NEON USER GUIDE TO AQUATIC MACROINVERTEBRATE COLLECTION (DP1.20120.001)

PREPARED BY	**ORGANIZATION**
Stephanie Parker	AQU
Tanya Chesney	DPS
Caren Scott	AQU

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	08/16/2017	Initial Release
B	10/15/2020	Included general statement about usage of neonUtilities R package and statement about possible location changes. Updated taxonomy information. Updated littoral sampling locations and figures. Added information about standard taxonomic effort and information about invertebrate bycatch in the fish data product.
C	04/08/2022	Added language in section 4 Taxonomy addressing RTE species obfuscation in the data. Updated section 5.3 Data Revision with latest information regarding data release
C.1	03/30/2023	Updated information about the inv_identificationHistory table
C.2	05/12/2023	Description of taxonomic identification target taxon ranks and taxonomic information in inv_taxonomyProcessed and inv_taxonomyRaw remarks fields.
C.3	08/15/2023	Removed reference to beetle sample tables.
D	04/09/2024	Minor formatting updates

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

1 DESCRIPTION

1.1 Purpose

This document provides an overview of the data included in this NEON Level 1 data product, the quality controlled product generated from raw Level 0 data, and associated metadata. In the NEON data products framework, the raw data collected in the field, for example,the macroinvertebrate samples collected in the field are considered the lowest level (Level 0). Raw data that have been quality checked via the steps detailed herein, as well as simple metrics that emerge from the raw data are considered Level 1 data products.

The text herein provides a discussion of measurement theory and implementation, data product provenance, quality assurance and control methods used, and approximations and/or assumptions made during L1 data creation.

1.2 Scope

This document describes the steps needed to generate the L1 data product Macroinvertebrate collection and associated metadata from input data. This document also provides details relevant to the publication of the data products via the NEON data portal, with additional detail available in the file, NEON Data Variables for Aquatic Macroinvertebrate Collection (DP1.20120.001) (AD[04]), provided in the download package for this data product.

This document describes the process for ingesting and performing automated quality assurance and control procedures on the data collected in the field pertaining to AOS Protocol and Procedure: Aquatic Macroinvertebrate Sampling (AD[06]). The raw data that are processed in this document are detailed in the file, NEON Raw Data Validation for Aquatic Macroinvertebrate Collection (DP0.20120.001) (AD[03]), provided in the download package for this data product. Please note that raw data products (denoted by 'DP0') may not always have the same numbers (e.g., '20120') as the corresponding L1 data product.

2 RELATED DOCUMENTS

2.1 Associated Documents

AD[01]	NEON.DOC.000001	NEON Observatory Design (NOD) Requirements
AD[02]	NEON.DOC.002652	NEON Products Catalog
AD[03]	Available with data download	Validation csv
AD[04]	Available with data download	Variables csv
AD[05]	NEON.DOC.001152	NEON Aquatic Sampling Strategy
AD[06]	NEON.DOC.003046	AOS Protocol and Procedure: Aquatic Macroinvertebrate Sampling
AD[07]	NEON.DOC.000008	NEON Acronym List
AD[08]	NEON.DOC.000243	NEON Glossary of Terms
AD[09]	NEON.DOC.004825	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[10]	Available on NEON data portal	NEON Ingest Conversion Language Function Library
AD[11]	Available on NEON data portal	NEON Ingest Conversion Language
AD[12]	Available with data download	Categorical Codes csv

3 DATA PRODUCT DESCRIPTION

Aquatic macroinvertebrate-related data products include taxonomy, abundance and density, and morphometrics (which can be used along with literature length/mass regressions to determine biomass), and give information related to the NEON Grand Challenge area of Biodiversity as well as additional data about the macroinvertebrate community in streams, lakes, and rivers. These data can be used to assess the health of aquatic ecosystems. Macroinvertebrates are sampled three times per year at each NEON aquatic site (AD[05]). Sampling dates are based on a combination of variables, including hydrology in streams or ice on/ice off dates in lakes, accumulated degree days (temperature), and riparian greenness (phenology). Samples are collected by field personnel, preserved in the field, and sent to expert taxonomists for identification. For additional information see sampling design NEON Aquatic Sampling Strategy (AD[05]) and protocol AOS Protocol and Procedure: Aquatic Macroinvertebrate Sampling (AD[06]).

Data are organized into tables for field data collected by NEON technicians and external lab data returned by the expert taxonomy lab(s). Field data contains metadata on sample time, location, type of habitat and substratum, and the type of sampler used, which determines the benthic area sampled. The lab data includes subsampling information, taxonomic analysis, count and size class data, and quality metrics. Lab data are corrected for subsampling, however the data user must use both the lab data and the field data to calculate counts per benthic area of habitat if a quantitative result is desired. See Section below for suggested calculations.

3.1 Spatial Sampling Design

Benthic invertebrates at NEON aquatic sites (Figure 1) are sampled using a percent-based macrohabitat approach (after Moulton et al. 2002). Habitats sampled focus on riffles, runs, pools, and step pools depending on the percent cover of each habitat within each 1 km-long NEON Aquatic wadeable stream site (NOTE: some NEON sites may be less than 1 km due to permitting restrictions), and benthic-pelagic and littoral samples in lakes and rivers. Five samples are collected in the dominant habitat type (wadeable stream) or littoral area (lake and river), and three samples are collected in the second-most dominant habitat type (wadeable stream) or pelagic area (lakes and river) for a total of eight samples on a given sampling date at a site.

Samplers used for macroinvertebrate collection are designed to work by disturbing the benthic sediments and catching invertebrates in an attached net or container, while delineating the benthic area sampled for a quantitative result. The sampler type chosen differs depending on the water depth, velocity, and substratum type in the chosen habitat (Hauer and Resh 2006). The collection method may differ depending on the habitat and substrate being sampled, however all samples are collected from the surface of the natural substratum in each habitat using a quantitative sampling method. See AOS Protocol and Procedure: Aquatic Macroinvertebrate Sampling (AD[06]) for additional details on sampling strategy and SOPs.

As much as possible, sampling occurs in the same locations over the lifetime of the Observatory. However, over time some sampling locations may become impossible to sample, due to disturbance or other local changes. When this occurs, the location and its location ID are retired. A location may also shift to slightly different coordinates. Refer to the locations endpoint of the NEON API for details about locations that have been moved or retired: https://data.neonscience.org/data-api/endpoints/locations/

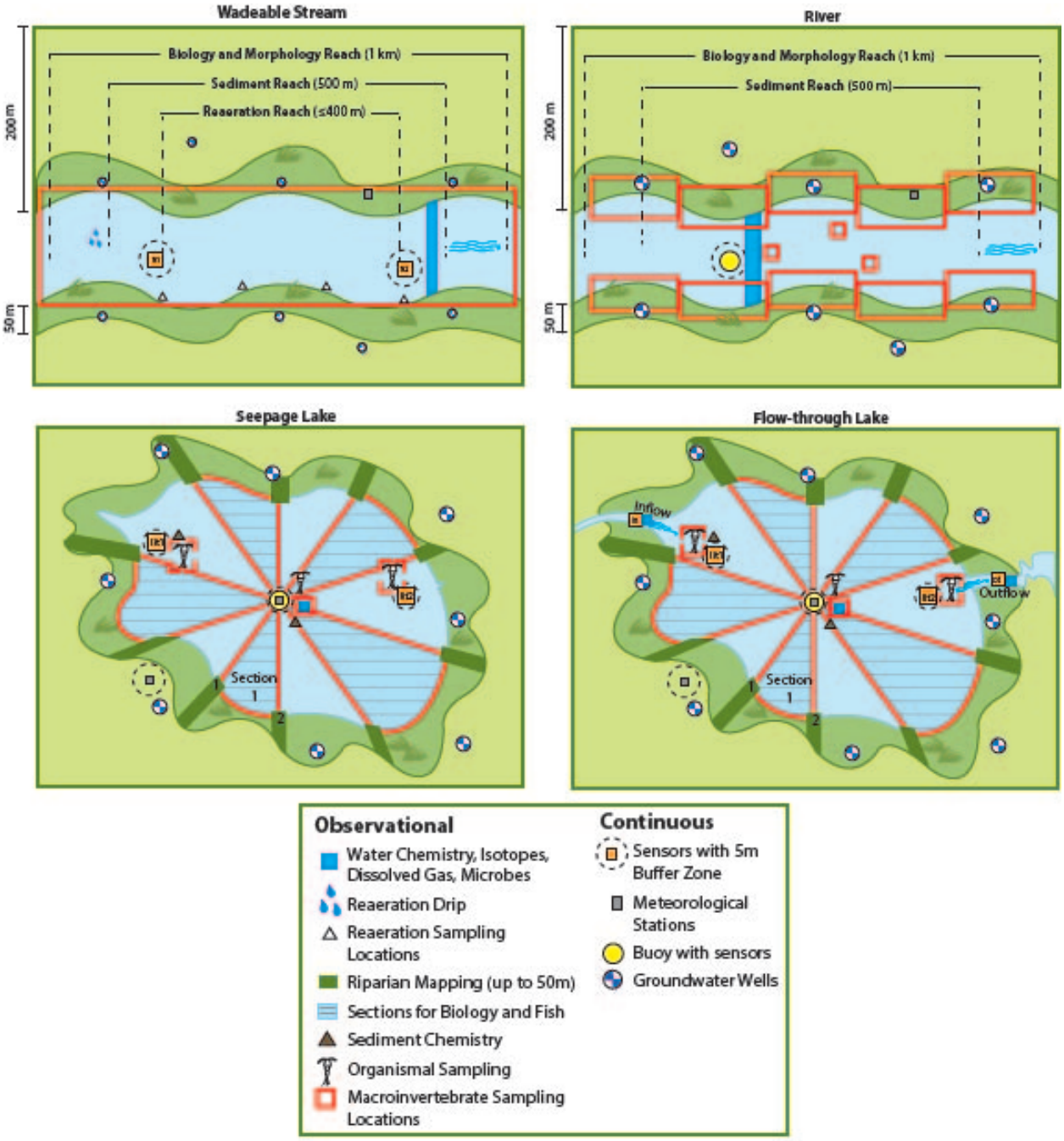

Figure 1: Generic aquatic site layouts (wadeable streams, rivers, and lakes) with macroinvertebrate sampling locations in red.

3.2 Temporal Sampling Design

Macroinvertebrate sampling occurs three times per year. Timing of sampling is site-specific and determined based on historical hydrological and meteorological data. Sample bout 1 is an early-season date, representing a period of rapid biomass accumulation after winter, typically prior to leaf out or ice-off where applicable. Sample bout 2 targets mid-summer baseflow conditions and sample bout 3 represents the late growing season (typically autumn) during leaf-fall where applicable. These dates differ on a site-by-site basis, but should always occur at, or near, baseflow conditions within the watershed. Sampling does not occur directly following a flood in wadeable streams (defined as >1.5 x base flow; Biggs et al. 1999). Should such a flood event occur on or prior to a target collection date, sampling is delayed 3 days-1 week (maximum 2 weeks, dependent on field schedule) to allow for invertebrates to recolonize the substratum (c.f. Brooks and Boulton 1991, Matthaei et al. 1996). Data collection for this data product occurs within one day per bout at a given site. See NEON Aquatic Sampling Strategy (AD[05]), AOS Protocol and Procedure: Aquatic Macroinvertebrate Sampling (AD[06]) for additional details.

3.3 Sampling Design Changes

Location names for the nearshore sensors in seepage lakes (lakes without a true inlet and outlet stream) were changed on January 1, 2021. The location previously known “inlet” changed to “littoral 1” (“lit2”) and “outlet” changed to “littoral 2” (“lit2”) to indicate that these locations are not near an inlet or outlet stream (Figure 1). Flow-through lakes (e.g., D18 TOOK) have sensors in the inflow and outflow streams, as well as the lit1 and lit2 locations in the lake.

3.4 Variables Reported

All variables reported from the field or laboratory technician (L0 data) are listed in the file, NEON Raw Data Validation for Aquatic Macroinvertebrate Collection (DP0.20120.001) (AD[03]). All variables reported in the published data (L1 data) are also provided separately in the file, NEON Data Variables for Aquatic Macroinvertebrate Collection (DP1.20120.001) (AD[04]).

Field names have been standardized with Darwin Core terms (http://rs.tdwg.org/dwc/; accessed 4 August 2017), the Global Biodiversity Information Facility vocabularies (http://rs.gbif.org/vocabulary/gbif/; accessed 16 February 2014), the VegCore data dictionary (https://projects.nceas.ucsb.edu/nceas/projects/bien/wiki/VegCore; accessed 16 February 2014), where applicable. NEON Aquatic Observation System (AOS) spatial data employs the World Geodetic System 1984 (WGS84) for its fundamental reference datum and Earth Gravitational Model 96 (EGM96) for its reference gravitational ellipsoid. Latitudes and longitudes are denoted in decimal notation to six decimal places, with longitudes indicated as negative west of the Greenwich meridian.

Some variables described in this document may be for NEON internal use only and will not appear in downloaded data.

3.5 Temporal Resolution and Extent

The finest temporal resolution that macroinvertebrate data will be tracked is per sampling day. All 8 samples are collected within a single day at a particular site. A suite of other biological sampling occurs at the

site during the same ~30 day bout. Three sampling bouts occur per site per year. The finest resolution at which temporal data are reported is at **collectDate**, the date and time of day when the samples were collected in the field.

The NEON Data Portal provides data in monthly files for query and download efficiency. Queries including any part of a month will return data from the entire month. Code to stack files across months is available here: https://github.com/NEONScience/NEON-utilities

3.6 Spatial Resolution and Extent

Each macroinvertebrate sample represents a patch of stream bottom within the 1 km permitted wadeable stream or river reach, or permitted lake area, and contains multiple individuals. The exact location (latitude and longitude) of each sample is not tracked as it is intended to represent the overall habitat. The **locationID** reported in wadeable streams represents a midpoint in the permitted reach, plus coordinate uncertainty surrounding that point. In lakes and rivers, some samples are collected near monumented locations associated with a more-specific **locationID**. Sampling locations are tracked by latitude and longitude and include an indication of **coordinateUncertainty**.

Up to two different habitats are sampled at each site to account for the variability or patchiness among habitats. Overall,this results in a spatial hierarchy of:

locationID (finest spatial resolution, ID of location within site) -> siteID (ID of NEON site) -> domainID (ID of a NEON domain)

Note that some **sampleID**s in the legacyData may start with “ST” (i.e., STMA = MAYF, STWA = WALK, STCU = CUPE, STKG = KING). These samples were collected in the STREON reach (the downstream 500 m of the 1 km reach) prior to descoping of the STREON experiment. Data from these samples may be used just as any other sample from the site.

3.7 Associated Data Streams

A subset of the macroinvertebrate field collection data are related to Macroinvertebrate metabarcoding (DP1.20126.001) samples collected at the same time and location, related samples share the same **parentSampleID**.

Macroinvertebrate collection data are also loosely related to Aquatic General Field Metadata collected on the same sampling day (NEON.DOC.001646). Data for Aquatic General Field Metadata are available in the NEON data product “Gauge Height” (DP1.20267.001). These data products are linked through the **siteID** field and local date in the NEON Data Publication Workbook for AOS Macroinvertebrate Collection (AD[04]).

Invertebrate data are also recorded in the fsh_invertBycatch table in the Fish electrofishing, gill netting, and fyke netting counts (DP1.20107.001) data product for select NEON domains where Decapods are a large part of the aquatic community.

3.8 Product Instances

At each aquatic site, there will be up to 24 samples collected per year (8 samples per bout). Each sample generates multiple records from the external lab on a per taxon, per size class basis.

3.9 Data Relationships

For each record collected in inv_fieldData a number of child records may be created. In the event that sampling is impractical (e.g., the location is dry, ice covered, etc.), there will be no child records. If a **sampleID** is recorded in inv_fieldData, there will be one corresponding record in inv_perSample (sample sorting and subsampling at the external lab). Each **sampleID** record in inv_fieldData may have multiple child records in inv_taxonomyRaw and inv_taxonomyProcessed, one record for each **scientificName** and **sizeClass** combination. A record from inv_fieldData may have multiple or no records in inv_perVial, as that table represents individuals removed from the final archived sample and placed in the external lab's in-house reference collection, records in this table are opportunistic and are organized by **sampleID** and **scientificName**. Duplicates and/or missing data may exist where protocol and/or data entry aberrations have occurred; users should check data carefully for anomalies before joining tables.

inv_fieldData.csv - > One record is created for each sample collected in the field, creating a **sampleID** which is linked to all subsequent tables. This table also indicates the field conditions, including **habitatType**, **samplerType**, **substratumSizeClass**, and sample depth if applicable (e.g., lake and river sites). For the macroinvertebrate collection data, **sampleID**s ending in '.DNA' can be filtered out from the table, they are used in the Macroinvertebrate Metabarcode Data product.

inv_perSample.csv - > One record (**sampleID**) is created for each sample processed at the expert taxonomy lab. Each sample is sorted (removing macroinvertebrates from organic and inorganic material) and may be subsampled if the number of macroinvertebrates in the sample appears to exceed 300. Data from this table are linked to the inv_fieldData and subsequent inv_taxonomyProcessed and inv_taxonomyRaw data through the **sampleID**s in each table.

inv_taxonomyRaw.csv - > One record is created for each taxonomic group identified in a sample created in inv_fieldData. Taxonomic identifications are made to the lowest practical taxonomic level (typically genus or species). The taxonomic nomenclature in this file reflects the verbatim identifications provided by the external taxonomist and may contain synonyms. Data are linked to the fieldData and perSample tables through the **sampleID**s in each table. Records in this table are unique by the combination of **sampleID**, **scientificName**, **morphospeciesID**, **sizeClass**, **immatureSpecimen**, **indeterminateSpecies**, and **identificationQualifier**. Records may include a **slideID** used to identify permanent slides used to facilitate the identification of difficult taxa, such as Chrionomids or Oligochates. Permanent slides will be archived using the **slideID**.

inv_taxonomyProcessed.csv - > One record is created for each taxonomic group identified in a sample created in inv_fieldData. Taxonomic identifications are made to the lowest practical taxonomic level (typically genus or species). The taxonomic nomenclature in this file has been standardized and desynonymized according to NEON's master taxonomy for macroinvertebrates and zooplankton. Data are linked to the fieldData and perSample tables through the **sampleID**s in each table. Records in this table are unique by the combination of **sampleID**, **scientificName**, **morphospeciesID**, **sizeClass**, **immatureSpecimen**, **indeterminateSpecies**, and **identificationQualifier**. Records may include a **slideID** used to identify

permanent slides used to facilitate the identification of difficult taxa, such as Chrionomids or Oligochates. Permanent slides will be archived using the **slideID**.

inv_perVial.csv - > One record is created for each taxonomic group removed from the final archive vial(s) from the subsamples created in inv_fieldData and inv_perSample. Individuals are removed from the archived sample to be kept at the expert taxonomy lab as part of the reference collection. Data are linked to the fieldData, perSample, and taxonomy tables through the **sampleID**s in each table. Records in this table are unique by the combination of **sampleID** and **scientificName**. Individual organisms documented in inv_taxonomyRaw are returned to a single vial per **sampleID** for archiving. Any individuals removed from that vial to be used by the external lab for the reference collection are documented in the inv_perVial table. The reference collection is housed at the external facility for the life of the contract, and is organized by **domainID** and **scientificName**. The **referenceCount** field indicates the number of organisms that have been removed from the inv_taxonomyRaw vial to be archived. Records may include a **referenceID** used to identify a vial added to the NEON reference collection for a given taxon. Vials labeled with a **referenceID** will be archived.

inv_identificationHistory.csv - > One or more records expected per identificationHistoryID. Records are only created when data corrections to taxonomic identifications are made. If errors in identification are detected through QAQC processes after data publication, then corrected taxonomy will be provided in the inv_taxonomyRaw and inv_taxonomyProcessed tables. The inv_identificationHistory table is populated with all prior names used for specimen(s) in the data product. When data are populated in the inv_identificationHistory table, **identificationHistoryID** is used as a linking variable between the inv_identificationHistory table and all other tables where updates were made.

Data downloaded from the NEON Data Portal are provided in separate data files for each site and month requested. The neonUtilities R package contains functions to merge these files across sites and months into a single file for each table described above. The neonUtilities package is available from the Comprehensive R Archive Network (CRAN; https://cran.r-project.org/web/packages/neonUtilities/index.html) and can be installed using the install.packages() function in R. For instructions on using neonUtilities to merge NEON data files, see the Download and Explore NEON Data tutorial on the NEON website: https://www.neonscience.org/download-explore-neon-data

3.10 Special Considerations

The macroinvertebrate taxonomic counts are reported per 1 mm size class. In some cases, the taxonomy lab reported counts from the same size class twice, and numbers should be summed. This is indicated in the field **sizeCategory**, where the **sizeClass** is appended with a, b, c, etc. Depending on the use case, data users may want to sum all *sizeClass **and** sizeCategory **per** sampleID **+** scientificName** prior to data analysis.

Count data can be found in the field **estimatedTotalCount**, which is corrected for subsampling at the external lab, but is NOT corrected for benthic area. Data users will need to refer to the **benthicArea** presented in the inv_fieldData table and apply this correction to get the number of organisms per square meter of stream, lake, or river bottom. All taxon records from a sample should be summed and divided by the **benthicArea** prior to reporting the total abundance per m^2.

$$macroinvertebrateAbundancePerM_i^2 = \frac{\sum_{i=1}^{n} inv\_taxonomyProcessed.estimatedTotalCount_i}{inv\_fieldData.benthicArea_i} \quad (1)$$

Where 'i' is a unique **sampleID**

See the external lab SOP (referenced in inv_perSample) for calculations applied to the data by the external laboratory.

Data users interested in calculating macroinvertebrate biomass should use length/mass relationships in published literature (e.g., Benke et al. 1999), along with the **scientificName** and **sizeClass** from the inv_taxonomyProcessed table and abundance estimates calculated in Equation 1 to calculate biomass per square meter. Biomass estimates are not provided in the published NEON data.

4 TAXONOMY

NEON manages taxonomic entries by maintaining a master taxonomy list based on the community standard, if one exists. Through the master taxonomy list, synonyms submitted in the data are converted to the appropriate name in use by the standard. The master taxonomy for macroinvertebrates and zooplankton was originally based on comprehensive taxonomy lists provided by expert taxonomy labs (EcoAnalysts, Inc. and GEI Consultants, Inc.) that were cross-referenced with taxonomic concepts from the Integrated Taxonomic Information System (ITIS, itis.gov) or Catalogue of Life (www.catalogueoflife.org) databases. Unique Taxon ID codes used to identify taxonomic concepts in the NEON master taxonomy list were generated for each taxon by concatenating the first three letters of the genus name together with the first three letters of the specific epithet to make a unique taxon ID for each scientific name. The list includes a variety of macroinvertebrate taxa, including mollusks, snails, worms, insects, mites, and crustaceans. NEON plans to keep the taxonomy updated in accordance with Merritt et al. (2019) and other current literature starting in 2020 and annually thereafter.

The master taxonomy list also indicates the expected geographic distribution for each species by NEON domain and whether it is known to be introduced or native in that part of the range. Given that the spatial distributions of many aquatic macroinvertebrate taxa are not well known, NEON assumes that all taxa are possible at all aquatic sites. As spatial resolution of distribution maps improves, NEON will update the taxon tables to generate errors if a species is reported at a location outside of its known range.

Prior to the 2022 data release, publication of species identifications were obfuscated to a higher taxonomic rank when the taxon was found to be listed as threatened, endangered, or sensitive at the state level where the observation was recorded. The state-level obfuscation routine was removed from the data publication process at all aquatic locations excluding sites located in D01, and data have been reprocessed to remove the obfuscation of state-listed taxa for all years. Federally listed threatened and endangered or sensitive species remain obfuscated at all sites and sensitive species remain redacted at National Park sites.

The full master taxonomy lists are available on the NEON Data Portal for browsing and download: http://data.neonscience.org/static/taxon.html.

4.1 Identification History

Beginning in 2023, the inv_identificationHistory table was added to track any changes to taxonomic identifications that have been published in NEON data. Such taxonomic revisions may be necessary when errors are found in QAQC checks, or when evidence from genetic analysis of samples or re-analysis of archived samples indicate a revision is necessary. Requests for taxonomic changes are reviewed by NEON science staff. Proposed changes are evaluated based on evidence in the form of photographs, existing samples, genetic data, consultation with taxonomic experts, or range maps. Upon approval, the existing record in the inv_taxonomyRaw and inv_taxonomyProcessed tables are updated with the new taxonomic information and a unique identifier is added to the **identificationHistoryID** field. A record with the same identificationHistoryID is created in the inv_identificationHistory table where the previous taxonomic information is archived along with the date the change was made.

4.2 Macroinvertebrate Taxon Rank and Targets

Taxonomic information from the expert taxonomists is recorded in the **scientificName** field of the inv_taxonomyProcessed and inv_taxonomyRaw tables. Data in this field are delivered at the taxon rank specified in the Standard Taxonomic Effort table (see Section 5.5). Some identification information may also be recorded in **identificationRemarks** for records where the taxonomist is 1) able to positively identify the specimen to a lower taxon rank than the target, or 2) provide plausible identifications for a higher level taxon rank when the taxon is not clear (e.g., family level recorded in **scientificName** and two potential genera provided in **identificationRemarks**). Other fields that indicate the reasons why specimens may not be resolved to the expected target taxon rank are **immatureSpecimen**, **indeterminateSpecies**, **distinctTaxon**, **specimenQualifier**, and **sampleCondition**.

5 DATA QUALITY

5.1 Data Entry Constraint and Validation

Many quality control measures are implemented at the point of data entry within a mobile data entry application or web user interface (UI). For example, data formats are constrained and data values controlled through the provision of dropdown options, which reduces the number of processing steps necessary to prepare the raw data for publication. The field data entry workflow for collecting macroinvertebrate field data is diagrammed in Figure 2.

An additional set of constraints are implemented during the process of ingest into the NEON database. The product-specific data constraint and validation requirements built into data entry applications and database ingest are described in the document NEON Raw Data Validation for Aquatic Macroinvertebrate Collection (DP0.20120.001) (AD[03]), provided with every download of this data product. Contained within this file is a field named 'entryValidationRulesForm', which describes syntactically the validation rules for each field built into the data entry application. Data entry constraints are described in Nicl syntax in the validation file provided with every data download, and the Nicl language is described in NEON's Ingest Conversion Language (NICL) specifications ([AD[10]).

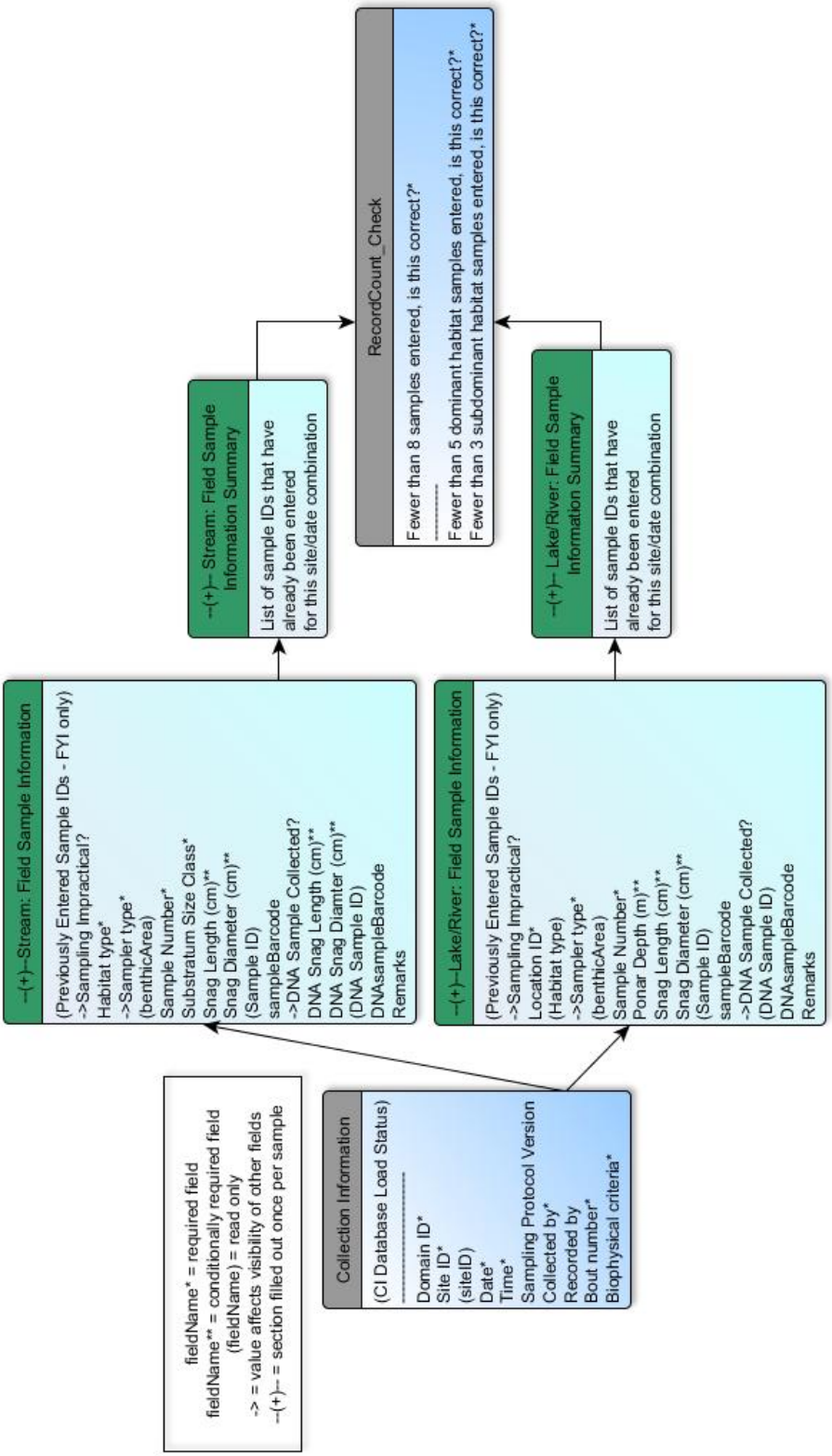

Figure 2: Schematic of the applications used by field technicians to enter macroinvertebrate field data

5.2 Automated Data Processing Steps

Following data entry into a mobile application of web user interface, the steps used to process the data through to publication on the NEON Data Portal are detailed in the NEON Algorithm Theoretical Basis Document: OS Generic Transitions (AD[09]).

5.3 Data Revision

All data are provisional until a numbered version is released. Annually, NEON releases a static version of all or almost all data products, annotated with digital object identifiers (DOIs). The first data Release was made in 2021. During the provisional period, QA/QC is an active process, as opposed to a discrete activity performed once, and records are updated on a rolling basis as a result of scheduled tests or feedback from data users. The Issue Log section of the data product landing page contains a history of major known errors and revisions.

5.4 Quality Flagging

The **dataQF** field in each data record is a quality flag for known errors applying to the record. Please see the table below for an explanation of **dataQF** codes specific to this product.

Table 1: Descriptions of the dataQF codes for quality flagging

fieldName	value	definition
dataQF	legacyData	Data recorded using a paper-based workflow that did not implement the full suite of quality control features associated with the interactive digital workflow

Records of land management activities, disturbances, and other incidents of ecological note that may have a potential impact are found in the Site Management and Event Reporting data product (DP1.10111.001)

5.5 Analytical Facility Data Quality

Data analyses conducted on macroinvertebrate community data conform to the current data quality standards used by practitioners. Ten percent of all samples are quality checked for taxonomic difference between two taxonomists at the external facility. These records are indicated by the fields **qcChecked**, **qcEnumerationDifference**, and **qcTaxonomicDifference** indicating Percent Difference in Enumeration (PDE) and Percent Taxonomic Difference (PTD) (Stribling et al. 2008). Details on the calculations of these fields can be found in the external lab SOP.

Standard taxonomic effort and measurement targets were defined by the contracted labs and used for data collected after October 2018. Taxonomists follow the most current nomenclature that is “accepted” in the literature for each taxon. The current list of standard taxonomic targets, references, and measurement guidelines can be found in the NEON document library:

https://data.neonscience.org/api/v0/documents/macroinvertebrate_standard_taxonomic_effort_vA

If errors in identification are detected through QAQC processes after data publication, then corrected taxonomy is provided in the inv_taxonomyProcessed and inv_taxonomyRaw tables and previous taxonomic information is preserved in the inv_identificationHistory table (see Sections 3.9 Data Relationships and 4.1 Identification History above for more details).

6 REFERENCES

Benke, A. C., A. D. Huryn, L. A. Smock, and J. B. Wallace. 1999. Length-mass relationships for freshwater macroinvertebrates in North America with particular reference to the southeastern United States. Journal of the North American Benthological Society 18: 308-343.

Biggs, B. J. F., R. A. Smith, and M. J. Duncan. 1999. Velocity and sediment disturbance of periphyton in headwater streams: biomass and metabolism. Journal of the North American Benthological Society 18: 222-241.

Brooks, S. S. and A. J. Boulton. 1991. Recolonization dynamics of benthic macroinvertebrates after artificial and natural disturbances in an Australian temporary stream. Australian Journal of Marine and Freshwater Research 42:295-308.

Hauer, F. R. and V. H. Resh. 2006. Macroinvertebrates. Pages 435-463 in F. R. Hauer and G. A. Lamberti, editors. Methods in Stream Ecology, Second Edition. Academic Press, Boston, MA.

Matthaei, C. D., U. Uhlinger, E. I. Meyer, and A. Frutiger. 1996. Recolonization by benthic invertebrates after experimental disturbance in a Swiss prealpine river. Freshwater Biology 35: 233-248.

Merritt, R. W., K. W. Cummins, and M. B. Berg. 2019. An Introduction to the Aquatic Insects of North America, 5th Edition. Kendall Hunt Publishing Company, Dubuque, Iowa. 1480 pp.

Moulton, S. R., II, J. G. Kennen, R. M. Goldstein, and J. A. Hambrook. 2002. Revised protocols for sampling algal, invertebrate, and fish communities as part of the National Water-Quality Assessment Program. Open-File Report 02-150. U.S. Geological Survey, Reston, VA.

Stribling, J. B., K. L. Pavlik, S. M. Holdsworth, and E. W. Leppo. 2008. Data quality, performance, and uncertainty in taxonomic identification for biological assessments. Journal of the North American Benthological Society. 27: 906-919.