NEON USER GUIDE TO PERIPHYTON AND PHYTOPLANKTON CHEMICAL PROPERTIES (NEON.DP1.20163)

PREPARED BY	ORGANIZATION	DATE
Tanya Chesney	DPS	12/16/2017
Stephanie Parker	AOS	12/16/2017
Caren Scott	AOS	12/16/2017

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	08/15/2017	Initial Release

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 . . . 4
3.3 Laboratory Quality Assurance and Uncertainty . . . 5
3.4 Variables Reported . . . 5
3.5 Spatial Resolution and Extent . . . 5
3.6 Temporal Resolution and Extent . . . 6
3.7 Associated Data Streams . . . 6
3.8 Product Instances . . . 7
3.9 Data Relationships . . . 7
3.10 Special Considerations . . . 7

4 DATA QUALITY **8**
4.1 Data Entry Constraint and Validation . . . 8
4.2 Automated Data Processing Steps . . . 8
4.3 Data Revision . . . 11
4.4 Quality Flagging . . . 11
4.5 Analytical Facility Data Quality . . . 11

5 REFERENCES **11**

LIST OF TABLES AND FIGURES

Figure 1 Generic aquatic site layouts for non-wadeable streams/rivers, wadeable streams, and lakes with periphyton, seston, and phytoplankton spatial sampling locations in red. . . . 4
Figure 2 Schematic of the applications used by field technicians to enter periphyton and phytoplankton field data . . . 9
Figure 3 Schematic of the applications used by field technicians to enter domain lab subsampling and biomass data . . . 10

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 concentration of 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, non-wadeable streams, 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 (NEON.DP1.20163) (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 (NEON.DP0.20166) (AD[03]) and NEON Raw Data Validation for Plant and Algae External Lab Chemistry (NEON.DP0.20065) (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 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 and Phytoplankton Collection, Level 0 (NEON.DP0.20166)
AD[04]	NEON.DP0.20065.001 _dataValidation.csv	NEON Raw Data Validation for Plant and Algae External Lab Chemistry (NEON.DP0.20065)
AD[05]	NEON.DP1.20163.001 _variables.csv	NEON Data Variables for Periphyton and Phytoplankton Chemical Properties (NEON.DP1.20163)
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]	OS_Generic_Transitions .pdf	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[11]	Nicl Language.pdf	NEON's Ingest Conversion Language (NICL) specifications

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*, 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[06]); 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 pheophyton. 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]).

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.

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. 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[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 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.

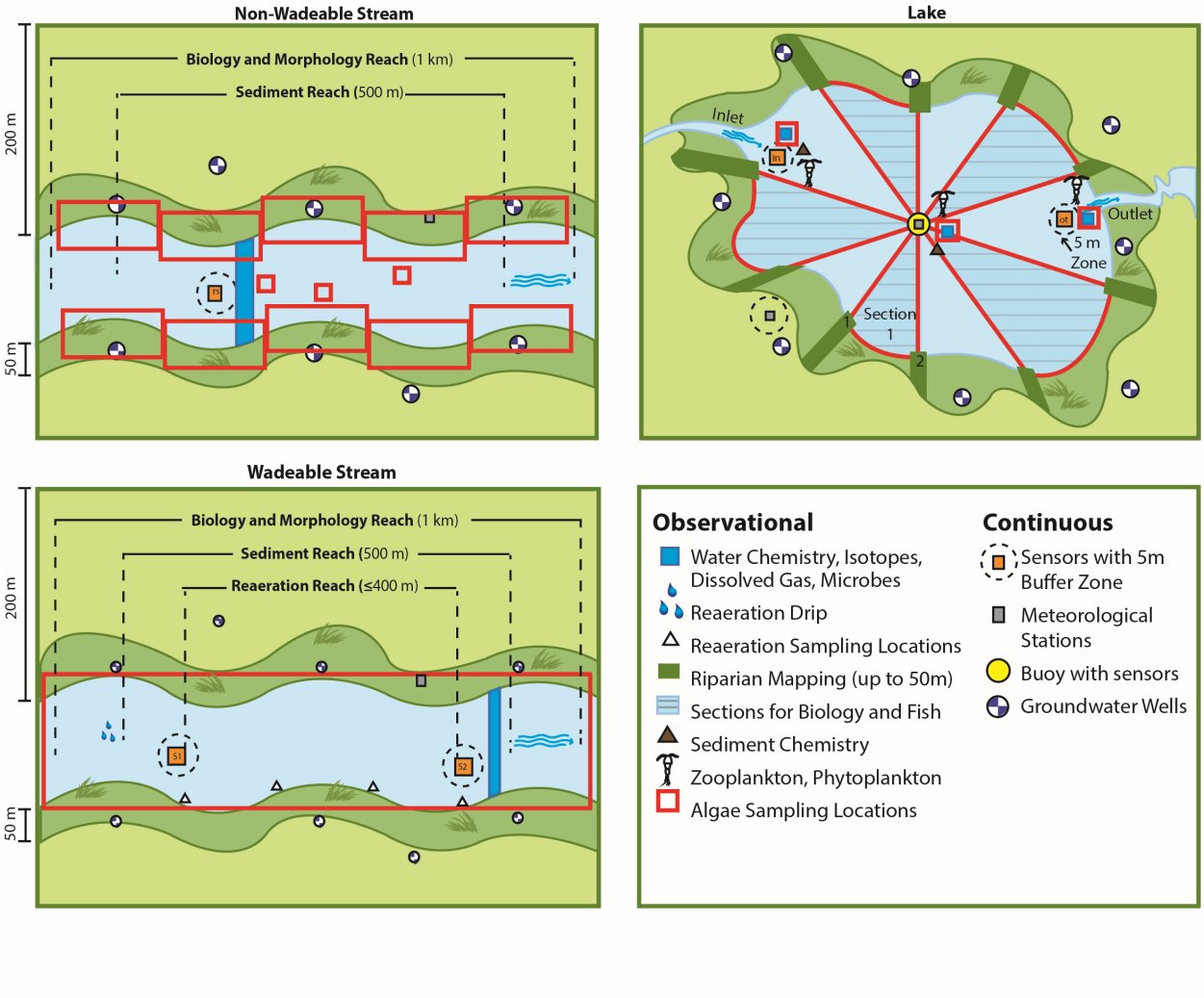

Figure 1: Generic aquatic site layouts for non-wadeable streams/rivers, wadeable streams, and lakes with periphyton, seston, and phytoplankton spatial sampling locations in red.

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 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 a summary file. 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.

3.4 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 (NEON.DP0.20166) (AD[03]) and NEON Raw Data Validation for Plant and Algae External Lab Chemistry (NEON.DP0.20065) (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 (NEON.DP1.20163) (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 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 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), 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).

3.6 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

3.7 Associated Data Streams

This data product is dependent on the field data collected in AOS Periphyton and Phytoplankton Collection (NEON.DP1.20166). 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 (NEON.DP1.20267). 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" (NEON.DP1.20252) and "Depth profile at specific depths" (NEON.DP1.20254). 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" (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**.

3.8 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) will 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).

3.9 Data Relationships

A record in alg_domainLabChemistry, alg_algaeExternalLabDataPerSample, or alg_externalLabPOMSummaryData 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_externalLabPOMSummaryData tables. 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 **parentSsampleID** which is linked to all subsequent tables (except asi_externalLabPOMSummaryData).

alg_domainLabChemistry.csv - > One record (**sampleID** plus **analysisType** plus **replicate**) is created for each subsample processed at the NEON domain lab.

alg_algaeExternalLabDataPerSample.csv - > One record is created for each analytical replicate of each analyte for a sample, resulting in multiple entries per sample. Data can be tracked to alg_fieldData through the parent **sampleID** or to alg_domainLabChemistry data through **sampleID** plus **analyte** plus **replicate**.

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 using the fields **laboratoryName**, **analysisDate**, and **analyte**.

3.10 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.

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

The periphyton and phytoplankton chemical analyses are reported in the field **analyteConcentration**. This field should be coupled with **plantAlgaeLabUnits** to find the units for each analyte. The analytical data are corrected for the **sampleVolumeFiltered**, but are NOT corrected for **benthicArea**. 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 δ^{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.

$$algalConcentrationPerM_i^2 = alg\_algaeExternalLabDataPerSample.analyteConcentration_i \times \frac{alg\_domainLabChemistry.fieldSampleVolume}{alg\_fieldData.benthicArea_i} \quad (1)$$

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

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 (NEON.DP0.20166), 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]).

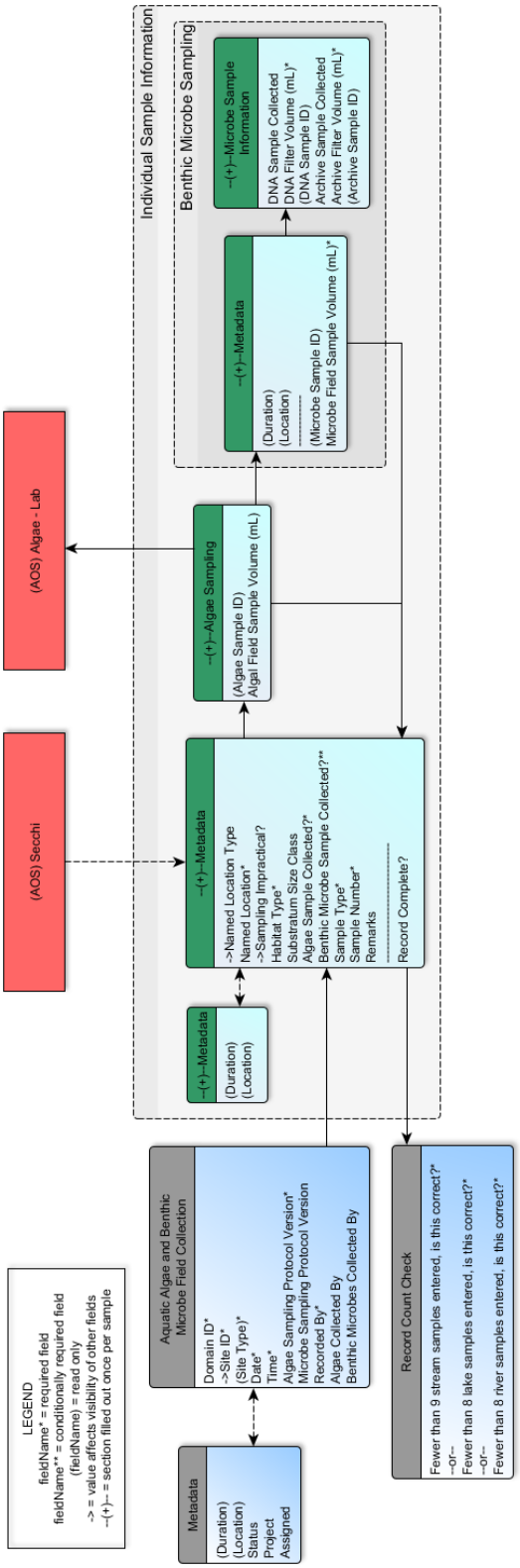

Figure 2: Schematic of the applications used by field technicians to enter periphyton and phytoplankton field data

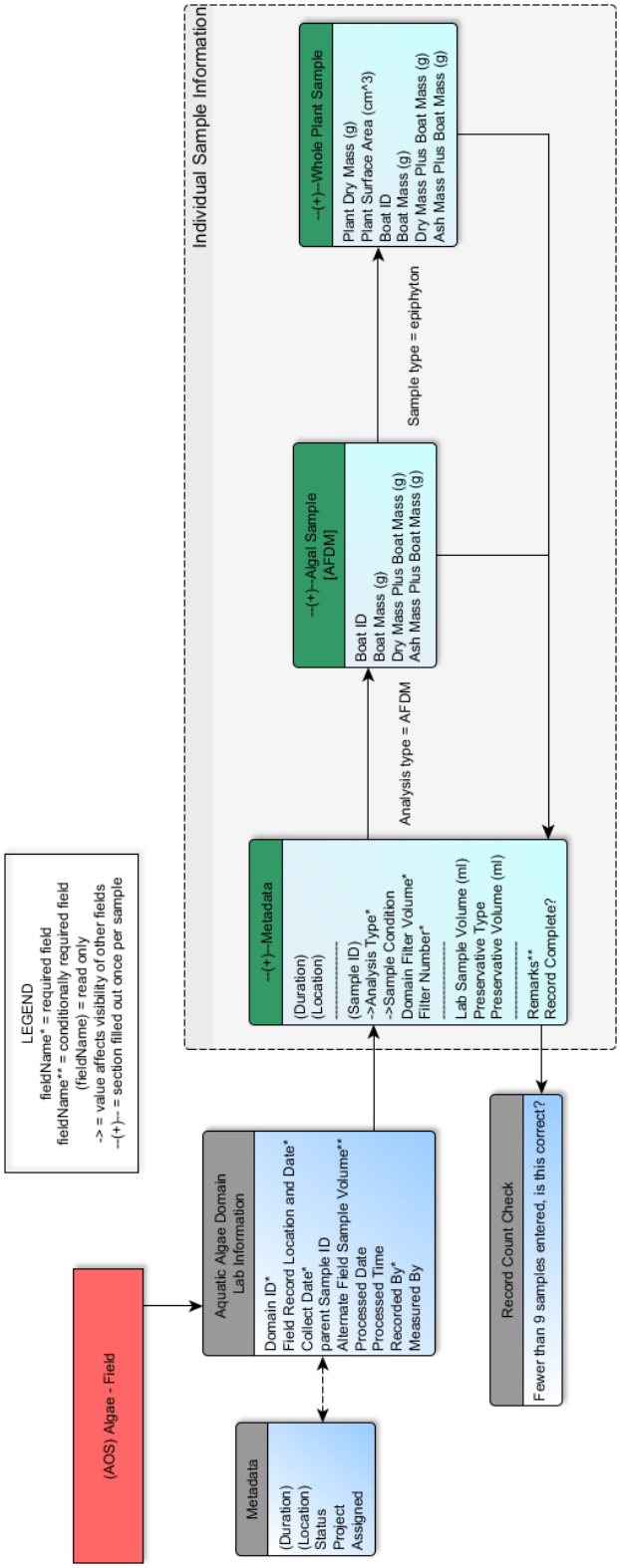

Figure 3: Schematic of the applications used by field technicians to enter domain lab subsampling and biomass data

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
dataQF	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

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. Lab quality data are presented in the table “asi_externalLabPOMSummaryData” that is included in this download package. Details on the on lab analyses and quality control these fields can be found in the external lab SOP(s).

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.