equivalent to the target count, either because there were not enough individuals to meet the target count (e.g. low-quality stream) or the sorter finished processing the final grid cell after the target count was met (making the actual count larger than the target). To derive an actual count, sum the values in the result value field, which contain a characteristic name of "count". Actual counts can be used to derive a population census whereas the target count is used to derive sampling method comparability. *Target count* is a **conditionally required** data element in WQX for benthic submissions. Although this field is required for benthic assemblages, if the Target Count is unknown for older data, a value of "Unknown" can be entered.

9. Proportion Sample Processed

Once a benthic macroinvertebrate sample is collected, the macroinvertebrates are sorted from the sample for identification. Prior to sorting, the sample is often placed onto a sampling grid where a sorter chooses a subset of the sample for identification by picking the benthic macroinvertebrates from random grid cells within the sample. The number of cells processed will depend on the protocol target count. For example, a sorter will pick specimens until their target count is met, thus leaving the remainder of the sample unidentified. This count needs to be accounted for in any calculations. This entry can be utilized to convert result to a standard calculation when aggregating data. The value is stored as a number between 0 and 1 where 1 represents the entire sample being processed. Some protocols also call for a large/rare count to be performed where just the large or rare benthic macroinvertebrates are picked across all the grids, this would also be documented as 1 in this field. It is important to note that the proportion of the sample processed does not consider any sieving or reducing large pieces of debris required to get the sample in the grid cells initially. *Proportion sample processed* is a **conditionally required** data element for benthic submissions. Although this field is required for benthic assemblages, it is understood for older data this metadata element may not be captured. Therefore, if the proportion of the sample processed is unknown, a value of 0.000000001 can be entered and it will be converted to Null. However, it is recommended this information be provided whenever possible Capturing and/or documenting this information along with your benthic results is strongly encouraged for future samples. Figure 1 shows a sampling grid where 4 of the 16 cells were sampled. This would be recorded at 0.25 indicating that a quarter of the sample was processed.

Figure 1 Grid Sampling Design

ø	✓	ø	ø
ø	ø	ø	✓
✓	ø	✓	ø
ø	ø	ø	ø

10. Target Taxonomic Levels

Organizations identify benthic macroinvertebrates to an established level of taxonomic resolution and life stage. To facilitate accurate reuse, secondary data users need to know whether a given life stage or taxon was not encountered, or whether it was deliberately excluded from results. For example, many organizations include larval gyrnid beetles but not adults whereas other organizations include the adults. Similarly, some organizations identify specimens down to the genus level while others only identify to the family level. To address this, it is recommended users provide the acceptable targeted levels of identification resolution, and which life stage and damaged or fragmented specimens are acceptable. These guidelines should be part of their standard operating procedures and specified in the *sample collection method description* within WQX. If target taxa differ based on project data quality objectives, create a new collection method documenting the different target taxa. It is always **required** that collection methods, analytical methods, and lab preparation methods be registered in WQX.

11. Subject Taxonomic Name

Subject taxonomic name is the name of the specimen identified from the sample. *Subject Taxonomic Name* is a **required** field and is populated from a domain list of allowable values for WQX.

12. Subject Taxonomic Name User Supplied

Organizations submitting to WQX sometimes have different taxonomic naming conventions than the WQX allowable values, such as bench names. For those users there is an **optional** *subject taxonomic name user supplied* data element. This is a text data element where the user can enter their preferred name. Providing the original taxa allows the data to be more representative of the actual taxa collected. It is also important to document the dichotomous key or other identification source used for the user supplied subject taxonomic name. This can be included in the **optional** *subject taxonomic name user supplied Reference Text* data element.

13. Uploading QA/QC Documents

A monitoring program should always have a Quality Assurance Program Plan (QAPP) or similar document. The QAPP outlines data and measurement quality objectives an organization's monitoring program must meet to ensure the data collected is of sufficient quantity and quality for their program objectives. Oftentimes, programs which collect benthic data will have metrics in their QAPP to quantify benthic data quality. Metrics may include percent taxonomic disagreement, which reflects the sample-based precision of taxonomic identifications and percent sorting efficiency, which quantifies sorter bias associated with finding macroinvertebrate specimens. Providing this information allows for secondary data users to screen not only for comparable protocols, but also for comparable data quality. Providing actual data and results of QC analyses is **optional** but encouraged to elevate the value of and confidence in the reported data, and to allow for their more informed reuse. For benthic macroinvertebrate data, QA/QC documentation can be uploaded in the following two fields at the Project or Activity level:

Document Attachment Type

This captures the type of file uploaded such as "pdf".

Document Attachment File Name

This is the name of the document and must exactly match the file name in the uploaded Zip file. The file name should also make it clear it is a sample and data QC analysis document. All files must be uploaded as Zip files.

Documenting Benthic Macroinvertebrate Results

Benthic Macroinvertebrate data are collected and shared to WQX by many organizations throughout the country. By clarifying ambiguous terms, providing user guidance, and improving WQX submission requirements, the WQX benthic macroinvertebrate data can be captured at sufficient detail to facilitate secondary data usage. A complete biological result submission includes the fields in Table 1:

Table 1 includes a complete biological result with required elements highlighted in green:

Table 1 Complete Biological Results Submission

Field Name	**Example Data**
Monitoring Location Name	Big Creek
Monitoring Location Type Name	River/Stream
Horizontal Collection Method Name	NAD83
Project Identifier	03-500
Project Name	Rotating Basin Monitoring Program
Activity Identifier	ML-06:20200301:1355:SR
Activity Media Name	Biological
Activity Start Date	3/1/20
Activity Type Code	Sample-Routine
Assemblage Sampled Name	Benthic Macroinvertebrates
Proportion of Sample Processed Numeric	0.25
Target Count	300
Characteristic Name	Count
Result Value	350
Subject Taxonomic Name	Ephemerella catawba
Activity Type	Sample-Routine
Activity Media Name	Biological
Biological Intent	Population Census
Habitat Selection Method	Targeted
Collection Duration Measure	30
Collection Duration Unit	seconds
Collection Area Measure	5
Collection Area Unit	meters
Gear Procedure Unit Measure	3
Gear Procedure Unit Code	# of net sweeps
Result Sampling Point Type	Line Transect
Result Sampling Point Place In Series Numeric	5
Subject Taxonomic Name User Supplied	Mayfly
Subject Taxonomic Name User Supplied Reference Text	NCBI
Net Mesh Size Measure	500
Net Mesh Size Unit	micron
Reach Length Measure	30
Reach Length Unit	feet
Reach Width Measure	10
Reach Width Unit	feet

Field Name	Example Data
Sample Collection Method ID	GRE:BEN-Kick
Sample Collection Equipment Name	D-Frame Net
Biological Intent	Frequency Class
Frequency Class Descriptor	Larva
Frequency Class Descriptor Unit	count

Any questions about this best practices guide, along with any technical questions can be directed to the WQX Helpdesk at wqx@epa.gov.