

<i>Title:</i> NEON User Guide to Periphyton and Phytoplankton Collection (NEON.DP1.20166)	<i>Date:</i> 02/12/2020
<i>Author:</i> Stephanie Parker	<i>Revision:</i> B

# NEON USER GUIDE TO PERIPHYTON AND PHYTOPLANKTON COLLECTION (NEON.DP1.20166)

<b>PREPARED BY</b>	<b>ORGANIZATION</b>	<b>DATE</b>
Stephanie Parker	AOS	02/12/2020
Tanya Vance	DPS	02/12/2020

<i>Title:</i> NEON User Guide to Periphyton and Phytoplankton Collection (NEON.DP1.20166)	<i>Date:</i> 02/12/2020
<i>Author:</i> Stephanie Parker	<i>Revision:</i> B

## CHANGE RECORD

<b>REVISION</b>	<b>DATE</b>	<b>DESCRIPTION OF CHANGE</b>
A	08/13/2017	Initial Release
B	01/07/2020	Sample design updates as a result of optimization, update Data Relationships

## TABLE OF CONTENTS

<b>1 DESCRIPTION</b>	<b>1</b>
1.1 Purpose . . . . .	1
1.2 Scope . . . . .	1
<b>2 RELATED DOCUMENTS AND ACRONYMS</b>	<b>2</b>
2.1 Associated Documents . . . . .	2
2.2 Acronyms . . . . .	2
<b>3 DATA PRODUCT DESCRIPTION</b>	<b>3</b>
3.1 Spatial Sampling Design . . . . .	3
3.2 Temporal Sampling Design . . . . .	5
3.3 Sampling Design Changes . . . . .	6
3.4 Variables Reported . . . . .	6
3.5 Spatial Resolution and Extent . . . . .	6
3.6 Sampling Design Changes . . . . .	7
3.7 Temporal Resolution and Extent . . . . .	7
3.8 Associated Data Streams . . . . .	7
3.9 Product Instances . . . . .	8
3.10 Data Relationships . . . . .	8
3.11 Special Considerations . . . . .	9
<b>4 DATA QUALITY</b>	<b>11</b>
4.1 Data Entry Constraint and Validation . . . . .	11
4.2 Automated Data Processing Steps . . . . .	11
4.3 Data Revision . . . . .	11
4.4 Quality Flagging . . . . .	11
4.5 Analytical Facility Data Quality . . . . .	11
<b>5 REFERENCES</b>	<b>14</b>

## LIST OF TABLES AND FIGURES

Figure 1 Generic aquatic site layouts for non-wadeable streams/rivers, wadeable streams, and lakes with periphyton and phytoplankton spatial sampling locations in red. In wadeable streams, benthic samples may be collected anywhere within the 1 km permitted reach (“algae sampling locations”), while the seston sample is collected at the S2 sensor location. . . . .	5
Figure 2 Schematic of the applications used by field technicians to enter periphyton and phytoplankton field data . . . . .	12
Figure 3 Schematic of the applications used by field technicians to enter domain lab subsampling and biomass data . . . . .	13

## 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 type of algal sample collected, 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 (AD[01]).

### 1.2 Scope

This document describes the steps needed to generate the L1 data product Periphyton and phytoplankton collection - the collection of periphyton using multiple benthic sampling methods in lakes, non-wadeable streams, and wadeable streams, as well as seston and phytoplankton collected in the water columns of wadeable streams (seston) and lakes or rivers (phytoplankton), 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 Periphyton and Phytoplankton Collection (NEON.DP1.20166) (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 for Periphyton and Phytoplankton Sampling (AD[06]). The raw data that are processed in this document are detailed in the file, NEON Raw Data Validation for AOS Periphyton, Seston, and Phytoplankton Collection, Level 0 (NEON.DP0.20166) (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., '10033') as the corresponding L1 data product.

## 2 RELATED DOCUMENTS AND ACRONYMS

### 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.DP0.20166.001_dataValidation.csv	NEON Raw Data Validation for AOS Periphyton, Seston, and Phytoplankton Collection, Level 0 (NEON.DP0.20166)
AD[04]	NEON.DP1.20166.001_variables.csv	NEON Data Variables for Periphyton and Phytoplankton Collection (NEON.DP1.20166)
AD[05]	NEON.DOC.001152	NEON Aquatic Sampling Strategy
AD[06]	NEON.DOC.003045	AOS Protocol and Procedure for Periphyton and Phytoplankton Sampling
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

### 2.2 Acronyms

Acronym	Definition
NAWQA	National Water Quality Assessment (USGS)
AFDM	Ash-free dry mass

### 3 DATA PRODUCT DESCRIPTION

This data product includes measurements for microalgae collection, including periphyton (benthic microalgae and biofilms, collected from a variety of benthic surfaces including cobble [epilithon], silt [epipelon], sand [epip-sammon], woody debris [epixylon], and plant surfaces[epiphyton]), seston (collected from the wadeable stream water column), and phytoplankton (collected from a lake or river water column). Periphyton and phytoplankton collection includes the field collection of microalgae (periphyton and phytoplankton) as well as ash-free dry mass (biomass), taxonomy, and abundance. Biovolume data were collected from 2014-2018, and discontinued in 2018. Algal chlorophyll, chemistry, and isotope data are related to this data product and can be found in “Periphyton and phytoplankton chemical properties” (NEON.DP1.20163; AD[02]). Starting in 2018, only chlorophyll data will be collected from seston samples in wadeable streams. Prior to 2018, seston data include taxonomy, biomass, and the full suite of chemistry data. These data are related to the NEON Grand Challenge areas of Biodiversity and Biogeochemistry and include additional data about the microalgal community in streams and lakes. These data can be used to assess the health of the aquatic ecosystem. Periphyton and phytoplankton 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). For additional information see sampling design NEON Aquatic Sampling Strategy (AD[05]) and protocol AOS Protocol and Procedure for Periphyton and Phytoplankton Sampling (AD[06]).

Microalgae samples are processed at the NEON Domain Support Facility and subsampled for “Periphyton and phytoplankton chemical properties” (NEON.DP1.20163; AD[02]) to be processed at an external facility, including chlorophyll, pheophytin, total carbon, total nitrogen, total phosphorus, 13C isotopes, 15N isotopes, and 34S isotopes. NEON personnel process a subsample for ash-free dry mass (AFDM) in the domain support facility. For epiphyte samples (microalgae from plant surfaces), domain personnel also estimate the plant surface area sampled as well as AFDM of the plant itself (**plantAdjAshFreeDryMass**). Finally, a subsample is also collected and preserved for taxonomic analysis at an external facility. Taxonomy samples are identified to the lowest practical taxon by a contracting taxonomist at an external facility.

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 and count data. 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 3.9 for suggested calculations.

Algae taxonomy data have a turnaround time of up to 6 months after collection. Field data will be available on the data portal before external lab data become available.

#### 3.1 Spatial Sampling Design

In wadeable streams, periphyton is 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 habitats present at each NEON Aquatic site (Figure 1). A minimum of three samples per habitat type are taken at each stream. All samples are collected from the surface of the natural substratum present in each macrohabitat. Field protocols differ depending on substratum being sampled. For example, riffles and runs often have cobble/pebble substratum, while pools may have silt or sand substrata. At some sites with sandy or silty bottoms, the major-

<i>Title:</i> NEON User Guide to Periphyton and Phytoplankton Collection (NEON.DP1.20166)	<i>Date:</i> 02/12/2020
<i>Author:</i> Stephanie Parker	<i>Revision:</i> B

ity of the periphyton community may be colonizing the leaves of aquatic plants (epiphytes) or woody debris at some sites, thus plant or woody debris substrata are sampled rather than sampling scarcely populated sandy/silty substrata. Appropriate site-specific sampling procedures are determined prior to sampling following NAWQA protocols (Moulton et al. 2002) and presented in site-specific AOS documents. See sampling design NEON Aquatic Sampling Strategy (AD[05]) and protocol AOS Protocol and Procedure for Periphyton and Phytoplankton Sampling (AD[06]) for additional details on strategy and SOPs.

In wadeable streams, non-wadeable streams, and lakes, seston and phytoplankton are collected from the water column at the water chemistry sampling locations (Figure 1). In wadeable streams, seston samples are collected near the S2 sensor in the thalweg of the stream. Starting in 2018, only chlorophyll is analyzed from seston samples in wadeable streams. In lakes, phytoplankton is collected at the central location (near the buoy) and the inlet and outlet sensor sets. In non-wadeable streams, phytoplankton is sampled near the sensor set (buoy), and from two other deep-water locations in the main channel. In lakes and non-wadeable streams, phytoplankton samples are composites of multiple depths depending on the depth of the euphotic zone and stratification. In a non-stratified system, the sample is composited from 1 surface sample, 1 sample near the bottom of the euphotic zone, and 1 mid-euphotic zone sample if the depth of the euphotic zone is >5m (see AD[06] for details). In a stratified system, the sample is composited from 1 surface sample, 1 sample in the metalimnion, and one sample near the bottom of the euphotic zone.

In wadeable streams, five periphyton samples are collected in the dominant habitat type and three samples are collected in the second-most dominant habitat type for a total of eight samples on a given sampling date at a site. Samples are spread out along the 1 km reach so that ideally no two samples are collected within the same habitat unit.

In lakes and non-wadeable streams, periphyton samples are collected in 5 of the 10 designated riparian sections following the divisions set forth in AOS Protocol and Procedure: Riparian Habitat Assessment in Lakes and Non-Wadeable Streams (NEON.DOC.001195). The most dominant substratum type in the littoral zone is chosen and samples are collected from each of five riparian sections. Field protocols differ depending on substrata being sampled and the sampler type used. Starting in 2019, subsamples are created for algal taxonomy from 3 of 5 parent samples. All 5 parent samples are still analyzed for AFDM, chlorophyll, and chemistry.

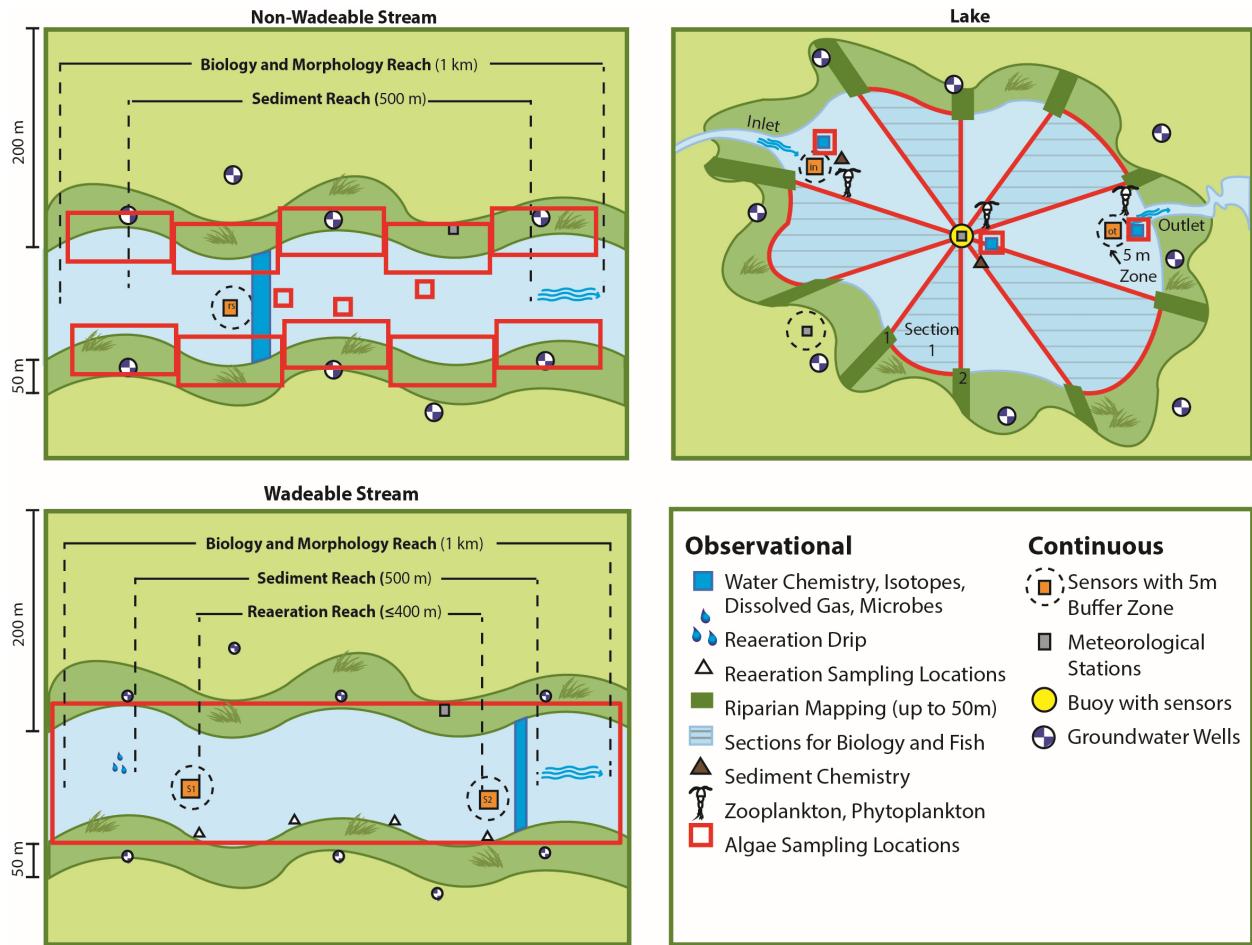


Figure 1: Generic aquatic site layouts for non-wadeable streams/rivers, wadeable streams, and lakes with periphyton and phytoplankton spatial sampling locations in red. In wadeable streams, benthic samples may be collected anywhere within the 1 km permitted reach ("algae sampling locations"), while the seston sample is collected at the S2 sensor location.

### 3.2 Temporal Sampling Design

Microalgae sampling occurs three times per year in order to capture presence and abundance of multiple species and growth forms. Timing of sampling is site-specific and determined based on historical data, including stream discharge, air temperature, and riparian greenness. Specific details on sample dates and strategy are provided in the NEON Aquatic Sample Strategy Document (AD[05]). Sample bout 1 is an early-season date, representing a period of rapid biomass accumulation after winter, typically after ice-off (where applicable) and prior to leaf out. Sample bout 2 targets low flows and high light (mid-summer) at each site. Sample bout 3 represents the late growing season (typically autumn) at each site during leaf-fall. These dates differ on a site-by-site basis but are always based on the same strategy. Sampling should occur at base-flow conditions, and will not occur directly fol-



lowing a flood in the stream (>1.5 x base flow; Biggs et al. 1999) or under ice. A period of 14-days is allowed after a flood event for periphyton to recolonize before sampling occurs. See NEON Aquatic Sampling Strategy (AD[05]), AOS Protocol and Procedure for Periphyton and Phytoplankton Sampling (AD[06]) for additional details.

### 3.3 Sampling Design Changes

- 2014-2017: Seston samples collected and analyzed for algal taxonomy, chlorophyll/pheophytin, AFDM, and chemistry.
- 2018-present: Seston samples are collected and analyzed for chlorophyll/pheophytin and AFDM only.
- 2014-2018: Algae taxonomy subsamples created from all parent samples.
- 2019-present: In lakes and rivers, algae taxonomy subsamples created from 3 phytoplankton samples and 3 littoral benthic samples. Five littoral benthic samples are still collected, and all are analyzed for chlorophyll/pheophytin, AFDM, and chemistry.
- 2014 - 2019: **processedDate** referred to the start date and time of all subsample processing in the domain support facility
- April 2019 - present: **processedDate** was made accurate as to the processing time of algal subsamples (e.g., processed date and time for taxonomy subsamples, chlorophyll filters, etc. )

### 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 AOS Periphyton, Seston, and Phytoplankton Collection, Level 0 (NEON.DP0.20166) (AD[03]). All variables reported in the published data (L1 data) are also provided separately in the file, NEON Data Variables for Periphyton and Phytoplankton Collection (NEON.DP1.20166) (AD[04]).

Field names have been standardized with Darwin Core terms (<http://rs.tdwg.org/dwc/>; accessed 12 August 2017), the Global Biodiversity Information Facility vocabularies (<http://rs.gbif.org/vocabulary/gbif/>; accessed 12 August 2017), the VegCore data dictionary (<https://projects.nceas.ucsb.edu/nceas/projects/bien/wiki/VegCore>; accessed 12 August 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.

### 3.5 Spatial Resolution and Extent

*Wadeable stream* Each periphyton sample will represent a patch of stream bottom within the 1 km permitted wadeable stream reach. The exact location of each sample will not be tracked as it is intended to represent the overall habitat (locationID = “reach”). Up to two different habitats are sampled at each site to account for the variability or “patchiness” among habitats. Field replicate samples are collected in each habitat, with five samples collected in the dominant habitat and three samples collected in the secondary habitat during each sampling bout (Figure 1).

<i>Title:</i> NEON User Guide to Periphyton and Phytoplankton Collection (NEON.DP1.20166)	<i>Date:</i> 02/12/2020
<i>Author:</i> Stephanie Parker	<i>Revision:</i> B

Each seston sample is collected in the water column near the wadeable stream S2 sensor location. Starting in 2018, seston samples are only analyzed for chlorophyll/pheophytin.

*Lake or River* Each periphyton sample will represent a patch of lake or river bottom in the littoral zone of the 1 km river permitted reach or the permitted littoral extent of the lake. Five samples are collected from the most dominant substratum type in the littoral zone.

Each phytoplankton sample represents a parcel of water from the water column in a lake or river. Three phytoplankton samples are collected at each lake or river site per bout. In a lake, one sample is collected at each of the following locations: the buoy (c0), inlet (in), and outlet (ot). In non-wadeable streams (rivers), one sample is collected near the sensor buoy (c0), and two other samples are collected in deep water locations to be determined by the field technician (locationID = reach). These samples do not require a fixed location.

Overall, this results in a spatial hierarchy of:

**habitatType** (habitat type sampled) → **locationID** (ID of the sampling location) → **siteID** (ID of NEON site) → **domainID** (ID of a NEON domain).

Note that some **sampleIDs** 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.

### 3.6 Sampling Design Changes

- 2014-2017: Seston algal taxonomy samples were processed through 2017. Algal taxonomy analysis from seston was halted due to sparse algal cells in the seston samples.
- 2014-2018: At lake and river sites, 5 littoral periphyton samples and 3 pelagic phytoplankton samples were analyzed for taxonomy.

### 3.7 Temporal Resolution and Extent

The finest temporal resolution that microalgae data are tracked is per sampling day. All 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>.

### 3.8 Associated Data Streams

All of the above data products 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” (NEON.DP1.20267). These data products are linked through the **siteID** field and local date in the NEON Data Variables for Periphyton and Phytoplankton Collection (AD[04]).

Phytoplankton data in lakes and rivers are related to data generated from AOS Secchi Depth and Depth Profiles (NEON.DOC.002792). These data products may be found in “Secchi depth” (NEON.DP1.20252) and “Depth profile at specific depths” (NEON.DP1.20254). These data products are linked through the **eventID** field in the NEON Data Publication Workbook for AOS Periphyton and Phytoplankton Collection (AD[04]).

Periphyton data in streams is related to data and samples generated for benthic microbes sampling in the following data products: “Benthic microbe community composition” (NEON.DP1.20086), “Benthic microbe group abundances” (NEON.DP1.20277), “Benthic microbe marker gene sequences” (NEON.DP1.20280), and “Benthic microbe metagenome sequences” (NEON.DP1.20279). Samples may be linked by **siteID**, **collectDate**, and **sampleNumber** or by parent **sampleID**.

Field data collected for of this data product are the parent samples to the algal chemistry data products found in “Periphyton and Phytoplankton Chemical Properties” (NEON.DP1.20163) and may be tracked using **parentSampleID** and **sampleID** between data products.

### 3.9 Product Instances

At each aquatic site, there are up to 27 parent samples collected per year (9 samples per bout) in wadeable streams, and 24 parent samples per year (8 per bout) in lakes and rivers.

Streams: Parent (field) seston samples generate 1 chlorophyll, 1 pheophytin record, and 1 AFDM record. Parent benthic samples generate 1 chlorophyll, 1 pheophytin record, and 1 AFDM record, as well as 1 each of carbon, nitrogen, carbon isotope, nitrogen isotope, and sulfur isotope records (NEON.DP1.20163). Each parent benthic samples generate multiple taxonomy records.

Lakes and rivers: Parent phytoplankton records and 3 parent benthic littoral records generate 1 chlorophyll, 1 pheophytin record, and 1 AFDM record, as well as 1 each of carbon, nitrogen, carbon isotope, nitrogen isotope, and sulfur isotope records (NEON.DP1.20163). Each parent benthic samples generate multiple taxonomy records. Two benthic littoral samples generate 1 chlorophyll, 1 pheophytin record, and 1 AFDM record, as well as 1 each of carbon, nitrogen, carbon isotope, nitrogen isotope, and sulfur isotope records (no taxonomy records).

- 2014 - 2017: Seston samples generate a full suite of records (taxonomy, chlorophyll/pheophytin, AFDM, and chemistry)
- 2014 - 2018: All 5 lake/river benthic littoral samples generate taxonomy records

### 3.10 Data Relationships

For each record that is collected in the field (**alg\_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 sampling is practical, there may be from 1 to 20 child records from a single **parentSampleID**. Child records will be found in **alg\_biomass** (initial subsampling information, including filter volumes, AFDM measurement, and taxonomy preservation), **alg\_taxonomyRaw** (raw external lab taxonomic data), and **alg\_taxonomyProcessed** (processed taxonomic data). Every record in **alg\_biomass**, **alg\_taxonomyRaw**, and **alg\_taxonomyProcessed** should have a corresponding record in **alg\_fieldData** describing field collection conditions, location, and metadata during sample collection. If **parentSampleID** is empty in **alg\_fieldData**, there will be no additional records in the **alg\_biomass**, **alg\_taxonomyRaw**, or **alg\_taxonomyProcessed**. There is one unique record for each **parentSampleID** - **sampleID** - **analysisType** - **filterNumber** combination for child records, and **alg\_taxonomyRaw** and **alg\_taxonomyProcessed**

will have multiple records per **sampleID** organized by **scientificName** and **algalParameter**. 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.

alg\_fieldData.csv - > One record is created for each sample collected in the field, creating a **parentSampleID** which is linked to all subsequent tables (except alg\_biovolume). This table also indicates the field conditions, including **habitatType**, **algalSampleType**, **substratumSizeClass**, and sample depth if applicable (e.g., lake and non-wadeable sites).

alg\_biomass.csv - > One record (**sampleID**) is created for each sample (**parentSampleID**) processed at the NEON domain lab. A sample aliquot is filtered onto a glass-fiber filter and dried and weighed to calculate **adjAsh-FreeDryMass**. Data from this table are linked to the fieldData, subsequent taxon data, and chemistry data in NEON.DP1.20163 through the **sampleIDs** in each table. Additionally, measurements for plant material sampled for epiphyton are presented here, including **plantSurfaceArea** and **plantAdjAshFreeDryMass**. Note that the field **fieldSampleVolume** presented in this table is corrected for gain or loss of water in the sample for epipsammon and epipelon, and is populated with the volume of water used to process epiphyton samples. For epilithon and epixylon,  $alg\_biomass.fieldSampleVolume == alg\_fieldData.fieldSampleVolume$ .

alg\_taxonomyRaw.csv - > One record is created for each taxonomic group identified in a sample created in alg\_fieldData. Taxonomic identifications are made to the lowest practical taxonomic level (typically 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 biomass tables through the **sampleIDs** in each table.

alg\_taxonomyProcessed.csv - > One record is created for each taxonomic group identified in a sample created in alg\_fieldData. Taxonomic identifications are made to the lowest practical taxonomic level (typically species). The taxonomic nomenclature in this file has been standardized and desynonymized according to NEON's master taxonomy for algae. Data are linked to the fieldData and biomass tables through the **sampleIDs** in each table.

alg\_archive.csv - > One record is created for each item deposited at the bioarchive. Glass slides, preserved soft algae samples in vials, and freeze-dried diatom subsamples are recorded here and can be linked to the biomass or taxonomy tables by **sampleID**.

alg\_biovolume.csv - > Biovolume data are reported by the external lab annually through 2018 on a per-taxon basis, and are not separated by site. These records are not directly related at the sample level, but may be linked to collection and taxonomy data through **scientificName**. One record is created for each site per year for each **scientificName** that is measured for biovolume analysis. Data in this file represent a combination of measurements from NEON samples and algae samples from other projects analyzed at the external lab.

### 3.11 Special Considerations

The periphyton and phytoplankton taxonomic counts come from an external lab, in the field **algalParameter-Value**. This field should be coupled with **algalParameterUnit**. The taxonomic count data with **algalParameterUnit = countPerBottle** are not corrected for the volume of preservative in the bottle or for benthic area. Data users will need to refer to the **benthicArea** presented in the alg\_fieldData table and apply this correction to get the number of algal cells per 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<sup>2</sup>.

*Step 1: Convert to mL and correct for preservative volume When **algalParameterUnit = countPerBottle***

- using the sample volume and preservative volume measured at the NEON Domain Support Facility (prior to shipping):

$$algalCountPerMl_i = \frac{alg\_taxonomyProcessed.algalParameterValue_i}{alg\_biomass.labSampleVolume_i + alg\_biomass.preservativeVolume_i} \quad (1)$$

- or use the lab-measured sample volume in the bottle (may be missing for some legacyData):

$$algalCountPerMl_i = \frac{alg\_taxonomyProcessed.algalParameterValue_i}{alg\_taxonomyProcessed.perBottleSampleVolume_i} \quad (2)$$

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

Step 2: Correcting for benthicArea Using algalCountPerMl from Step 1:

$$algalAbundancePerM_i^2 = \sum_{i=1}^n \frac{alg\_taxonomyProcessed.algalCountPerMl_i \times alg\_biomass.fieldSampleVolume}{alg\_fieldData.benthicArea_i} \quad (3)$$

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

See the external lab SOP (referenced in **method** in alg\_taxonomyProcessed and alg\_taxonomyRaw) for calculations applied to the data by the external laboratory.

Additionally, taxonomy data are repeated in different types of counts in the alg\_taxonomyProcessed and alg\_taxonomyRaw tables based on **algalParameter**, depending on how the data user wants to process data. NEON suggests using "cell density" for the calculations above. See below for descriptions.

algalParameter	algalParameterUnit	description
number of valves	count	Number of diatom valves in analyzed subsample
number of natural units	count	Number of natural units (cells or colonies of cells) in analyzed subsample
number of cells	count	Number of cells in analyzed subsample
cell density	cellsPerBottle	Number of cells in the sample bottle (not corrected for preservative volume)
biovolume density	cubicMicrometersPerBottle	Biovolume of cells in the sample bottle (not corrected for preservative volume)*
subjective rating	none	Rating indicating relative density in macroalgae

\*biovolume data collection was discontinued at the end of the 2018 field season

## 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 aquatic periphyton and phytoplankton data is diagrammed in Figure 2, and the domain lab workflow is diagrammed in Figure 3. 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 AOS Periphyton, Seston, and Phytoplankton Collection, Level 0 (NEON.DP0.20166) (AD[03]), and 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 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[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.

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 periphyton and phytoplankton community data conform to the current data quality standards used by practitioners. Ten percent of all samples are quality checked for taxonomic difference between

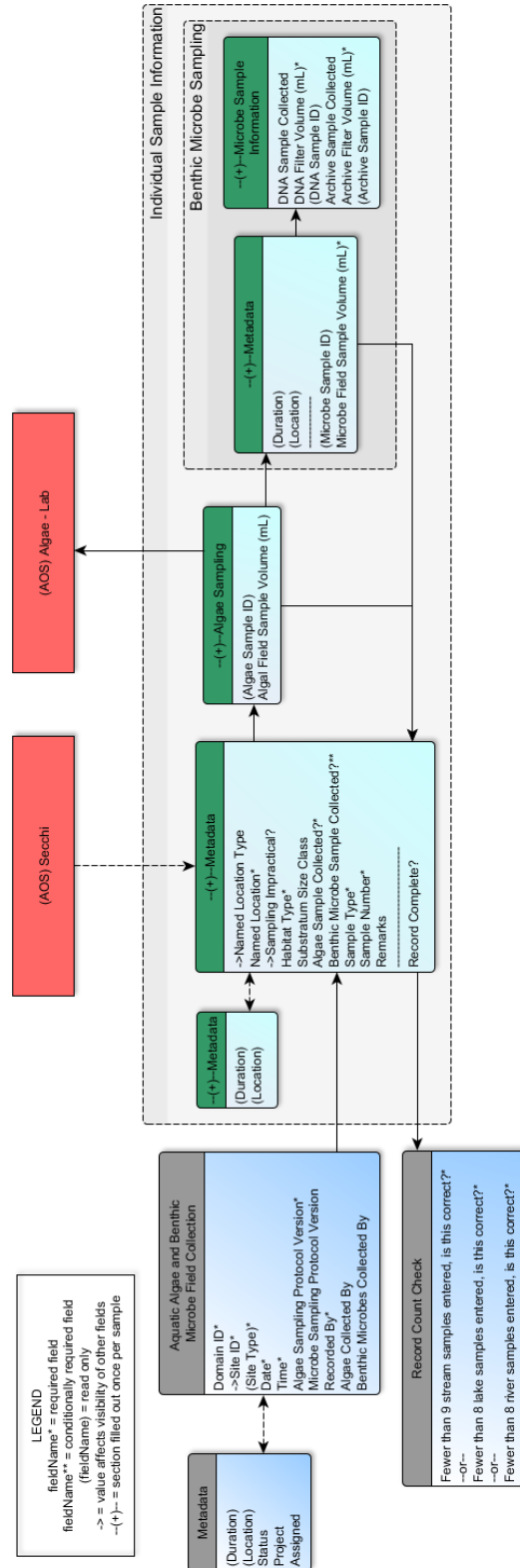


Figure 2: Schematic of the applications used by field technicians to enter periphyton and phytoplankton field data

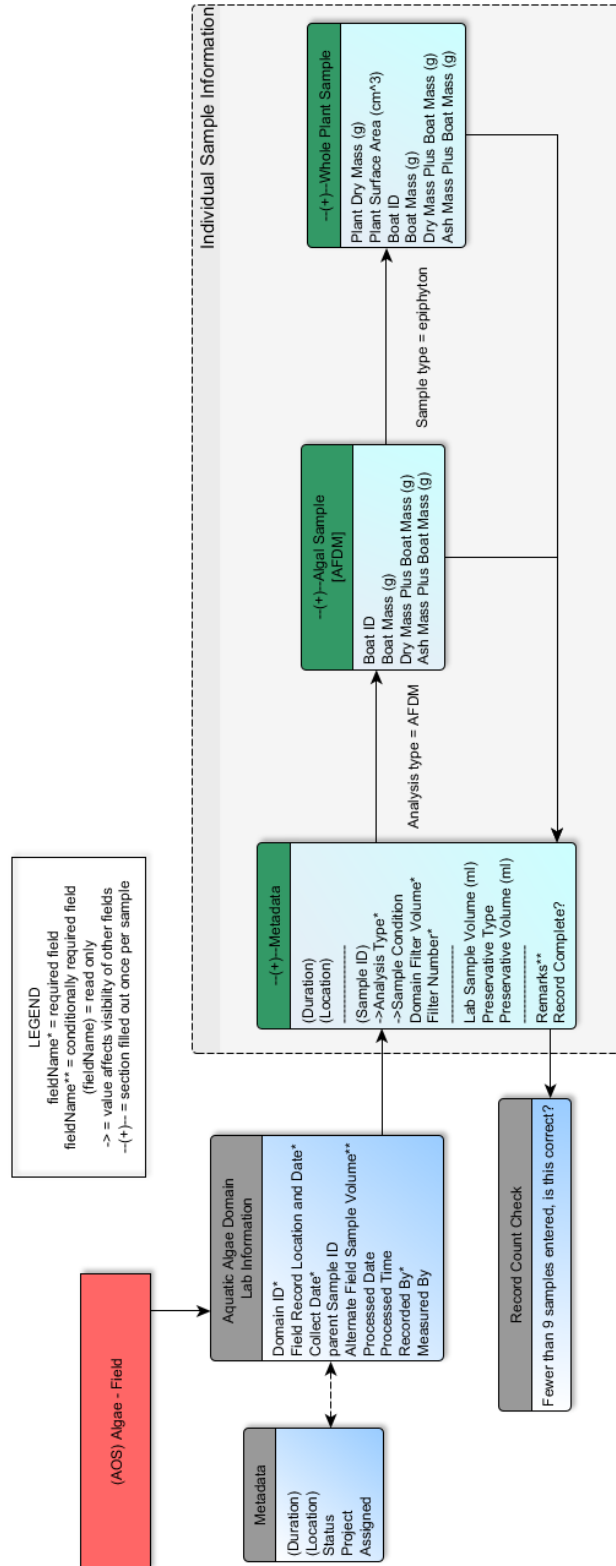


Figure 3: Schematic of the applications used by field technicians to enter domain lab subsampling and biomass data



<i>Title:</i> NEON User Guide to Periphyton and Phytoplankton Collection (NEON.DP1.20166)	<i>Date:</i> 02/12/2020
<i>Author:</i> Stephanie Parker	<i>Revision:</i> B

two taxonomists at the external facility. These records are indicated by the field **qcTaxonomyStatus**, indicating whether the sample was chosen randomly for a taxonomic quality check. Details on the quality control of algal taxonomy can be found in the external lab SOP.

## 5 REFERENCES

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.

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.