NEON USER GUIDE TO FISH ELECTROFISHING, GILL NETTING, AND FYKE NETTING COUNTS (NEON.DP1.20107)

PREPARED BY	ORGANIZATION	DATE
Brandon Jensen	AQU	11/09/2017
Stephanie Parker	AQU	11/09/2017
Caren Scott	AQU	11/09/2017

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	11/08/2017	Initial Release

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 (NEON.DP1.20107) (AD[05]), 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: 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 (NEON.DP0.20107) (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 Level 1, Level 2 and Level 3 Data Products Catalog
AD[04]	NEON.DP0.20107.001 _dataValidation.csv	NEON Raw Data Validation for Fish electrofishing, gill netting, and fyke netting counts (NEON.DP0.20107)
AD[05]	NEON.DP1.20107.001 _variables.csv	NEON Data Variables for Fish electrofishing, gill netting, and fyke netting counts (NEON.DP1.20107)
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]	OS_Generic_Transitions .pdf	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[12]		NEON's Ingest Conversion Language (NICL) specifications

3 DATA PRODUCT DESCRIPTION

The fish electrofishing, gill netting, and fyke netting counts data product includes information on fish taxonomy, abundance, morphometrics and community structure in streams and lakes. After processing (weighing and measuring) a minimum of 50 fish per taxon, the remaining specimens are bulk counted, either by actually counting each individual or (in cases where hundreds or more fish captured), estimating. Estimated counts are achieved by scooping and counting the total number of specimens in one dip net and then multiplying the total number of scoops of captured fish by the fish counts from the first scoop. This method assumes a homogenous composition of species. Data on life stage, deformities, eroded fins, lesions, tumors, and parasites are also recorded. Voucher specimens and tissue samples may also be collected. These data can be used to assess the health of aquatic ecosystems. Fishes are sampled two times per year at each NEON aquatic site (AD[06]). 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). 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 in up to ten replicate, non-overlapping 80-120 m reaches. Three of these reaches are designated as “fixed” reaches, which are sampled twice every year using a three-pass electrofishing depletion approach (Moulton et al. 2002 and Peck et al. 2006). Fixed reaches include representative habitat features that are found throughout the entire 1 km NEON aquatic biological sampling reach (NOTE: some NEON sites may be less than 1 km due to permitting restrictions). 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. Capture probabilities estimated from fixed three-pass depletion reaches may be used to estimate population sizes of fish within the random single-pass reaches. 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 10 segments established within the lake for studying 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. The sampling segments are established using the 10 pre-determined lake riparian habitat transect locations. Lake fish sampling segments occur between two riparian transects and converge at the approximate center of the lake (Figure 1). For example, fish segment 1 occurs between riparian transect 1 and 2. Fish segment 2 occurs between riparian transect 2 and 3. Three permanent “fixed” segments are established which include habitat features most representative of the entire lake. The fixed segments are 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 sampled 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 appro-

priate 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.

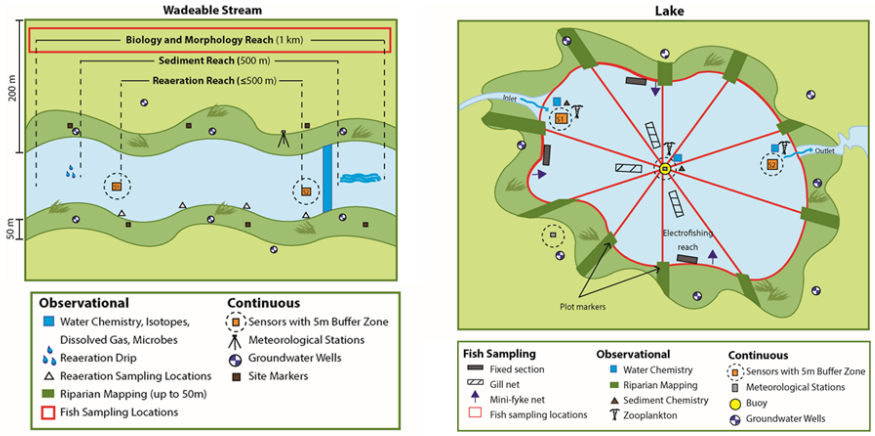

Figure 1: Generic aquatic site layouts (wadeable streams, and lakes) with fish sampling locations in red.

3.2 Temporal Sampling Design

Fish sampling occurs two times per year corresponding with the first and third sampling bout windows (roughly spring and autumn) 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, water temperature (or accumulated degree days), and riparian peak greenness. Sample timing will be refined on a site-by-site basis by Science Operations based on data collected by the aquatic sensors and Field Operations. Sampling bouts are targeted to be completed within 5 days. All three 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 Sampling in Lakes (AD[08])for additional details.

3.3 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 (NEON.DP0.20107) (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 (NEON.DP1.20107) (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 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.

3.4 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 (80-120 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 **locationID** reported in wadeable streams represents the fish reach locations within the NEON aquatic reach. The **locationID** 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:

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

3.5 Temporal Resolution and Extent

The finest resolution at which temporal data are reported is per **eventID** (a single sampling pass at a single reach or segment), 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 stack files across months is available here: https://github.com/NEONScience/NEON-utilities

3.6 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 (NEON.DP1.20105) 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.

3.7 Product Instances

At each aquatic site, there will be up to 2 fish sampling bouts per year during which 6 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 5 passes during each bout, yielding up to 60 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.

3.8 Data Relationships

The protocol dictates that reach is sampled within a single day of year (local time), yielding a unique **reachID**. A record from fsh_fieldData may have zero (if sampling is impractical; e.g. the location is dry, ice-covered, etc) or up to 5 child records in fsh_perPass, depending on whether the reach is being sampled using multiple electrofishing passes, and/or multiple sampler types. Each record from fsh_perPass, with a unique eventID, may have zero (if **targetTaxaPresent** = 'No') or up to ~50 child records per taxonID in fsh_perFish and 0 or multiple child records in fsh_bulkCount, depending on the taxonomic diversity and abundance at the site. 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.

fsh_fieldData.csv - > One record is created for each sampling activity per reach, creating a **reachID** which is linked to the fsh_perPass table. This table also indicates the reach conditions **reachCondition**, the length of the reach **measuredReachLength**, if it is a fixed or random reach **fixedRandomReach**, and the habitat type **habitatType**.

fsh_perPass.csv - > One record is created for each pass **passNumber** for each **reachID**. There may be up to 3 electrofishing passes for fixed reaches/segments and just one for random reaches/segments at wadeable stream and

lake sites. One record for gill net and mini-fyke net efforts are created at the pass level per reach or segment. This table includes water quality measurements and other metadata about the sampling pass. Data from this table are linked to the fieldData through the **reachID**.

fsh_perFish.csv - > One record is created for each individually-measured fish collected. For fish, taxonomic identifications are made to the lowest practical taxonomic level (typically genus or species). Data are linked to the fsh_perPass table through the **eventID** in each table. Records in this table are unique by the combination of **specimenNumber** and **eventID**.

fsh_bulkCount.csv - > One record is created for each bulk count of fish collected in the field. These data are linked to the fsh_perPass through the **eventID** in each table.

fsh_morphospecies.csv - > 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 unique names (**morphospeciesID**) within a domain per year. 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.

4 DATA QUALITY

4.1 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 (NEON.DP0.20107), 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.

4.2 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]).

4.3 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.

4.4 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.

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

5 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.

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/.

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.