NEON USER GUIDE TO PERIPHYTON AND PHYTOPLANKTON CHEMICAL PROPERTIES (DP1.20163.001)

PREPARED BY	ORGANIZATION
Stephanie Parker	AOS
Tanya Vance	DPS
Caren Scott	AOS
Zachary Nickerson	AOS

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	08/15/2017	Initial Release
B	10/22/2020	Included general statement about usage of neonUtilities R package and statement about possible location changes. Add data quality tables, update spatial figure and lake littoral locations.
B.1	06/01/2021	Added quality flagging choices for external laboratory data.
B.2	08/17/2021	Edited units in equation 2.
C	02/08/2022	Updated section 4.3 Data Revision with latest information regarding data release
D	10/11/2022	Added information about the decarbonation via acid digestion SOP developed by the external lab for the carbon (C) isotope samples
E	03/25/2024	Update with new information about composite sampling and new tables
F	04/17/2025	Update equations to include conversion from mL to L in fieldSampleVolume. Added information about the new neonUtilities Python package.
G	04/06/2026	Add alg_archive table to this data product. Updated spatial and Fulcrum figures, standard updates for Biorepository and automated QC.
G.1	09/08/2026	Add alg_blank table to the published data product.
G.2	10/14/2026	Adding averaged data from 2014 to 2023 to the alg_domainLabChemistryComp, alg_algaeDataPerSampleCompChl, and alg_algaeDataPerSampleCompCNPS tables. Adding the analyteSampleID for archived metabarcode samples to the alg_domainLabChemistryComp table.

TABLE OF CONTENTS

1 DESCRIPTION **1**
1.1 Purpose . . . 1
1.2 Scope . . . 1

2 RELATED DOCUMENTS AND ACRONYMS **2**
2.1 Associated Documents . . . 2
2.2 Acronyms . . . 2

3 DATA PRODUCT DESCRIPTION **3**
3.1 Spatial Sampling Design . . . 3
3.2 Temporal Sampling Design . . . 6
3.3 Sampling Design Changes . . . 6
3.4 Laboratory Quality Assurance and Uncertainty . . . 7
3.5 Variables Reported . . . 9
3.6 Spatial Resolution and Extent . . . 9
3.7 Temporal Resolution and Extent . . . 10
3.8 Associated Data Streams . . . 10
3.9 Product Instances . . . 10
3.10 Data Relationships . . . 10
3.11 Special Considerations . . . 12

4 DATA QUALITY **14**
4.1 Data Entry Constraint and Validation . . . 14
4.2 Automated Data Processing Steps . . . 14
4.3 Data Revision . . . 14
4.4 Quality Flagging . . . 14
4.5 Analytical Facility Data Quality . . . 17

5 REFERENCES **17**

LIST OF TABLES AND FIGURES

Table 1 Acid treatment status for each NEON Aquatic site . . . 8
Table 2 Descriptions of the dataQF and externalLabDataQF codes for quality flagging . . . 17

Figure 1 Generic aquatic site layouts for wadeable streams, rivers, and lakes with periphyton, seston, and phytoplankton spatial sampling locations in red. . . . 5
Figure 2 Schematic of the applications used by field technicians to enter periphyton and phytoplankton field data . . . 15
Figure 3 Schematic of the applications used by field technicians to enter domain lab subsampling and biomass data . . . 16

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, for example, the concentration of an analyte, 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 Periphyton and phytoplankton chemical properties - the carbon, nitrogen, phosphorus, stable isotopes, chlorophyll *a*, and pheophytin of microalgae from water column and benthic samples in lakes, rivers, and wadeable streams 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 Chemical Properties (DP1.20163.001) (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 for Periphyton and Phytoplankton Sampling (AD[07]). The raw data that are processed in this document are detailed in the file, NEON Raw Data Validation for AOS Periphyton and Phytoplankton Collection, Level 0 (DP0.20166.001) (AD[03]) and NEON Raw Data Validation for Plant and Algae External Lab Chemistry (DP0.20065.001) (AD[04]), provided in the download package for this data product. Please note that raw data products (denoted by 'DP0') may not always have the same numbers (e.g., '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	NEON Raw Data Validation for AOS Periphyton and Phytoplankton Collection, Level 0 (DP0.20166.001)
AD[04]	Available with data download	NEON Raw Data Validation for Plant and Algae External Lab Chemistry (DP0.20065.001)
AD[05]	Available with data download	NEON Data Variables for Periphyton and Phytoplankton Chemical Properties (DP1.20163.001)
AD[06]	NEON.DOC.001152	NEON Aquatic Sampling Strategy
AD[07]	NEON.DOC.003045	AOS Protocol and Procedure for Periphyton and Phytoplankton Sampling
AD[08]	NEON.DOC.000008	NEON Acronym List
AD[09]	NEON.DOC.000243	NEON Glossary of Terms
AD[10]	NEON.DOC.004825	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[11]	Available on NEON data portal	NEON Ingest Conversion Language Function Library
AD[12]	Available with data download	Categorical Codes csv
AD[13]	NEON.DOC.005424	NEON Algorithm Theoretical Basis Document: OS Data Quality Control

2.2 Acronyms

Acronym	Definition
NAWQA	National Water Quality Assessment (USGS)

3 DATA PRODUCT DESCRIPTION

The aquatic periphyton and phytoplankton chemistry data product provides chlorophyll *a*, ash-free dry mass, pheophytin, total carbon, total nitrogen, total phosphorus, δ^{13}C, δ^{15}N and δ^{34}S of the bulk algal community. Microalgae are sampled three times per year at each NEON aquatic site (AD[07]); periphyton are sampled at all sites, while seston are sampled in wadeable streams and phytoplankton are sampled in lakes and rivers. Starting in 2018, seston samples are only analyzed for chlorophyll and pheophytin. 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[06]) and protocol AOS Protocol and Procedure for Periphyton and Phytoplankton Sampling (AD[07]). Starting in 2018, only chlorophyll, pheophytin, and ash-free dry mass data are collected from seston samples in wadeable streams. Prior to 2018, seston records include the full suite of chemistry data. Starting in 2024, field samples are composited by habitat type prior to laboratory analysis.

Preserved and freeze-dried algae subsamples and permanent slides (up to 12/2023), preserved whole samples from 2024-2025, and metabarcode archive filters are stored at the NEON Biorepository and are available by request for further study and analysis. Contact the Biorepository for detailed information about sample availability.

3.1 Spatial Sampling Design

In wadeable streams, periphyton are 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 majority 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[06]) and protocol AOS Protocol and Procedure for Periphyton and Phytoplankton Sampling (AD[07]) for additional details on strategy and SOPs.

Seston and phytoplankton are collected from the water column at water chemistry sampling locations (Figure 1). In wadeable streams, seston samples are collected near the S2 sensor in the thalweg of the stream. In lakes, phytoplankton is collected at the central location (near the buoy) and the littoral sensor sets. In rivers, phytoplankton is sampled near the sensor set (buoy), and from two other deep-water locations in the main channel. In lakes and rivers, 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[07] for details). In a stratified system, the sample is composited from 1 surface sample, 1 sample in the metalimnion, and one sample near the bot-

tom of the euphotic zone. Starting in 2024, the 3 phytoplankton samples collected in lakes and rivers are compisuted prior to analysis.

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. Starting in 2024, the 5 dominant field samples are composited into 1 sample and the 3 subdominant field samples are composited into a second sample used for analysis.

In lakes and rivers, 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 (NEON.DOC.003826). 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 2024, the 5 littoral field samples are composited prior to analysis.

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/

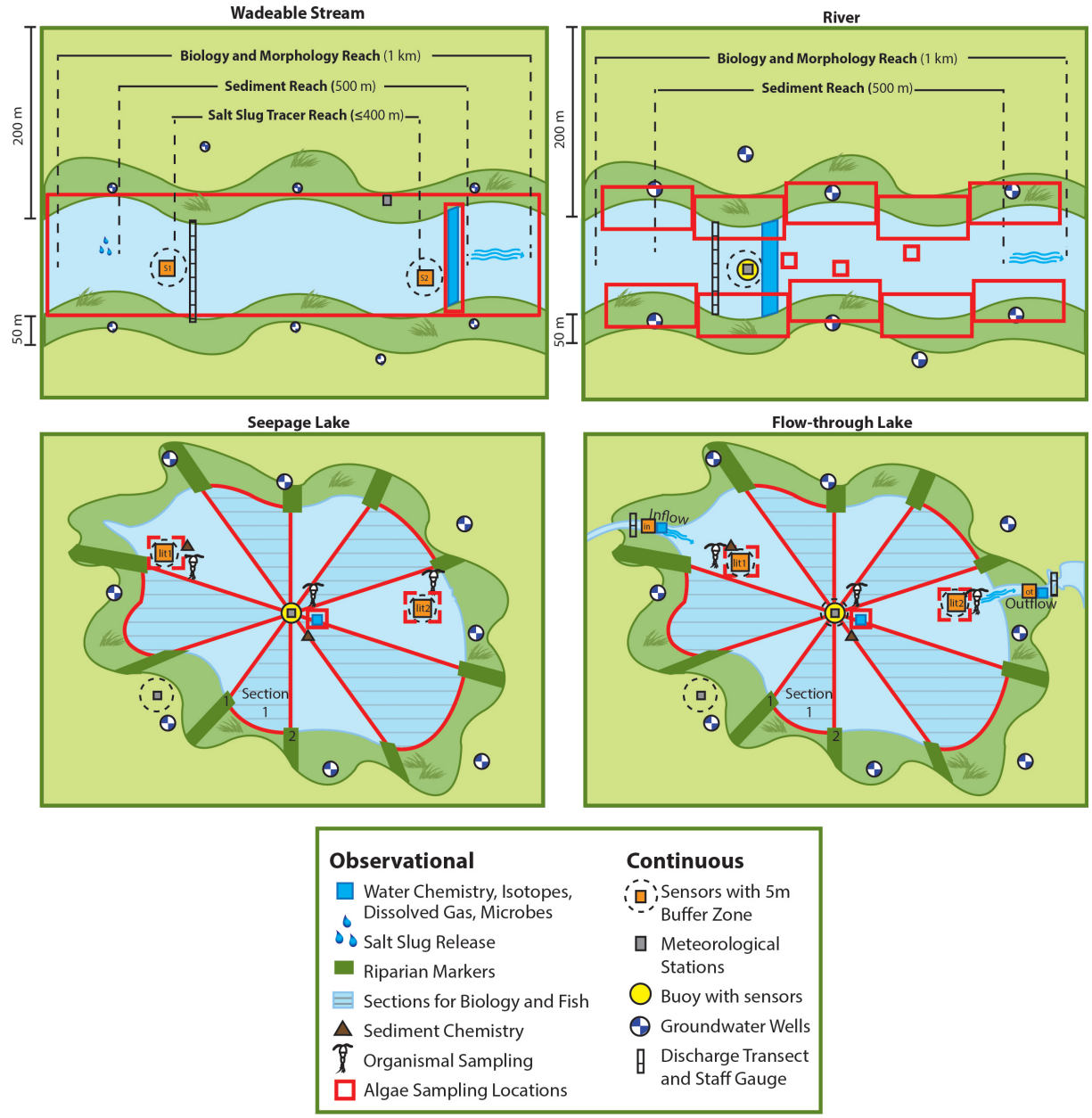

Figure 1: Generic aquatic site layouts for wadeable streams, rivers, and lakes with periphyton, seston, and phytoplankton spatial sampling locations in red.

3.2 Temporal Sampling Design

Algae sampling occurs three times per year in order to capture presence and abundance of multiple species and growth forms. The three bouts are intended to capture significant phases of the growing season; exact 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[06]). Sample bout 1 is in the early season, 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 will be allowed after a flood event for periphyton to recolonize before sampling occurs. See NEON Aquatic Sampling Strategy (AD[06]), AOS Protocol and Procedure for Periphyton and Phytoplankton Sampling (AD[07]) for additional details.

3.3 Sampling Design Changes

Location names for the nearshore sensors in seepage lakes (lakes without a true inlet and outlet stream) were changed on January 1, 2021. The location previously known “inlet” changed to “littoral 1” (“lit2”) and “outlet” changed to “littoral 2” (“lit2”) to indicate that these locations are not near an inlet or outlet stream. Flow-through lakes (e.g., D18 TOOK) have sensors in the inflow and outflow streams, as well as the lit1 and lit2 locations in the lake.

The Periphyton, seston, and phytoplankton collection data product (DP1.20166.001) was suspended in January 2026. Due to that suspension, the alg_archive table was moved to this data product to document movement of the archive samples from the field to the NEON Biorepository. Records detailing the archived algae metabarcode sample are also documented in this data product due to the suspension, in **alg_domainLabChemistry.analyteSampleID**.

Chlorophyll and pheophytin

- Chlorophyll and pheophytin data are populated in tables starting in October 2017. Prior to this date, there was 1) an audit failure at the contracted lab so analyte concentrations were removed, and 2) a gap in time when there was no contract in place.
- Chlorophyll data reported as analyte = **total chlorophyll a** represent values that are not corrected for pheophytin. Data reported as analyte = **chlorophyll a** have been corrected for pheophytin. See lab SOPs referenced in the data, located in the NEON document library for more details.
- Raw chlorophyll data are provided starting in January 2019, including dilution factor, solvent volume, and pre and post acidification fluorescence. Details about each lab’s calculations can be found in their SOPs, referenced in the NEON data and in the document library.

C, N, P, and isotopes

- Analysis of C, N, P, and C, N, S isotopes was discontinued in wadeable stream seston in 2018 due to low algal cell concentration in seston samples. Chlorophyll *a* and pheophytin analysis in seston is still reported.

- C, N, P, and C, N, S isotope filters were frozen at -20 C prior to 10/19/2017. After 10/19/2017, filters are dried overnight at 65 C prior to analysis to align with standard methods and save shipping costs.
- It was discovered that some C isotope data may have been impacted by the presence of inorganic carbon. Scientists at NEON and the contracted external lab worked together to develop a method to remove inorganic carbon from samples via acidification. Beginning in 08/2020, sites that were determined to contain high inorganic carbon are acidified prior to analysis. Users can identify acidified samples using the alg_algaeExternalLabDataPerSample:**acidTreatment** field.

Composites - Starting in 2024, composites are created from field samples collected in each habitat type, resulting in 1 composite taxonomy sample from the dominant habitat type and 1 composite sample from the subdominant habitat type from each **eventID**. - Data collected from 2014-2023 were not collected as composites, however records from this time period are averaged and presented in the alg_domainLabChemistryComp, alg_algaeDataPerSampleCompChl, and alg_algaeDataPerSampleCompCNPS tables with **dataQF** = “means of uncomposited data”.

3.4 Laboratory Quality Assurance and Uncertainty

External laboratory facilities have been chosen for their use of analytical methods widely adopted by the scientific community. Labs report the long-term analytical precision and uncertainty of standard reference materials analyzed as unknowns for each analyte in the data quality table (alg_algaeExternalLabQA). This allows users to interpret and model the periphyton and phytoplankton chemical properties data in the context of its uncertainty range. Contracted external facilities upload a summary file (asi_externalLabSummaryData) when they begin work for NEON, then again once per year or whenever their information changes (for example, a new instrument is acquired or a change is detected in analytical precision). Additionally, NEON’s Calibration/Validation department has regular procedures for auditing the quality assurance of external laboratories and their reports are available to data users.

Filters for carbon, nitrogen, and sulfur are created in duplicate at the domain support facility, indicated by the **analysisType** CNSa and CNSb (this naming convention began around 2017-10-01). The lab will analyze one of these filters per parent sample, reserving the second filter to be analyzed in the case of equipment failure. Approximately 25% of filters or less require the reserve filter to be analyzed. In data downloads, users should expect to see analytical data for either the CNSa or CNSb filter for a given parent sample, but not both.

Due to high levels of inorganic carbon in the surface water of some NEON sites, samples from certain sites are decarbonated via acid digestion (1N HCl for 30 min) prior to analysis. The following table reports the acid treatment status for each NEON Aquatic site.

Table 1: Acid treatment status for each NEON Aquatic site

NEON Domain ID	NEON Site ID	Acid Treatment	Notes
D01	HOPB	Y	
D02	LEWI	Y	
D02	POSE	Y	
D03	BARC	Y	Only phytoplankton samples
D03	FLNT	Y	Only phytoplankton samples
D03	SUGG	Y	Only phytoplankton samples
D04	CUPE	N	
D04	GUIL	Y	
D05	CRAM	N	
D05	LIRO	N	
D06	KING	Y	
D06	MCDI	Y	
D07	LECO	N	
D07	WALK	Y	
D08	BLWA	N	
D08	MAYF	N	
D08	TOMB	N	
D09	PRLA	Y	
D09	PRPO	Y	
D10	ARIK	Y	
D11	BLUE	Y	
D11	PRIN	Y	
D12	BLDE	Y	
D13	COMO	N	
D13	WLOU	Y	
D14	SYCA	N	
D15	REDB	Y	
D16	MART	Y	
D16	MCRA	N	
D17	BIGC	Y	
D17	TECR	Y	
D18	OKSR	N	
D18	TOOK	Y	
D19	CARI	Y	

3.5 Variables Reported

All variables reported from the field technician or laboratory (L0 data) are listed in the files, NEON Raw Data Validation for AOS Periphyton and Phytoplankton Collection, Level 0 (DP0.20166.001) (AD[03]) and NEON Raw Data Validation for Plant and Algae External Lab Chemistry (DP0.20065.001) (AD[04]). All variables reported in the published data (L1 data) are also provided separately in the file, NEON Data Variables for Periphyton and Phytoplankton Chemical Properties (DP1.20163.001) (AD[05]).

Field names have been standardized with Darwin Core terms (http://rs.tdwg.org/dwc/; accessed 16 February 2014), 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 AOS spatial data employs the World Geodetic System 1984 (WGS84) for its fundamental reference datum and Geoid12A geoid model for its vertical reference surface. Latitudes and longitudes are denoted in decimal notation to six decimal places, with longitudes indicated as negative west of the Greenwich meridian.

Some variables described in this document may be for NEON internal use only and will not appear in downloaded data.

3.6 Spatial Resolution and Extent

Wadeable stream Each periphyton sample represents a patch of stream bottom within the 1 km permitted wadeable stream reach. The exact location of each sample is not 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).

Each seston sample is collected in the water column near the wadeable stream S2 sensor location.

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. 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 or c1), littoral 1 (lit1), and littoral 2 (lit2). In rivers, one sample is collected near the sensor buoy (c0 or c1), 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).

3.7 Temporal Resolution and Extent

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

This data product is dependent on the field data collected in AOS Periphyton and Phytoplankton Collection (DP1.20166.001). Data can be linked to the parent sample through the **sampleID** field.

All of the above data products are also loosely related to gauge height data and associated metadata collected on the same sampling day (DP1.20267.001). These data products are linked through the **siteID** and **collectDate**.

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.

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 **siteID**, **collectDate**, and **sampleNumber**, or by parent **sampleID**.

3.9 Product Instances

At each aquatic site, there will be 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. Because data are reported in long format (as opposed to wide), each sample generates records for each analyte measured, ~8 analytes per sample. Across the observatory, this leads to up to ~7300 data records per year. Early data (**collectDate** prior to March 2017) may show additional analytical replicates per analyte.

Ten percent of samples will be haphazardly selected for analytical replicates on all analyses. If replicate samples are taken, there will be one unique sample per **replicate** per **sampleID** per sampling event, and the sample ID(s) of the replicate sample(s) will have the **replicate** appended to the end (for samples with a **collectDate** later than March 2017).

3.10 Data Relationships

A record in alg_domainLabChemistry, alg_algaeExternalLabDataPerSample, and alg_algaeExternalLabQA should have a corresponding record in alg_fieldData describing field collection conditions, location, and metadata during sample collection. If **sampleID** is empty in alg_fieldData, there will be no additional records in the alg_domainLabChemistry, alg_algaeExternalLabDataPerSample, or alg_algaeExternalLabQA

tables. Records in alg_externalLabPOMSummaryData are entered about once per year, or anytime the lab equipment or standards change. 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 the alg_domainLabChemistry table.

alg_domainLabChemistry.csv - > One record (**sampleID** = **parentSampleID** plus **analysisType** plus **replicate**) is created for each subsample processed at the NEON domain lab. alg_domainLabChemistry can be linked to alg_fieldData via the **parentSampleID**, and downstream lab tables via the **sampleID**. Subsamples created include 2 filters for carbon, nitrogen, and sulfur (CNSa and CNSb), 1 filter for phosphorus (P) and 1 filter for chlorophyll/pheophytin. This table does not contain data after 12/31/2023.

alg_domainLabChemistryComp.csv - > One record (**compositeSampleID**) is created for each composite sample (pooling of multiple **parentSampleID**) processed at the NEON domain lab. alg_domainLabChemistryComp can be linked to alg_fieldData via the **parentSampleID**, and downstream lab tables via **analyteSampleID** to **sampleID**. Subsamples created include 2 filters for carbon, nitrogen, and sulfur (CNSa and CNSb), 1 filter for phosphorus (P), 1 filter for chlorophyll/pheophytin, and data for 1 ash-free dry mass (AFDM) filter. Sample information for the metabarcode archive filter is also contained here. Records collected from 2014-2023 represent averages of uncomposited samples, records starting in 2024 represent composited samples.

alg_labBlank.csv - > One record is created for each **domainID** and **eventID** when chemistry filters are created in alg_domainLabChemistry or alg_domainLabChemistryComp. Records represent DI water filtered in the Domain Support Facility and are linked via the **filterBlankRep1** and **filterBlankRep2** fields. Filters are analyzed by the external lab and results are published in the alg_algaeExternalLabQA table.

alg_algaeExternalLabDataPerSample.csv - > One record is created for each analytical replicate of each analyte for a sample, for each analyte in a long-format table, resulting in multiple entries per sample in a long-format table, resulting in multiple entries per sample. Data can be joined to alg_domainLabChemistry through the **sampleID**, then from that table to the alg_fieldData through the **parentSampleID**. Note that data from only 1 replicate carbon, nitrogen, sulfur filter (CNSa or CNSb) are expected in this table, as the other filter may be used for instrument calibration. This table does not contain data after 12/31/2023.

alg_algaeDataPerSampleCompCNPS - > One record is created for each analytical replicate of each analyte for a CNS or P sample, for each analyte in a long-format table, resulting in multiple entries per sample in a long-format table, resulting in multiple entries per sample. Data can be joined to alg_domainLabChemistry through **sampleID** = **analyteSampleID**, then from that table to the alg_fieldData through the **parentSampleID**. Note that data from only 1 replicate carbon, nitrogen, sulfur filter (CNSa or CNSb) are expected in this table, as the other filter may be used for instrument calibration. Records collected from 2014-2023 represent averages of uncomposited samples, records starting in 2024 represent composited samples.

alg_algaeDataPerSampleCompChl.csv - > One record is created for each analytical replicate of each analyte for a chlorophyll/pheopytin sample, for each analyte in a long-format table, resulting in multiple entries per sample in a long-format table, resulting in multiple entries per sample. Data can be joined to

alg_domainLabChemistry through **sampleID** = **analyteSampleID**, then from that table to the alg_fieldData through the **parentSampleID**. Records collected from 2014-2023 represent averages of uncomposited samples, records starting in 2024 represent composited samples.

alg_algaeExternalLabQA.csv - > One record is created per blank or standard and are not directly tied to the NEON **sampleID**. Records in this table can be linked to alg_algaeExternalLabDataPerSample through the **batchID** field.

alg_archive.csv - > Summary information for subsamples shipped toe the NEON Biorepository, includes samples from the Periphyton, seston, and phytoplankton collection (DP1.20166.001) data product. Tied to the **sampleID** and **parentSampleID**.

asi_externalLabPOMSummaryData.csv - > Summary information for each analytical method are recorded in this table with **startDate** and **endDate**. These dates can be used to apply to the data in alg_algaeExternalLabDataPerSample and alg_algaeExternalLabQA using the fields **laboratoryName** and **analyte** that have a **analysisDate** falling between the **startDate** and **endDate** in this table.

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

3.11 Special Considerations

Aquatic periphyton and phytoplankton data are generated from analyses that take place at external labs. Labs use standard machines and techniques, and use replicate samples for several analytes. Algal chemistry samples/filters are destroyed during analysis and are not archived. See the lab SOPs for more information.

Data users may wish to use average analytical replicate data for each unique **sampleID** + **analyte** + **replicate** combination, where replicates exist.

The periphyton and phytoplankton chemical analysis data are reported in the field **analyteConcentration**. This field should be coupled with **analyte** and **plantAlgaeLabUnits** to find the units for each analyte. The analytical data are corrected for the **sampleVolumeFiltered**, but are NOT corrected for **benthicArea** in the case of periphyton samples. Data users will need to refer to the **benthicArea** presented in the alg_fieldData table and apply this correction to get the concentration of analyte per stream, lake, or river bottom. All analyte records, with the exception of $\delta^{13}C$, ^{15}N, or ^{34}S records, from a sample should be divided by the **benthicArea** prior to reporting the concentration per m^2.

Periphyton (benthic algae: epilithon, epixylon, epipelon, epipsammon, epiphyton)

alg_domainLabChemistry and alg_algaeExternalLabDataPerSample

$$algalConcentrationPerM_i^2 = alg_algaeExternalLabDataPerSample.analyteConcentration_i \times \frac{alg_domainLabChemistry.fieldSampleVolume}{alg_fieldData.benthicArea_i \times 1000} \quad (1)$$

Where *i* = **sampleID** + **analyte** + **replicate**. Note that **fieldSampleVolume** is in mL and **analyteConcentration** is in ug/L.

alg_domainLabChemistryComp and alg_algaeDataPerSampleCompCNPS

$$algalConcentrationPerM_i^2 = alg_algaeDataPerSampleCompCNPS.analyteConcentration_i \times \frac{alg_domainLabChemistryComp.fieldSampleVolume}{alg_fieldData.benthicArea_i \times 1000} \quad (2)$$

Where *i* = **analyteSampleID**. Note that **fieldSampleVolume** is in mL and **analyteConcentration** is in ug/L.

alg_domainLabChemistryComp and alg_algaeDataPerSampleCompChl

$$algalConcentrationPerM_i^2 = alg_algaeDataPerSampleCompChl.analyteConcentration_i \times \frac{alg_domainLabChemistryComp.fieldSampleVolume}{alg_fieldData.benthicArea_i \times 1000} \quad (3)$$

Where *i* = **analyteSampleID**. Note that **fieldSampleVolume** is in mL and **analyteConcentration** is in ug/L.

Phytoplankton or Seston

alg_domainLabChemistry and alg_algaeExternalLabDataPerSample

$$algalConcentrationPerL_i = alg_algaeExternalLabDataPerSample.analyteConcentration_i \quad (4)$$

Where *i* = **sampleID** + **analyte** + **replicate**. Note that **analyteConcentration** is in ug/L.

alg_domainLabChemistryComp and alg_algaeDataPerSampleCompCNPS

$$algalConcentrationPerL_i = alg_algaeDataPerSampleCompCNPS.analyteConcentration_i \quad (5)$$

Where *i* = **analyteSampleID**. Note that **fieldSampleVolume** is in mL and **analyteConcentration** is in ug/L

alg_domainLabChemistryComp and alg_algaeDataPerSampleCompChl

$$algalConcentrationPerL_i = alg_algaeDataPerSampleCompChl.analyteConcentration_i \quad (6)$$

Where *i* = **analyteSampleID**. Note that **analyteConcentration** is in ug/L.

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, seston, 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 and Phytoplankton Collection, Level 0 (DP0.20166.001), 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[11]).

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[10]).

Published data are reviewed for completeness, timeliness, and validity using an internal set of tests and metrics, as detailed in the NEON Algorithm Theoretical Basis Document: OS Data Quality Control (AD[13]). These quality tests are used to guide process improvements, audits of analytical facilities, and data updates, but do not generate quality flags in published data.

4.3 Data Revision

All data are provisional until a numbered version is released. Annually, NEON releases a static version of all or almost all data products, annotated with digital object identifiers (DOIs). The first data Release was made in 2021. 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 Issue Log section of the data product landing page contains a history of major known errors and revisions.

4.4 Quality Flagging

The **dataQF** and **externalLabDataQF** fields in each data record of the alg_fieldData and alg_algaeExternalLabDataPerSample tables, respectively, are quality flags for known errors applying to the record. Please see the table below for an explanation of quality flagging codes specific to this product.

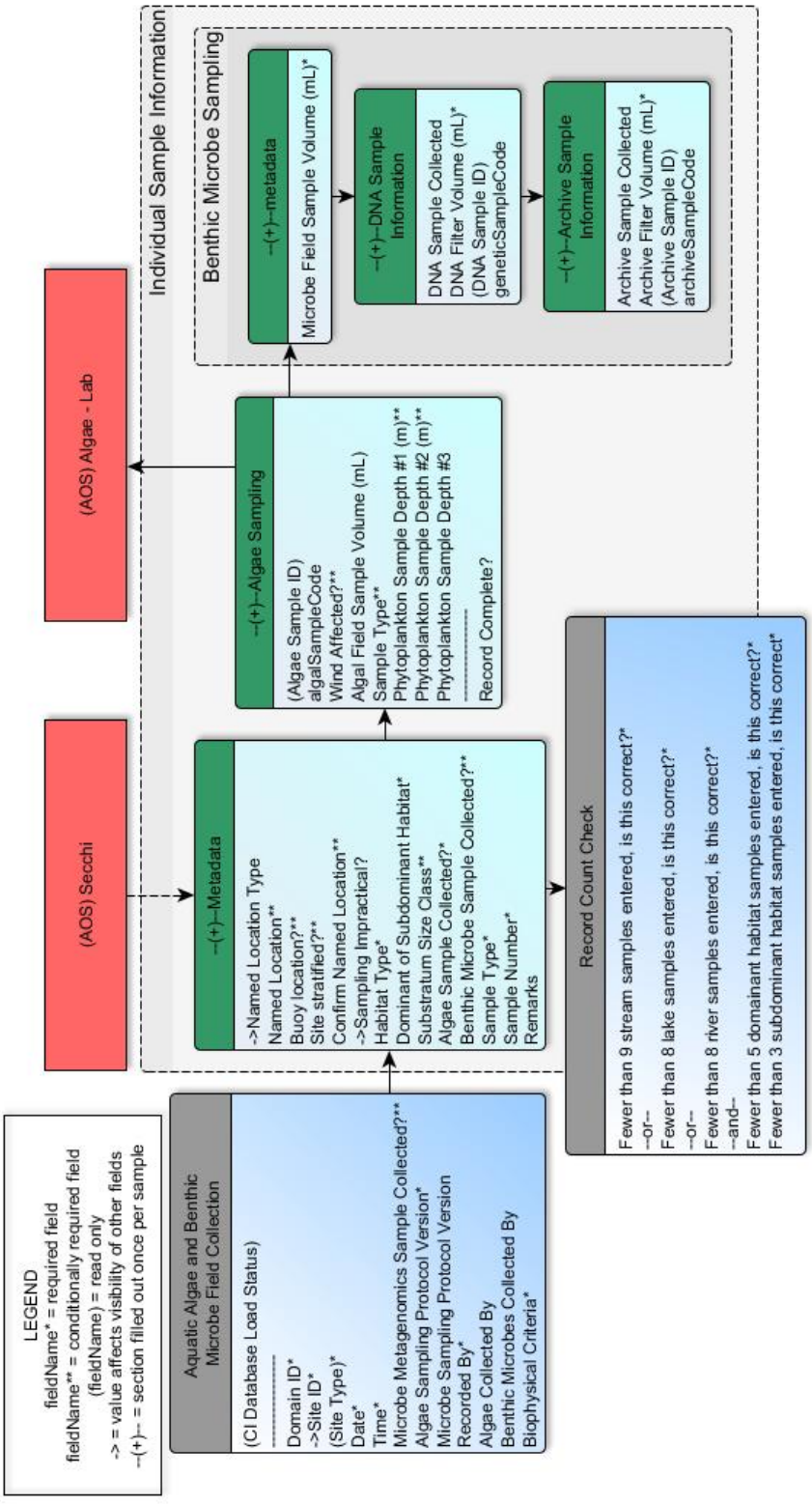

Figure 2: Schematic of the applications used by field technicians to enter periphyton and phytoplankton field data

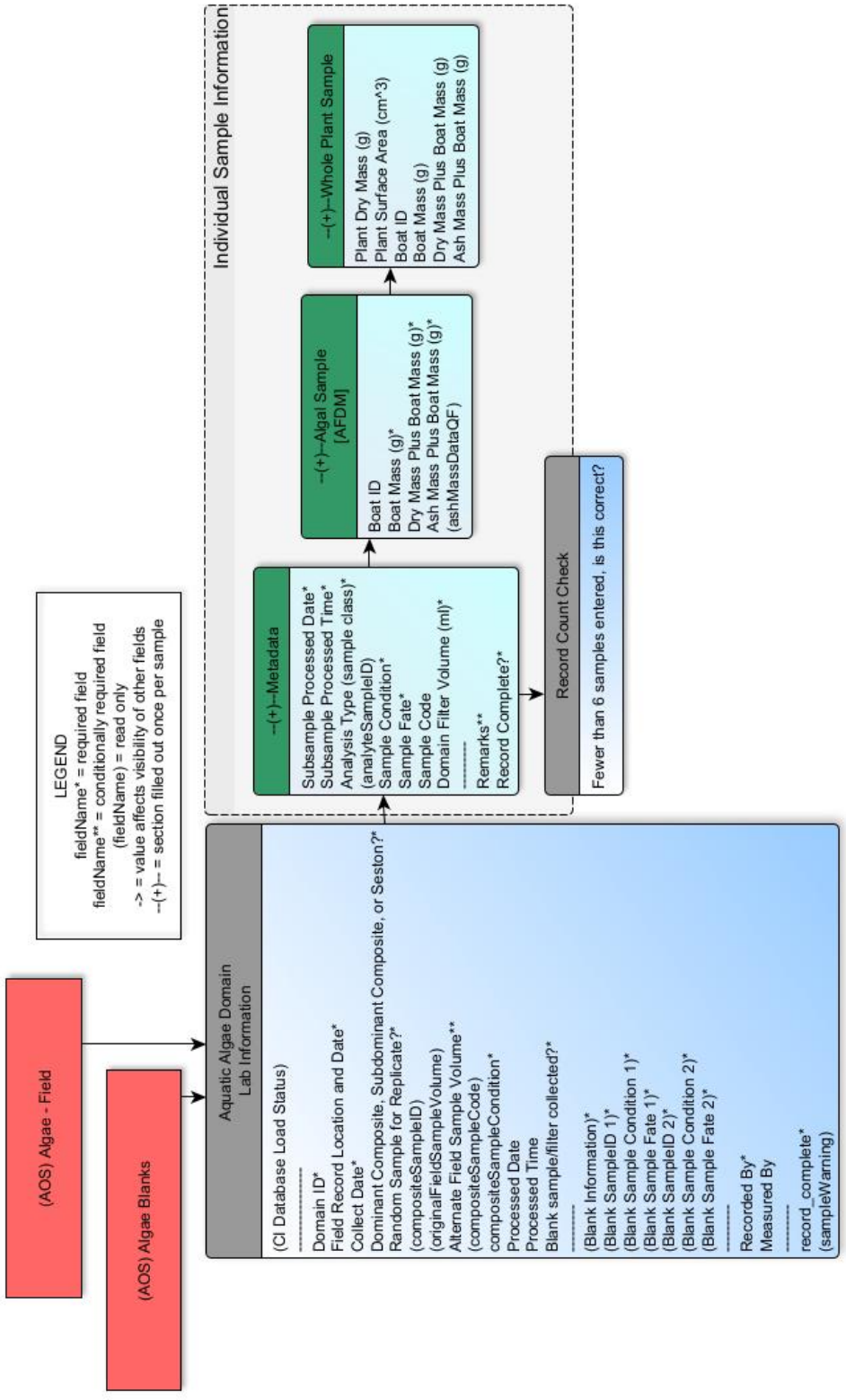

Figure 3: Schematic of the applications used by field technicians to enter domain lab subsampling and biomass data

Table 2: Descriptions of the dataQF and externalLabDataQF codes for quality flagging

fieldName	value	definition
dataQF / externalLabDataQF	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
externalLabDataQF	Did not meet quality audit requirements for analysis audit	The external lab did not meet the requirements of the NEON external facility audit for the year the data were generated
externalLabDataQF	acidTreatmentSOPNotFollowed	The external lab did not follow the standard operating procedure that indicated samples were decarbonated via acid fumigation

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)

4.5 Analytical Facility Data Quality

Data analyses conducted on algal chemistry data conform to the current data quality standards used by practitioners. Prior to 2017, replicate filters were analyzed for each analysis except for P and 34S. Starting in 2017, replicate filters are created and analyzed for each analysis for every ~10% of samples. These records are indicated by the same parent sample ID and “rep2”. In addition, secondary standards or reference material are analyzed in every batch of NEON data. Data for standards and blanks are presented in the table “alg_algaeExternalLabQA.csv”. Annual information on changing instrumentation or methods is recorded in “asi_externalLabPOMSummaryData”, both tables are included in this download package. Details on lab analyses and quality control can be found in the external lab SOPs.

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.