NEON USER GUIDE TO FISH ELECTROFISHING, GILL NETTING, AND FYKE NETTING COUNTS (DP1.20107.001)

PREPARED BY	ORGANIZATION
Ryan McClure	AQU
Dylan Monahan	AQU
Brandon Jensen	AQU
Stephanie Parker	AQU
Caren Scott	AQU

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	11/08/2017	Initial Release
B	10/26/2020	Included general statement about usage of neonUtilities R package and statement about possible location changes. Updated taxonomy information. Changes captured revB include the addition of macroinvertebrate bycatch data collection during fish sampling at D04. Also included in revB is the addition of methodology allowing for the identification of fish caught above and below permanent fish barriers that bisect fish sampling transects at MCDI and TECR.
C	03/05/2024	Updated information about the fsh_identificationHistory table
D	03/13/2025	Updated document wording for clarity. Added a bullet in the Sampling Design Changes section describing how the sampling bout window had been reduced from five to three days, and the expectation for fixed reaches to be sampled has been updated from three to one. Added an additional bullet in the Sampling Design Changes section describing the update sampling design in the stream sites with barriers. Added page breaks between sections. Added information about the new neonUtilities Python package.

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, fish length and weight measurements, 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 Fish electrofishing, gill netting, and fyke netting counts 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 Fish electrofishing, gill netting, and fyke netting counts (DP1.20107.001) (AD[05]), provided in the download package for this data product.

This document also 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: Fish Sampling in Wadeable Streams (AD[07]) and AOS Protocol and Procedure: Fish Sampling in Lakes (AD[08]). The raw data that are processed in this document are detailed in the file, NEON Raw Data Validation for Fish electrofishing, gill netting, and fyke netting counts (DP0.20107.001) (AD[04]), 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., '20107') 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.001152	NEON Aquatic Sampling Strategy
AD[03]	NEON.DOC.002652	NEON Data Products Catalog
AD[04]	Available with data download	Validation csv
AD[05]	Available with data download	Variables csv
AD[06]	NEON.DOC.001152	NEON Aquatic Sampling Strategy
AD[07]	NEON.DOC.001295	AOS Protocol and Procedure: Fish Sampling in Wadeable Streams
AD[08]	NEON.DOC.001296	AOS Protocol and Procedure: Fish Sampling in Lakes
AD[09]	NEON.DOC.000008	NEON Acronym List
AD[10]	NEON.DOC.000243	NEON Glossary of Terms
AD[11]	NEON.DOC.004825	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[12]	Available on NEON data portal	NEON Ingest Conversion Language Function Library
AD[13]	Available on NEON data portal	NEON Ingest Conversion Language
AD[14]	Available with data download	Categorical Codes csv

3 DATA PRODUCT DESCRIPTION

Data products from fish sampling (electrofishing, gill netting, and fyke netting) include information on fish taxonomy, abundance, morphometrics, and community structure in streams and lakes. Each individual fish captured during fish sampling is identified to the lowest common taxonomic level possible. A minimum of 50 fish per taxon/per reach (30 fish at BLUE, and all lake sites) are weighed to 0.1 grams (with any fish below 0.5 grams recorded as weighing 0.3 to account for the scales minimum detection range), and total length is measured in millimeters. Also recorded for the first 50 fish per species/per reach: life stage (if it can be determined in the field) and the presences or absence of deformities, eroded fins, lesions, tumors, and parasites. Voucher specimens and DNA Fin clip samples may also be collected.

After the first 50 individuals of a species/per reach are weighed and measured, the remaining specimens of that species are bulk counted and tallied, either by actually counting each individual or estimating in cases where hundreds or more fish are captured. Bulk fish are identified to the lowest taxonomic level possible, and tallied, no other data is collected for bulk species.

All fish specimens must have a taxon assigned to it. For fish who species assignment cannot be determined in the field, field ecologist will assign that individual or group either an Identification Qualifier or a morphospecies designation. Identification Qualifier's are indicators to the data user of the taxonomic level at which the field ecologist is uncertain about, but reserved for species where the NEON field ecologist can narrow the possible identification to one of three species. Table 1, is a list of identification qualifier code and the codes description used by NEON and published in NEON data.

Morphospecies is a designation for individuals and groups that cannot be assigned a species assignment, when the field ecologist determines the potential species is one of more than three possibilities. Morphospecies are named using the following format: the **domainID** and the **siteID** where the specimen was captured, the year of capture, and the word "Morph" followed by one or more unique letters. For example, "D15.REDB.2014.MorphA". The same letter will be given to all fish with the same morphological traits, in all the succeeding bouts that those morphospecies are captured and require a morphospecies code, only the year will change to reflect the current year in which the morphospecies is caught. When morphospecies identification is resolved to species or the lowest known taxonomic level, the resolution will be recorded in the fsh_morphospecies table. The morphospecies ID will stay with the original record but can be corrected by the user using the morphospecies identification in the fsh_morphospecies table.

For each sampling event, the total seconds the button was depressed during electrofishing and total net set time for fyke and gill nets is recorded. These data can be used to calculate relative abundances and catch per unit effort (CPUE).

At NEON's D04 Rio Cupeyes and Rio Guilarte invertebrate bycatch caught during electrofishing is also tallied, recorded, and identified to the lowest possible taxonomic level.

Fish are sampled twice annually at most NEON aquatic sites (AD[06]). All three river sites (TOMB, BLWA, FLNT), the two D03 lakes (SUGG and BARC), and D13's COMO creek are not sampled for fish. D16's MART creek is only sampled once annually in October. Sampling dates are based on a combination of variables, including hydrology in streams or ice on/ice off dates in lakes, the riparian greenness, and accumulated degree days (°C). Measurements are collected by field personnel employing various sampling techniques depending on the aquatic system being studied. For additional information see sampling design NEON

Aquatic Sampling Strategy (AD[06]) and protocols AOS Protocol and Procedure: Fish Sampling in Wadeable Streams (AD[07]) and AOS Protocol and Procedure: Fish Sampling in Lakes (AD[08]).

3.1 Spatial Sampling Design

Fish sampling at NEON wadeable stream aquatic sites (Figure 1) is conducted seasonally in up to six of ten non-overlapping, 40-120 meter permanently established fish sampling reaches nested in the larger ~1km NEON sampling reach (NOTE: some NEON sites may be less than 1 km due to permitting restrictions). Three of the ten reaches are designated as "fixed" reaches, and at least one "fixed" reach is sampled twice annually using a three-pass electrofishing depletion method (Moulton et al. 2002 and Peck et al. 2006). The remaining seven reaches are designated as "random" reaches, which are sampled following a rotating pattern to ensure that the variety of habitat types are sampled equally over time. Specifically, three random reaches at each site are sampled twice in any given year then a new set of three random reaches are selected and sampled in each subsequent year. Random reaches are sampled via a single-pass depletion approach. See AOS Protocol and Procedure: Fish Sampling in Wadeable Streams (AD[07]) for additional details on sampling strategy and SOPs.

Fish sampling at NEON lake aquatic sites (Figure 1) include up to ten individual segments delineated within the lake and including both nearshore and offshore fish populations to estimate species composition, species diversity, relative abundance, and an indication of the distribution of species within the NEON lake. Lake fish sampling segments occur between two riparian transects and converge at the approximate center of the lake (Figure 1). From these ten fish sampling segments, three permanent "fixed" segments were selected prior to the first sampling bout at the site. At least one of the fixed segments is sampled twice per year with a backpack electrofisher using a three-pass electrofishing depletion approach (Moulton et al. 2002 and Peck et al. 2006). Additionally, each fish sampling segment reach includes one mini-fyke net set and one gill net set (Baker et al. 1997). The remaining seven lake segments are established as "random" segments and sampled following a rotating design to ensure that the variety of habitat types are sampled equally over time. A rotating sampling design with initial random selection of shoreline segments ensures appropriate spatial coverage of habitat types within the lake (Baker et al. 1997). The same random segments are sampled in spring and fall of a given year. Random segments are sampled with a single electrofishing pass (without block nets), one mini-fyke net set, and one gill net set. See AOS Protocol and Procedure: Fish Sampling in Lakes (AD[08]) for additional details on sampling strategy and SOPs.

Sampling occurs in the same locations over the lifetime of the NEON Observatory. However, over time some sampling locations may become impossible to fish due to disturbance or other local changes. When this occurs, the reach and its location identifier are retired or updated (i.e., a location identifier may also shift to slightly different coordinates). Refer to the locations endpoint of the NEON API for details about locations that have been updated or retired: https://data.neonscience.org/data-api/endpoints/locations/

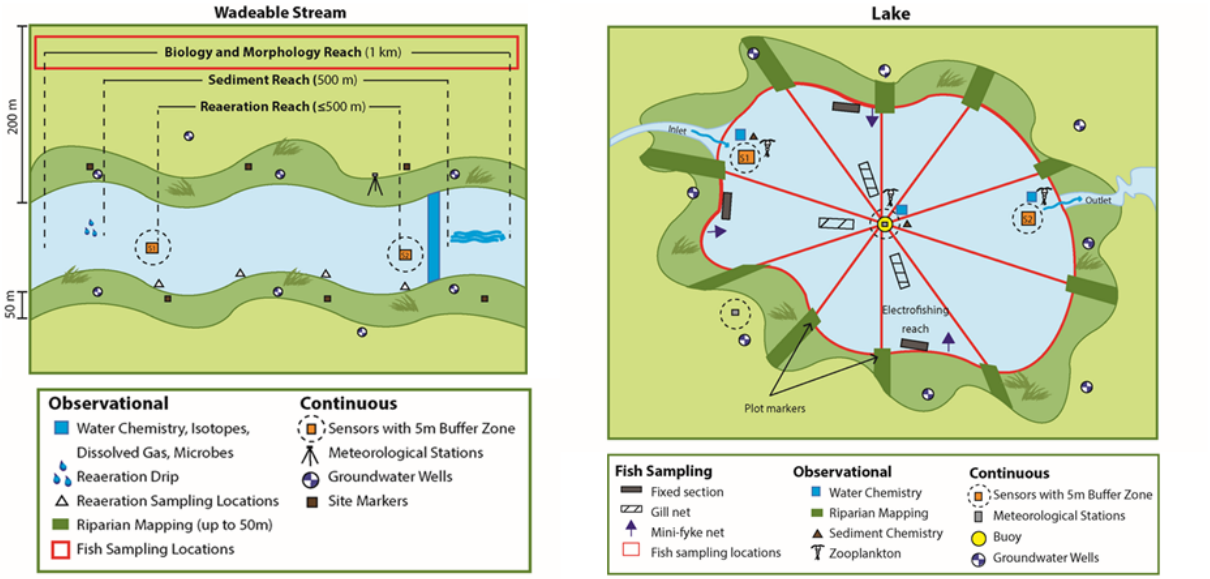

Figure 1: Generic aquatic site layouts (wadeable streams, and lakes) with fish sampling locations in red.

3.2 Temporal Sampling Design

Fish sampling occurs two times per year corresponding with the first and third sampling bout windows (roughly spring and fall for the northern hemisphere) at wadeable stream and lake sites, with a minimum of two weeks between sampling bouts. The initial sample timing is determined for each site using historical data including ice-out, the riparian greenness, and water temperature (or accumulated degree days, °C). Sample timing will be refined on a site-by-site basis by NEON Science using NEON aquatic sensors and aquatic observation data. Sampling bouts are targeted to be completed within three days. All three depletion passes in a fixed sampling segment must be sampled within the same day, with at least 30 minutes between passes to allow fish to resettle in the reach. Weather conditions at the site may push sampling outside of the bout window. See AOS Protocol and Procedure: Fish Sampling in Wadeable Streams (AD[07]) and AOS Protocol and Procedure: Fish Sampling in Lakes (AD[08]) for additional details on sampling strategy and SOPs.

Lake fish assemblage characterization requires multiple sampling methods that are optimal for sampling fish at different times of the day. Electrofishing occurs at night, starting 30 minutes after sunset and ceasing 30 minutes before sunrise (or during lowest-light hours at Arctic sites). Gill nets are set and sampled during daylight hours, with a preferred set time of up to 1 hour and maximum set time of 2 hours. Gill nets are set in the morning or early afternoon to allow for processing time. Mini-fyke nets are set before sunset and allowed to remain in the water until after sunrise the following morning.

Adjustments to the planned temporal sampling design may occur due to heavy rainfall, flooding, or unsafe wading conditions (Lane and Fay 1997). See NEON Aquatic Sampling Strategy (AD[06]), AOS Protocol and Procedure: Fish Sampling in Wadeable Streams (AD[07]) and AOS Protocol and Procedure: Fish Sam-

pling in Lakes (AD[08])for additional details.

3.3 Sampling Design Changes

- At lake sites, the pass number assigned to fyke and gill net sets in 2016 and 2017 is pass number one. Starting in 2018 fyke net sets are numbered pass 4, and gill net sets are numbered pass 5
- Fish bulk count records started after 100 individuals of each taxon were measured for length and weight before May 2017. After May 2017, fish bulk counts started after 50 individuals of each taxon were measured for length and weight.
- In March 2023, the number of individual fish taxa measured for length and weight before starting the bulk count records were lowered further from 50 to 30 individuals of each taxon. This additional reduction has only been implemented at BLUE, CRAM, LIRO, TOOK, PRPO, and PRLA sites. All remaining stream and lake sites should have 50 individual taxa per reach measured for length and weight before bulk records are generated.
- In August 2023, SYCA adopted a sampling design change and lowered the number of individual fish taxa measured for length and weight before starting the bulk count records from 50 to 30 when water temperatures in SYCA are between 22°C and 26°C. This is specific only to SYCA and minimizes the handling time and stress to fish. The maximum temperature limits for electrofishing in lakes and streams across all of NEON are >18°C for sites with salmonids and >26°C for non-salmonid sites, unless stated otherwise in a site specific permit.
- The original fish sampling design occurred over five days, and field scientists were targeting three fixed reaches and three random reaches. Starting in 2020, resource limitations prevented this full sampling design across most of the observatory, with inconsistent strategies for how to reduce the sampling design. Starting in January 2025, the fish sampling design was formally reduced, and field scientists now target sampling one fixed reach and three random reaches (reducing the overall effort from 5 days to 3). If scheduling allows, or if fishing is faster, some sites will still have the ability to follow the original sampling design. The fixed reaches are fished in the order from the most to least representative of the site.
- Three of NEON's stream sites include one or more barriers. This includes one permanent fish barrier bisecting a fish sampling reach at McDiffett Creek (MCDI, D06) in Kansas, one permanent barrier at NEON's Walker Branch (WALK, D07) in Tennessee, and two permanent fish barriers bisecting fish sampling transects at NEON's Teakettle Creek (TECR, D17) site in California. The reaches with the barriers include reach five at MCDI, reach six at WALK, and reach one and reach seven at TECR. Before 2019, fish sampling was conducted without identifying the existence of the barriers. Thus, there is no clear indication whether the fish enumerated in the fsh_perFish or fsh_bulkCount tables came from above or below the barriers. From 2019 to 2024, fish were captured and enumerated above and below the barriers and identified whether they were caught above or below. However, quality control in 2024 revealed that this approach led to sampling bias toward the longest section of reach, where data in the fsh_perFish table predominantly came from below the barrier and the data from the fsh_bulkCount data came from above. Additionally, it is unclear whether sampling teams prevented downstream fish passage across the barrier during the sampling. Thus, species richness estimates should be unaffected from the uncertainty in downstream fish passage, but

abundance is likely to be inaccurate. Starting in 2025 the reaches with barriers have been reduced in length such that fishing now only occurs in the longest, or safest if necessary (e.g., reach one in TECR), fishable sections of the original reach.

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 Fish electrofishing, gill netting, and fyke netting counts (DP0.20107.001) (AD[04]). All variables reported in the published data (L1 data) are also provided separately in the file, NEON Data Variables for Fish electrofishing, gill netting, and fyke netting counts (DP1.20107.001) (AD[05]).

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 Geoid12A geoid model for its vertical reference surface. 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 Spatial Resolution and Extent

The finest resolution at which spatial data are reported is a single fish sampling reach or segment (Figure 1). At wadeable stream sites, ten consecutive fish sampling reaches (50-120 m) are established throughout the ~1 km aquatic biological sampling reach. Lake fish sampling segments occur between two riparian transects and converge at the approximate center of the lake (Figure 1). The exact location (latitude and longitude) of each sample collected is not tracked as it is intended to represent the overall habitat at the reach (wadeable streams) or segment (lakes) level. The **namedLocation** reported in wadeable streams represents the fish reach locations within the NEON aquatic reach. The **namedLocation** reported in the lakes represents the segment locations within NEON lake sites. Sampling locations are tracked by latitude and longitude and include an indication of **coordinateUncertainty**.

Overall,this results in a spatial hierarchy of:

namedLocation (finest spatial resolution, ID of location within site) →siteID (ID of NEON site) →domainID (ID of a NEON domain)

3.6 Temporal Resolution and Extent

The finest resolution at which temporal data are reported is per **eventID**, **namedLocation**, and **passNumber** (a single sampling pass at a single reach or segment for a fishing bout), where both a **passStartTime** and **passEndTime** are reported.

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 download and stack files across months is available here: https://github.com/NEONScience/NEON-utilities.

3.7 Associated Data Streams

The **dnaSampleID** in the fsh_perFish table is a linking variable that can be used to tie specific samples and associated metadata to the Fish sequences DNA barcode (DP1.20105.001) product.

Fish electrofishing, gill netting, and fyke netting counts 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** and date fields.

3.8 Product Instances

At each aquatic site, there will be up to two fish sampling bouts per year during which up to six reaches or segments are sampled, generating a maximum of 12 records in fsh_fieldData per site per year. Each reach/segment is sampled using up to five passes during each bout, yielding up to 48 records in fsh_perPass per site per year. Up to 50 fsh_perFish records are expected per taxonID, per site, per pass, per year, with the remainder of fish enumerated by bulk counting. As a result, the number of instances of fsh_perFish and fsh_bulkCount varies with the diversity and abundance of the site.

3.9 Data Relationships

The protocol dictates that a reach is sampled within a single day of the year (local time), yielding a unique record for that reach. A record from fsh_fieldData will have up to five child records in fsh_perPass. Each record from fsh_perPass, with a unique eventID, will have up to 50 child records per taxonID in fsh_perFish and zero or multiple child records in fsh_bulkCount, depending on the taxonomic diversity and abundance at the site.

Some records that are unique to a reach may also yield incomplete downstream records. Data from the fsh_fieldData table that have the **samplingImpractical** field identified (e.g. the location is dry, ice-covered, etc) will not have matchable records in downstream tables (i.e. fsh_perPass, fsh_perFish, fsh_bulkCount, etc). Additionally, records from the fsh_perPass table, with a unique eventID that matches the fsh_fieldData table, will not have matchable records in downstream tables (i.e. fsh_perFish, fsh_bulkCount, etc) if **targetTaxaPresent** = 'No'. 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 and remove records that are **samplingImpractical** or **targetTaxaPresent** = 'No'.

fsh_fieldData →One record is created for each sampling activity per reach, creating a unique **eventID** and **namedLocation** that corresponds with a bout and the reaches that were sampled. Records in this table are unique by the combination of **eventID** and **namedLocation**. There may be up to six records for each bout at wadeable stream and lake sites including three fixed reaches and three random reaches. This table also indicates the reach conditions **reachCondition**, the length of the reach **measuredReachLength**, if

it is a fixed or random reach **fixedRandomReach**, whether the record occurred in the spring or fall in the Northern Hemisphere **boutID**, and the habitat type **habitatType**.

fsh_perPass →One record is created for each **passNumber** for each **namedLocation** (i.e. reach) within a corresponding **eventID**. Records in this table are unique by the combination of **eventID**, **namedLocation**, and **passNumber**. There may be up to three electrofishing passes for fixed reaches and just one for random reaches at wadeable stream and lake sites. One record for each gill net and mini-fyke net efforts are created at the pass level per reach. The fsh_perPass table also includes water quality measurements and other metadata regarding the electrofishing units and the net integrity for each **passNumber**. Data from this table are joined to the upstream fsh_fieldData table using the **eventID** and **namedLocation** fields and joined to the downstream fsh_perFish and fsh_bulkCount table through the **eventID**, **namedLocation**, and **passNumber** fields.

fsh_perFish →One record is created for each unique taxonomic identification **specimenNumber** with a corresponding, **fishTotalLength**, **fishWeight**, and **fishLifeStage**. Records in this table are unique by the combination of **eventID**, **namedLocation**, **passNumber**, and **specimenNumber**. Taxonomic identifications are made to the lowest practical taxonomic level (typically genus or species) with a corresponding **taxonID** and **scientificName**. Data are linked to the upstream fsh_perPass table through the **eventID**, **namedLocation**, and **passNumber** fields and joined to the downstream fsh_bulkCount table through the **eventID**, **namedLocation**, **passNumber**, and **taxonID** or **scientificName** fields.

fsh_bulkCount →One record with a taxonomic identification and a raw bulk count for fish that exceeded 50 fish processed from the fsh_perFish table. Records in this table are unique by the combination of **eventID**, **namedLocation**, **passNumber** and **taxonID**. **bulkCount** values may either be *actual* or *estimated* depending on the number of bulk fish that need to be counted (e.g., **bulkCount** data with ≤500 fish will have *actual* counts while **bulkCount** data with >500 will generally have *estimated*). These data must be summed with individual fish from the fsh_perFish table in order to generate tallied taxonomic-level fish counts per pass. Data from the fsh_bulkCount table can be linked with the fsh_perFish table through the **eventID**, **namedLocation**, **passNumber**, and **taxonID** fields.

fsh_morphospecies →One record is created for each individual or group of fish that are morphologically similar but technicians are unable to identify. Technicians are instructed to use the same unique name for each morphospecies type, but change the year in the ID to reflect the year of capture (**morphospeciesID**). Thus, these data may be linked to the fsh_perFish and fsh_bulkCount through the **morphoSpeciesID** and dates. Not all morphospecies are subsequently resolved to a formal taxonomic designation; so **morphospeciesID** values in fsh_perFish and fsh_bulkCount may lack corresponding records in fsh_morphospecies and the fsh_morphospecies table will only have resolved morphospecies.

fsh_invertBycatch →At D04 one record is created for each bulk count of macroinvertebrates (one per taxon) captured as bycatch during fishing. These data can be linked to the other fish data tables through **eventID**, **namedLocation**, and **passNumber**.

Data downloaded from the NEON Data Portal are provided in separate data files for each site and month requested. The neonUtilities package in R and the neonutilities package in Python contain functions to merge these files across sites and months into a single file for each table. The neonUtilities R 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. The neonutilities

package in Python is available on the Python Package Index (PyPi; https://pypi.org/project/neonutilities/) and can be installed using pip. For instructions on using the package in either language 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.

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 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.

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 Fish Taxonomy

The master taxonomy for fishes is derived from Froese and Pauly (2016) and the Integrated Taxonomic Information System (ITIS) on-line database (http://www.itis.gov). Taxon ID codes used to identify taxonomic concepts in the NEON master taxonomy list were generated for each fish by concatenating the first three letters of the genus name together with the first three letters of the species name to make a unique taxon ID for each scientific name. Where such concatenation would produce duplicate taxon ID codes, numbers were appended to the taxon ID until it was unique within the NEON database. For species in which the lowest possible taxonomic level identification is higher than species the first three letter of that lowest taxonomic level are used followed by SP or SPP (e.g. the code for Fundulus sp. is FUNSP). NEON plans to keep the taxonomy updated in accordance with the current literature starting in 2020 and annually thereafter.

Errors are generated if a species is reported at a location outside of its known range. If the record proves to be a reliable report, the master taxonomy table is updated to reflect the distribution change. Geographic ranges and native statuses used in this data product are from Froese and Pauly (2016) and the Integrated Taxonomic Information System (ITIS) on-line database (http://www.itis.gov).

4.2 Macroinvertebrate Taxonomy

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.

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.

4.3 Identification History

Beginning in 2022, the fsh_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 fsh_perFish and fsh_bulkCount 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 fsh_identificationHistory table where the previous taxonomic information is archived along with the date the change was made.

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. 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 Fish electrofishing, gill netting, and fyke netting counts (DP0.20107.001), 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[12]).

Data collected prior to 2017 were processed using a paper-based workflow that did not implement the full suite of quality control features associated with the interactive digital workflow.

5.2 Automated Data Processing Steps

Following data entry into a mobile application or 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[11]).

5.3 Data Revision

All data are provisional until a numbered version is released; the first release of a static version of NEON data, annotated with a globally unique identifier, is planned to take place in 2020. 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 Change Log section of the data product readme, provided with every data download, 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 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
dataQF	reachContainsBarrier–this section not fished after 2024	Identifies reaches that contain barriers and then sectionalizes what section of the reach is no longer going to be fished
dataQF	reachContainsBarrier-useWithCaution	Identifies reaches that contain barriers and then disseminates to users that data prior to 2024 for this reach should be used with caution due to in-cosistent sampling trechniques above and below barriers
dataQF	efTime estimated - use with caution for CPUE and efTimes for these records	efTimes for each pass were not recorded, however the total efTime for all three passes combined of that reach was recorded and was used to estimate the efTime per pass by dividing this total equally into the three passes.

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)

6 REFERENCES

Baker, J. R., D. V. Peck, and D. W. Sutton (editors). 1997. Environmental Monitoring and Assessment Program Surface Waters: field operations manual for lakes. EPA/620/R-97/001. U.S. Environmental Protection Agency, Washington D.C.

Froese, R. and D. Pauly. FishBase. World Wide Web electronic publication. (https://www.fishbase.de/home.htm) 10/2016

Lane, S .L. and R. G. Fay. 1997. Safety in field activities: U.S. Geological Survey techniques of water-resources investigations, Book 9, Chapter A9. Accessed 11 April 2013, at http://pubs.water.usgs.gov/twri9A9/.

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., 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, Reston, Virginia.

Peck, D. V., A. T. Herlihy, B. H. Hill, R. M. Hughes, P. R. Kaufmann, D. J. Klemm, J. M. Lazorchak, F. H. McCormick, S. A. Peterson, P. L. Ringold, T. Magee, and M. R. Cappaert. 2006. Environmental Monitoring and Assessment Program-Surface Waters: Western pilot study field operations manual for wadeable streams. EPA/620/R-06/003. USEPA. Washington, DC.