NEON USER GUIDE TO PERIPHYTON AND PHYTOPLANKTON COLLECTION (DP1.20166.001)

PREPARED BY	**ORGANIZATION**
Stephanie Parker	AOS
Tanya Vance	DPS

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	08/13/2017	Initial Release
B	01/07/2020	Sample design updates as a result of optimization, update Data Relationships
C	10/26/2020	Included general statement about usage of neonUtilities R package and statement about possible location changes. Updated taxonomy information. More sample design updates and lab data quality clarifications. Updated lake littoral sensor locations and figures.
C.1	9/22/2021	Adding calculation information for valves from diatom slides

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, 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, rivers, and wadeable streams, as well as phytoplankton collected in the water columns of lakes or rivers 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 (DP1.20166.001) (AD[04]), provided in the download package for this data product.

This document describes the process for ingesting and performing automated quality assurance and control procedures on the data collected in the field pertaining to AOS Protocol and Procedure 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 (DP0.20166.001) (AD[03]), provided in the download package for this data product. Please note that raw data products (denoted by 'DP0') may not always have the same numbers (e.g., '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 Data Products Catalog
AD[03]	Available with data download	Validation csv
AD[04]	Available with data download	Variables csv
AD[05]	NEON.DOC.001152	NEON Aquatic Sampling Strategy
AD[06]	NEON.DOC.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]	NEON.DOC.004825	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[10]	Available on NEON data portal	NEON Ingest Conversion Language Function Library
AD[11]	Available on NEON data portal	NEON Ingest Conversion Language
AD[12]	Available with data download	Categorical Codes csv

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 periphyton and phytoplankton collection, with periphyton collected from a variety of benthic surfaces including cobble [epilithon], silt [epipelon], sand [epipsammon], 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 algal biomass (ash-free dry mass), taxonomy, and abundance or cell density. Algal chlorophyll, chemistry, and isotope data are related to this data product and can be found in "Periphyton and phytoplankton chemical properties" (DP1.20163.001; 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, which can be used to assess the health of the aquatic ecosystem. Periphyton and phytoplankton samples are collected 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]).

Periphyton and phytoplankton samples are processed at the NEON Domain Support Facility and subsampled for taxonomic analysis and preserved and processed for ash-free dry mass (AFDM). Subsample for "Periphyton and phytoplankton chemical properties" (DP1.20163.001; AD[02]) to be processed at at this time, including chlorophyll, pheophytin, total carbon, total nitrogen, total phosphorus, 13C isotopes, 15N isotopes, and 34S isotopes. For epiphyte samples (microalgae from plant surfaces), domain personnel estimate the plant surface area sampled as well as AFDM of the plant itself (**plantAdjAshFreeDryMass**). Preserved taxonomy subsamples are shipped to an external taxonomy facility, identified to the lowest practical taxon, and enumerated. Permanent slides and preserved soft algae samples are also prepared for storage and curations at the NEON Biorepository.

Data are organized into tables for field data collected by NEON field ecologists and external lab data are 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 for periphyton. The lab data include subsampling information, taxonomic analysis, count data, and quality metrics. Lab data are corrected for subsampling (cellsPerBottle), however the data user must use both the lab data and the field data to calculate counts per mL or 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 5 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 collected at each stream (up to 6 samples total per site per sampling day). 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 majority of the periphyton community may be colonizing the leaves and stems 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. Samples are spread out along the 1 km reach so that ideally no two samples are collected within the same habitat unit.

In rivers and lakes, periphyton samples are collected from three benthic littoral locations selected from the 10 riparian divisions set forth in AOS Protocol and Procedure: Riparian Habitat Assessment (NEON.DOC.003826). The most dominant substratum type in the littoral zone is chosen and as with streams, field protocols differ depending on substrata being sampled. Phytoplankton are collected from the water column near the water chemistry sampling locations and sensor infrastructure (Figure 1). In lakes, phytoplankton is collected at the central location (near the buoy) and the two littoral sensor sets. In rivers, phytoplankton is sampled near the sensor set (buoy), and from two other deep-water locations in the main channel. Phytoplankton samples are composites of multiple depths depending on the depth of the euphotic zone and stratification at the buoy location 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.

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

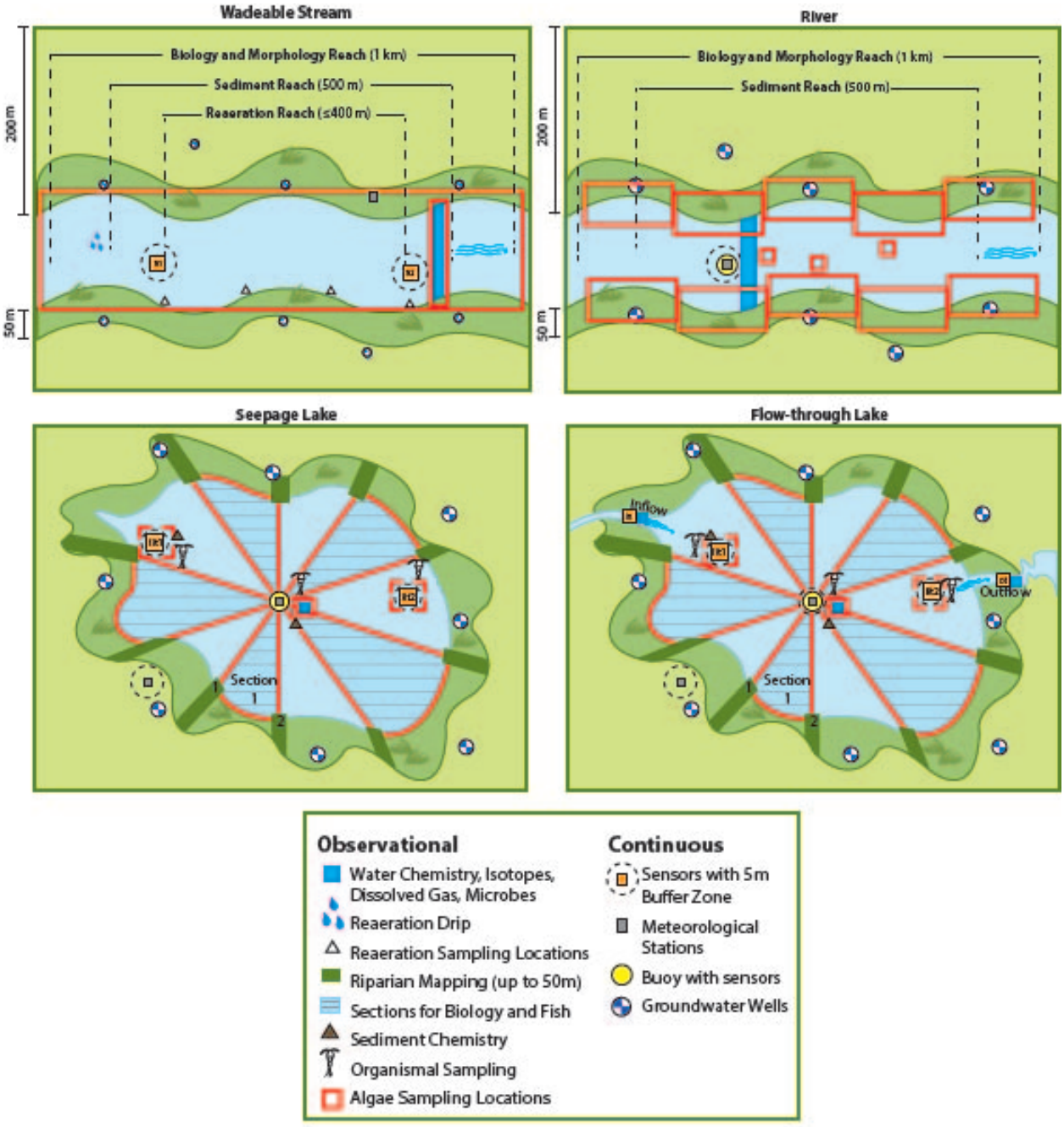

Figure 1: Generic aquatic site layouts for wadeable streams, rivers, and lakes with periphyton and phytoplankton spatial sampling locations in red. In wadeable streams, benthic samples may be collected anywhere within the 1 km permitted reach.

3.2 Temporal Sampling Design

Periphyton and phytoplankton 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 following 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. Due to low cell density, taxonomy was discontinued at the beginning of 2018 in seston samples.
- 2018-present: Seston samples are collected and analyzed for chlorophyll/pheophytin and AFDM only to provide supporting data for the chlorophyll sensor.
- 2014-2018: Algae taxonomy subsamples created from all parent benthic samples (8 samples per site in streams, 5 samples per site in lakes/rivers).
- 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. This was a result of optimization analysis that showed that 3 samples per habitat type are sufficient for diversity and abundance metrics.
- 2019-present: In wadeable streams, algae taxonomy subsamples created from 8 benthic samples, however only 6 samples are sent to the external taxonomist (3 from the dominant habitat and 3 from the subdominant habitat). This follows the optimization analysis for lakes and rivers above, optimization analyses for wadeable streams are in process to justify this change.
- 2014-2018: Biovolume data were collected and reported. Due to low interest and availability of metrics in the literature, these data were discontinues.
- 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 (DP0.20166.001) (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 (DP1.20166.001) (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 three samples collected in the dominant habitat and three samples collected in the secondary habitat during each sampling bout (Figure 1).

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 represents 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. Three 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 the the buoy (c0 or c1) and the two littoral sensor sets (lit1, lit2). In 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 **parentSampleID**s and **sampleID**s in the legacyData may start with "ST" (i.e., STMA = MAYF, STWA = WALK, STCU = CUPE, STKG = KING). These samples were collected in the STREON reach (the downstream 500 m of the 1 km reach) prior to descoping of the STREON experiment. Data from these samples may be used just as any other sample from the site.

3.6 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.

3.7 Associated Data Streams

All of the above data products are loosely related to Aquatic General Field Metadata collected on the same sampling day (NEON.DOC.001646). Data for Aquatic General Field Metadata are available in the NEON data product "Gauge Height" (DP1.20267.001). These data products are linked through the **siteID** field and local date in the NEON Data 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" (DP1.20252.001) and "Depth profile at specific depths" (DP1.20254.001). 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" (DP1.20086.001), "Benthic microbe group abundances" (DP1.20277.001), "Benthic microbe marker gene sequences" (DP1.20280.001), and "Benthic microbe metagenome sequences" (DP1.20279.001). Samples may be linked by **parentSampleID**.

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" (DP1.20163.001) and may be tracked using **parentSampleID** between data products.

3.8 Product Instances

At each aquatic site, there are up to 18 parent samples collected per year (6 samples per bout). Seston samples generate 1 AFDM record in this data product. Parent benthic and phytoplankton samples generate 1 AFDM record as well as multiple taxonomy records. Additional chemistry records generated by the parent samples can be found in (DP1.20163.001).

3.9 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 1 record created per missed sample in alg_fieldData and no downstream data populated in the tables below. If sampling is practical, there will be 2 child records from a single **parentSampleID** in alg_biomass. Multiple records per **sampleID** (taxonomy subsample) will be entered in 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. There is one unique record for each **parentSampleID** - **sampleID** 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 alg_biomass. For instances where sampling is impractical, one **samplingImrpactical** record is created per missed sample. This table also indicates the field conditions, including **habitatType**, **algalSampleType**, **substratumSizeClass**, and sample depth in lake and river 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 **adjAshFreeDryMass**. A second subsample for algal taxonomy is created and preserved. Data from this table are linked to the fieldData and subsequent taxon data. 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_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 biomass table through the **sampleID**s in each table. Quantitative data are returned as **algalParameterUnit** = *cellsPerBottle*.

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 **sampleID**s in each table. Quantitative data are returned as **algalParameterUnit** = *cellsPerbottle*.

alg_qualityCheck.csv - > Records are created in this table for the 10% of samples selected for quality analysis. Two records are expected per taxon sample, one for the diatom slide workflow and one for the soft algae workflow. Data can be linked to alg_taxonomyProcessed and alg_taxonomyRaw through the **sampleID** field.

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.

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

3.10 Special Considerations

The periphyton and phytoplankton taxonomic counts come from an external lab in the field **algalParameterValue**. This field should be coupled with **algalParameterUnit**.

Algal taxonomy data are analyzed in two ways: diatom valves are identified from permanent slides, and soft algae + diatom cells are identified in a counting chamber. Diatom valves can provide data on relative abundance (**algalAnalysisMethod** = “diatom slide”) and typically include data identified to the species level. Soft algae + diatom data (**algalAnalysisMethod** = “soft algae”) are quantitative and can be used to determine cells per square meter (periphyton) or cells per L (phytoplankton), however data may not consistently achieve the taxonomic level of species. Suggested calculations are provided below.

Diatom slides

1. Subset by **algalAnalysisMethod** = “diatom slide” or **algalParameter** = “valves”
2. Suggestion 1: Use species data in relative abundance calculations
3. Suggestion 2: Calculate the relative proportion of diatoms belonging to each species or group of interest following Stevenson and Bahls (1999).

a. Example: total valves counted in Sample 1 = 600, 300 of which are *A. minutissimum*. The relative proportion of *A. minutissimum* in the sample = 300/600, or 0.5.
b. Using the quantitative counts below in Step 3, replace the diatom counts with the relative proportions from the diatom slides. For example, if 100 of 300 total cells in the soft algae + diatom count are diatoms, then 50 of them would be *A. minutissimum*, and the remaining 50 diatoms cells would be split by the relative proportions of the other diatom species in the slide.

Soft algae + diatom quantitative

These data are not corrected for preservative volume or 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^2.

1. Subset by **algalParameterUnit** = "cellsPerBottle".
2. Convert to mL and correct for preservative volume - stop here to calculate phytoplankton density (cells per volume), continue to Step 3 to calculate periphyton density (cells per benthic area).

When **algalParameterUnit** = *cellsPerBottle*

- 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**

3. Correcting for benthicArea in periphyton samples

Using algalCountPerMl from Step 2:

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

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.

Table 1: Descriptions of the algalParamater counts

algalParameter	algalParameterUnit	description
valves	count	Number of diatom valves in analyzed subsample
natural units	count	Number of natural units (cells or colonies of cells) in analyzed subsample
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

Biomass calculations

Step 1: Phytoplankton and seston ash-free dry mass (g/mL)

$$ashFreDryMassPerMl_i = \frac{alg\_biomass.adjAshFreeDryMass_i}{alg\_biomass.domainFilterVolume} \quad (4)$$

Where *i* = **sampleID** + **analyte**

Step : Benthic periphyton ash-free dry mass (g/m^2)

$$ashFreDryMassPerM_i^2 = \frac{alg\_biomass.adjAshFreeDryMass_i}{alg\_biomass.domainFilterVolume} \times \frac{alg\_biomass.fieldSampleVolume}{alg\_fieldData.benthicArea_i} \quad (5)$$

Where *i* = **sampleID** + **analyte**

4 TAXONOMY

NEON manages taxonomic entries by maintaining a master taxonomy list based on the community standard, if one exists. Through the master taxonomy list, synonyms submitted in the data are converted to the appropriate name in use by the standard. The master taxonomy for algae was originally based on the USGS taxonomy lists maintained by the Patrick Center for Environmental Research Phycology Section at Academy of Natural Sciences of Drexel University (Academy of Natural Sciences of Drexel University and collaborators, 2011 – 2016). Prior to 2020, unique Taxon ID codes were generated by concatenating the string 'NEONDREX' with the North American Diatom Ecological Database (NADED) ID for taxa identified to

species, including non-diatom algae. Taxon IDs for taxa identified to the genus level or higher used a concatenated string including the taxon name (e.g., genus name or family name) and “sp” or “spp”. Starting in January 2020, unique Taxon ID codes used to identify taxonomic concepts in the NEON master taxonomy list are 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 diatom, soft algae, and macroalgae taxa. NEON plans to keep the taxonomy updated in accordance with Diatoms of North America (diatoms.org), the USGS BioData Database, and other current literature starting in 2020 and annually thereafter.

The master taxonomy list also indicates the expected geographic distribution for each species by NEON domain and whether it is known to be introduced or native in that part of the range. Given that reported spatial distributions for many algal taxa are unknown or have low precision, NEON assumes that all taxa are possible at all aquatic sites. As spatial resolution improves, NEON will update these taxon tables to generate errors if a taxon is reported at a location outside of its known range.

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

5 DATA QUALITY

5.1 Data Entry Constraint and Validation

Many quality control measures are implemented at the point of data entry within a mobile data entry application or web user interface (UI). For example, data formats are constrained and data values controlled through the provision of dropdown options, which reduces the number of processing steps necessary to prepare the raw data for publication. The field data entry workflow for collecting 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 (DP0.20166.001) (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]).

5.2 Automated Data Processing Steps

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

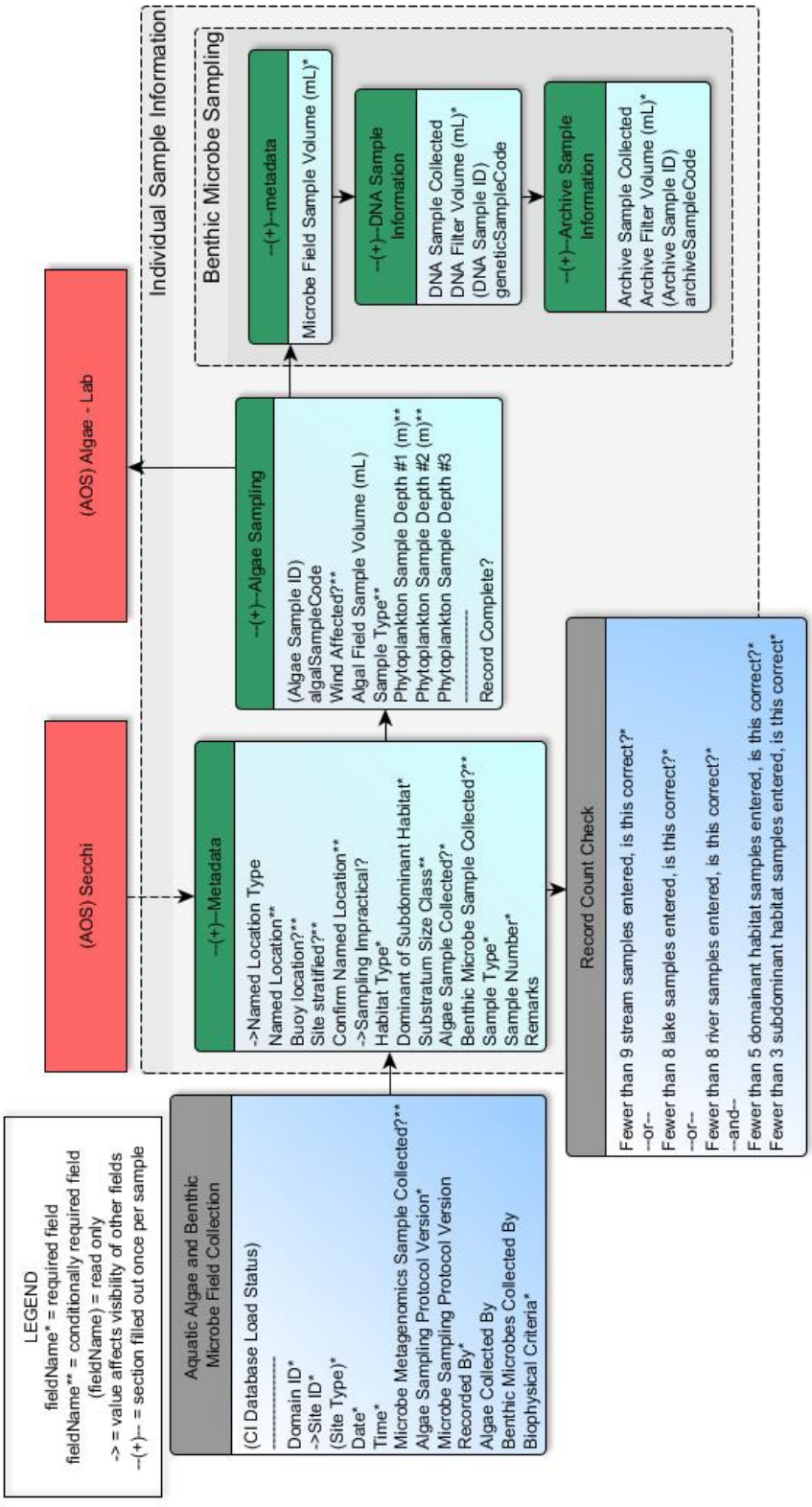

Figure 2: Schematic of the applications used by field technicians to enter periphyton and phytoplankton field data

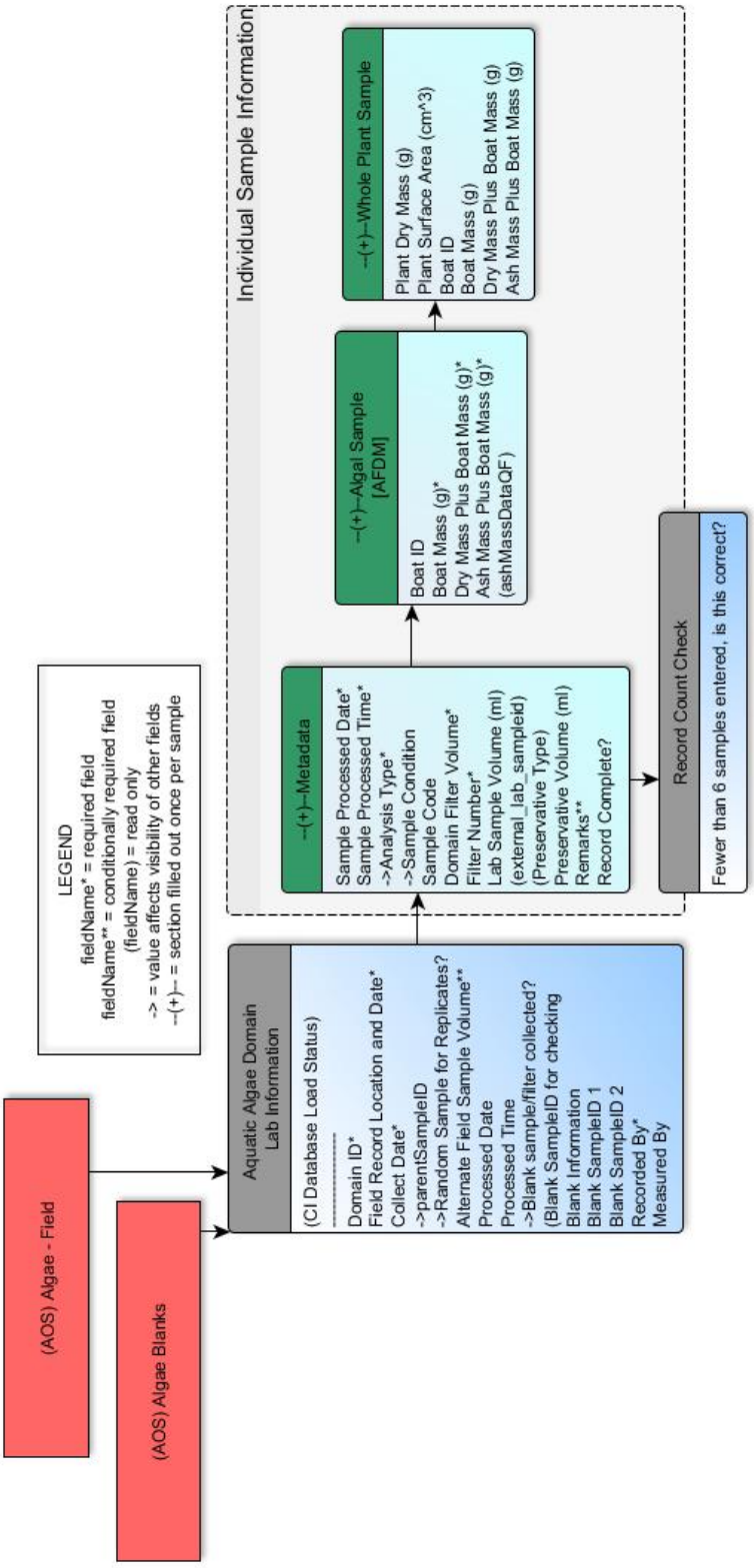

Figure 3: Schematic of the applications used by field technicians to enter domain lab subsampling and biomass data

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

Table 2: Descriptions of the dataQF codes for quality flagging

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

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

5.5 Analytical Facility Data Quality

Data analyses conducted on 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 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. Percent similarity (**qcPercentSimilarity**) and percent difference in enumerate (**enumerationDifference**) are reported in the alg_qualityCheck table. More details on the quality control of algal taxonomy can be found in the external lab SOP.

6 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.

Stevenson, R. J. and L. L. Bahls, 1999. Chapter 6: Periphyton Protocols. Pages 6-1 through 6-22 in M.T. Barbour, J. Gerritsen, B.D. Snyder, and J.B. Stribling (eds.). Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish, 2nd ed., Report EPA 841-B-99-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC.