NEON USER GUIDE TO AQUATIC ZOOPLANKTON COLLECTION (NEON.DP1.20219)

PREPARED BY	ORGANIZATION	DATE
Stephanie Parker	AQU	12/18/2017
Caren Scott	AQU	12/18/2017
		12/18/2017

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	09/25/2017	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, the zooplankton 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 Zooplankton 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 Zooplankton Collection (NEON.DP1.20219) (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: Zooplankton Sampling in Lakes (AD[06]). The raw data that are processed in this document are detailed in the file, NEON Raw Data Validation for Aquatic Zooplankton Collection (NEON.DP0.20219) (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 Level 1, Level 2 and Level 3 Data Products Catalog
AD[03]	NEON.DP0.20219.001 _dataValidation.csv	NEON Raw Data Validation for Aquatic Zooplankton Collection (NEON.DP0.20219)
AD[04]	NEON.DP1.20219.001 _variables.csv	NEON Data Variables for Aquatic Zooplankton Collection (NEON.DP1.20219)
AD[05]	NEON.DOC.001152	NEON Aquatic Sampling Strategy
AD[06]	NEON.DOC.001194	AOS Protocol and Procedure: Zooplankton Sampling in Lakes
AD[07]	NEON.DOC.000008	NEON Acronym List
AD[08]	NEON.DOC.000243	NEON Glossary of Terms
AD[09]	OS_Generic_Transitions .pdf	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[10]	Nicl Language.pdf	NEON's Ingest Conversion Language (NICL) specifications

3 DATA PRODUCT DESCRIPTION

The zooplankton collection data product includes taxonomy, abundance and morphometrics. The type of sampler used to collect zooplankton in the field is determined by the water depth at the sampling location. Locations deeper than 4 m are sampled using a vertical tow net, while locations shallower than 4 m are sampled using a Schindler-Patalas sampler (USEPA 2012a, 2012b). Typically, multiple (up to 3) tows or Schindler traps are collected and composited into a single sample.

Samplers used for zooplankton collection are designed to work by collecting a parcel of the water column at the sampling location. The volume collected with a vertical tow net is the area of the tow net opening multiplied by the depth of the tow, multiplied by the number of tows composited in one sample. The Schindler-Patalas trap is a clear Lexan box that contains 12 L of water, so the volume is 12 L times the number of traps composited in one sample.

Samples are preserved in ETOH and sent to professional taxonomic laboratories for taxonomic identification. For additional information see sampling design NEON Aquatic Sampling Strategy (AD[05]) and protocol AOS Protocol and Procedure: Zooplankton Sampling in Lakes (AD[06]).

These data inform the NEON Grand Challenge area of Biodiversity as well as provide specific information about the zooplankton community in streams and lakes. These data can be used to assess the health of aquatic ecosystems.

3.1 Spatial Sampling Design

Zooplankton at NEON aquatic sites (Figure 1) are sampled from the water column near the NEON sensor sets in lakes. Samples are collected near the inlet, outlet, and buoy (deepest) sampling locations.

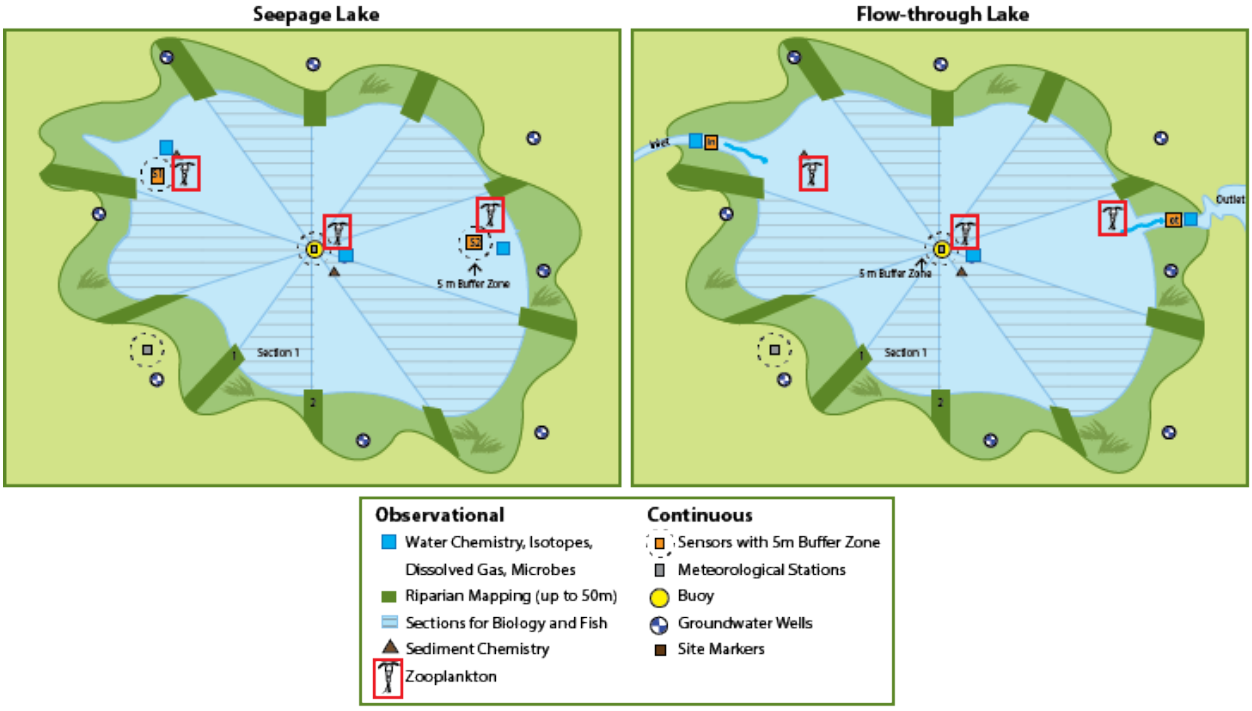

Figure 1: Generic aquatic site layout for lakes with zooplankton sampling locations in red.

3.2 Temporal Sampling Design

Zooplankton sampling occurs three times per year at each NEON lake site (AD[05]). 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 after 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 rain or wind event that causes turbidity in the water column. Sampling at each site is completed within a single day for each bout. See NEON Aquatic Sampling Strategy (AD[05]), AOS Protocol and Procedure: Zooplankton Sampling in Lakes (AD[06]) 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 Aquatic Zooplankton Collection (NEON.DP0.20219) (AD[03]). All variables reported in the published data (L1 data) are also provided separately in the file, NEON Data Variables for Aquatic Zooplankton Collection (NEON.DP1.20219) (AD[04]).

Field names have been standardized with Darwin Core terms (http://rs.tdwg.org/dwc/; accessed 7 September 2017), the Global Biodiversity Information Facility vocabularies (http://rs.gbif.org/vocabulary/gbif/; accessed 7 September 2017), the VegCore data dictionary (https://projects.nceas.ucsb.edu/nceas/projects/bien/wiki/VegCore; accessed 7 September 2017), 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 Temporal Resolution and Extent

The finest temporal resolution that zooplankton data will be tracked is per sampling day. All 3 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.

3.5 Spatial Resolution and Extent

Each zooplankton sample represents a location in a lake near one of the NEON sensor installations (inlet, outlet, or buoy), 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 represents the location of the NEON sensor infrastructure near the sampling location, plus coordinate uncertainty surrounding that point. The protocol dictates that samples are collected approximate 5 m from the sensor infrastructure to minimize effects on the profiling data, so standard **coordinateUncertainty** is 10 m to represent the normal sampling distance from the sampling location. If, for some reason, sampling cannot occur within 10 m of the named location, technicians will enter **additionalCoordinateUncertainty**.

Up to two different habitats (littoral and pelagic) are sampled at each site to account for the variability or patchiness among habitats. 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 Associated Data Streams

A subset of the zooplankton field collection data are related to Zooplankton DNA Barcode (NEON.DP1.20221) samples collected at the same time and location. Related samples share the same **eventID** and **namedLocation**.

Zooplankton collection data are also loosely related to Aquatic General Field Metadata (NEON.DOC.001646) and Secchi Disk and Depth Profile Sampling in Lakes and Non-wadeable Streams (NEON.DOC.002792) collected on the same sampling day. Data for Aquatic General Field Metadata are available in the NEON data product "Gauge Height" (DP1.20267.001). Depth profile and secchi data are available in Depth profile at specific depths (DP1.20254.001) and "Secchi Depth" (DP1.20252.001). These data products are linked through the **siteID** field and local date in the NEON Data Publication Workbook for AOS Zooplankton Collection (AD[04]).

3.7 Product Instances

At each aquatic lake site, there will be up to 9 samples collected per year (3 samples per bout). Each sample generates multiple records from the external lab on a per taxon basis.

3.8 Data Relationships

For each record collected in zoo_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. Each **sampleID** record in zoo_fieldData may have multiple child records in zoo_taxonomyRaw and zoo_taxonomyProcessed, one record for each **scientificName**. A record from zoo_fieldData may have multiple or no records in zoo_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.

zoo_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 **samplerType**, number of tows or traps collected (**towsTrapsNumber**), and sampling depth (**zooDepth1**, **zooDepth2**, **zooDepth3**).

zoo_taxonomyRaw.csv - > One record is created for each taxonomic group identified in a sample created in zoo_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 table through the **sampleID**s in each table. Records in this table are unique by the combination of **sampleID**, **scientificName**, and **morphospeciesID**.

zoo_taxonomyProcessed.csv - > One record is created for each taxonomic group identified in a sample created in zoo_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 table through the **sampleID**s in each table. Records in this table are unique by the combination of **sampleID**, **scientificName**, and **morphospeciesID**.

zoo_perVial.csv - > One record is created for each taxonomic group removed from the final archive vial from the sample created in zoo_fieldData. 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 and taxonomy tables through the **sampleID**s in each table. Records in this table are unique by the combination of **sampleID** and **scientificName** or **morphospeciesID**. Individual organisms documented in zoo_taxonomy are returned to a single container or slide 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 zoo_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 zoo_taxonomy vial to be archived.

3.9 Special Considerations

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 the type of sampler used, which determines the volume of lake water sampled. The lab data includes subsampling information, taxonomic analysis, count and size data. Lab data values in **adjCountPerBottle** are corrected for subsampling in the lab, but are NOT corrected field sample volume (i.e., the volume of water sampled by the tow net or Schindler-Patalas trap). Data users will need to refer to the **towTrapsVolume** presented in the zoo_fieldData table and apply this correction to get the standardized density of organisms per liter. All taxon records from a sample should be summed and divided by the **towTrapsVolume** prior to reporting the total abundance per L.

$$zooplanktonAbundancePerL_i = \frac{\sum_{i=1}^{n} zoo\_taxonomyProcessed.adjCountPerBottle_i}{zoo\_fieldData.towTrapsVolume_i} \quad (1)$$

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

See the external lab SOP (**testProtocolVersion**) in the zoo_taxonomy tables for calculations applied to the data by the external laboratory.

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. The field data entry workflow for collecting zooplankton 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 Zooplankton Collection (NEON.DP0.20219) (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]).

4.2 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]).

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 table below for an explanation of **dataQF** codes specific to this product.

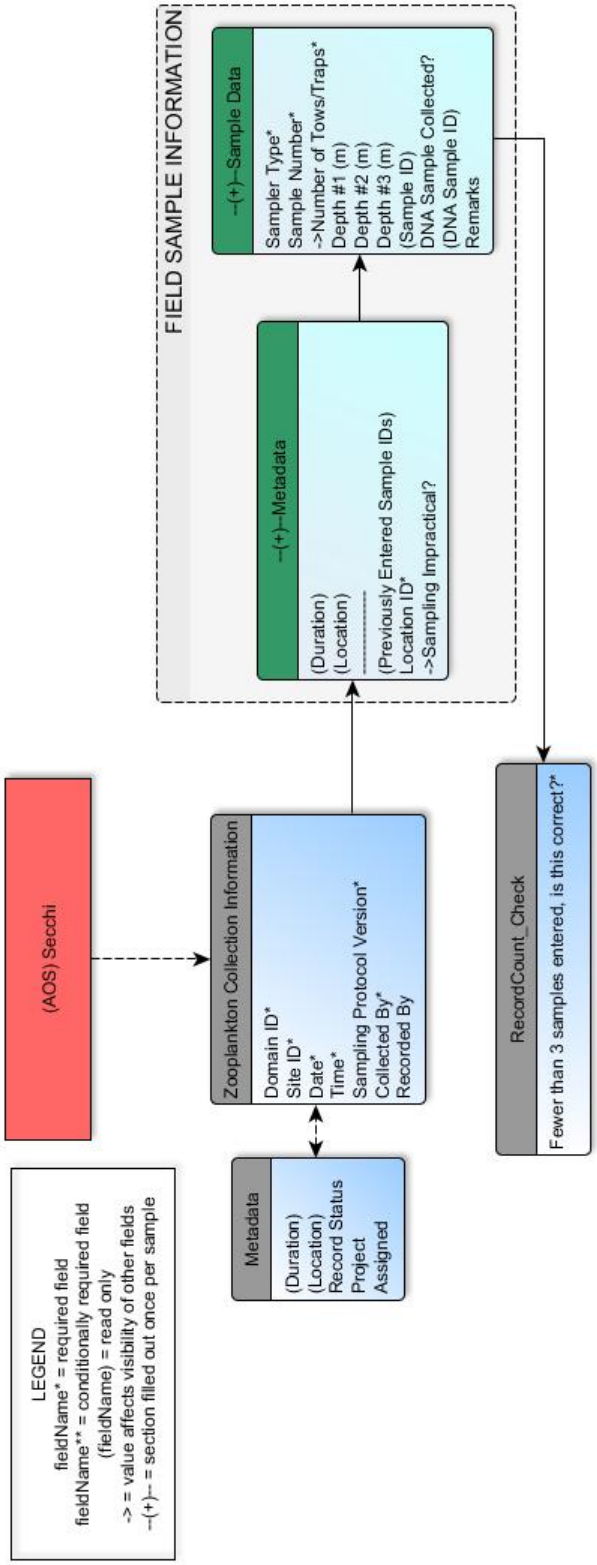

Figure 2: Schematic of the applications used by field technicians to enter zooplankton field data

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

4.5 Analytical Facility Data Quality

Data analyses conducted on zooplankton 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.

5 REFERENCES

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.

USEPA. 2012a. National Lakes Assessment Program, Field Operations Manual.

USEPA. 2012b. Sampling Procedures for the Great Lakes.