



|  |                         |
|--|-------------------------|
| <i>Title:</i> NEON User Guide to Aquatic Plant Bryophyte Macroalgae Clip Harvest (DP1.20066.001) | <i>Date:</i> 12/02/2025 |
| <i>Author:</i> Stephanie Parker  | <i>Revision:</i> E.2    |

# NEON USER GUIDE TO AQUATIC PLANT BRYOPHYTE MACROALGAE CLIP HARVEST (DP1.20066.001)

| <b>PREPARED BY</b> | <b>ORGANIZATION</b> |
|--------------------|---------------------|
| Stephanie Parker   | AOS                 |
| Caren Scott        | AOS                 |

## CHANGE RECORD

| REVISION | DATE       | DESCRIPTION OF CHANGE  |
|----------|------------|--|
| A        | 07/03/2017 | Initial Release  |
| B        | 10/12/2020 | Included general statement about usage of neonUtilities R package and statement about possible location changes. Updated taxonomy information and added morphospecies clarification. Included temporal optimization and updated spatial figure to indicate changes to lake littoral and inflow/outflow sensor locations. |
| B.1      | 04/22/2021 | Edit formatting in Sampling Design Changes.  |
| C        | 04/08/2022 | Added language in section 4 Taxonomy addressing RTE species obfuscation in the data. Updated section 5.3 Data Revision with latest information regarding data release.   |
| C.1      | 08/10/2023 | Updated information about the apl_identificationHistory table.   |
| D        | 04/09/2024 | Minor formatting updates.  |
| E        | 04/17/2025 | Added information about the new neonUtilities Python package.  |
| E.1      | 09/19/2025 | Added information about the apc_voucher tables.  |
| E.2      | 12/2/2025  | Updated section 5.4 to include new dataQF code.  |

## TABLE OF CONTENTS

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>DESCRIPTION</b>                             | <b>1</b>  |
| 1.1      | Purpose . . . . .                              | 1         |
| 1.2      | Scope . . . . .                                | 1         |
| <b>2</b> | <b>RELATED DOCUMENTS AND ACRONYMS</b>          | <b>2</b>  |
| 2.1      | Associated Documents . . . . .                 | 2         |
| 2.2      | Acronyms . . . . .                             | 2         |
| <b>3</b> | <b>DATA PRODUCT DESCRIPTION</b>                | <b>3</b>  |
| 3.1      | Spatial Sampling Design . . . . .              | 3         |
| 3.2      | Temporal Sampling Design . . . . .             | 5         |
| 3.3      | Sampling Design Changes . . . . .              | 5         |
| 3.4      | Variables Reported . . . . .                   | 6         |
| 3.5      | Temporal Resolution and Extent . . . . .       | 6         |
| 3.6      | Spatial Resolution and Extent . . . . .        | 6         |
| 3.7      | Associated Data Streams . . . . .              | 7         |
| 3.8      | Product Instances . . . . .                    | 7         |
| 3.9      | Data Relationships . . . . .                   | 7         |
| <b>4</b> | <b>TAXONOMY</b>                                | <b>9</b>  |
| 4.1      | Plant Taxonomy . . . . .                       | 9         |
| 4.2      | Algae Taxonomy . . . . .                       | 9         |
| 4.3      | Plant Morphospecies . . . . .                  | 10        |
| 4.4      | Identification History . . . . .               | 10        |
| <b>5</b> | <b>DATA QUALITY</b>                            | <b>11</b> |
| 5.1      | Data Entry Constraint and Validation . . . . . | 11        |
| 5.2      | Automated Data Processing Steps . . . . .      | 11        |
| 5.3      | Data Revision . . . . .                        | 11        |
| 5.4      | Quality Flagging . . . . .                     | 11        |
| 5.5      | Analytical Facility Data Quality . . . . .     | 14        |



|   |                         |
|---|-------------------------|
| <i>Title:</i> NEON User Guide to Aquatic Plant Bryophyte Macroalgae Clip Harvest<br>(DP1.20066.001) | <i>Date:</i> 12/02/2025 |
| <i>Author:</i> Stephanie Parker   | <i>Revision:</i> E.2    |

**6 REFERENCES**

**14**

**LIST OF TABLES AND FIGURES**

Table 1 Descriptions of the dataQF codes for quality flagging . . . . . 14

Figure 1 Aquatic plant sampling locations in all types of aquatic sites (rivers, wadeable streams, and lakes). . . . . 4

Figure 2 Aquatic plant sampling locations in wadeable streams, showing quadrat locations in proximity to established transects . . . . . 5

Figure 3 Schematic of the applications used by field technicians to enter aquatic plant clip harvest field data . . . . . 12

Figure 4 Schematic of the applications used by field technicians to measure aquatic plant biomass data and subsamples in the domain lab . . . . . 13

## 1 DESCRIPTION

### 1.1 Purpose

This document provides an overview of the data included in this NEON Level 1 data product, the quality controlled product generated from raw Level 0 data, and associated metadata. In the NEON data products framework, the raw data collected in the field, for example, the clip harvest field samples from a single collection event 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 Aquatic Plant Bryophyte Macroalgae Clip Harvest and associated metadata from input data. This document also provides details relevant to the publication of the data products via the NEON data portal, with additional detail available in the file, NEON Data Variables for Aquatic Plant Clip Harvest (DP1.20066.001) (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: Aquatic plant, bryophyte, lichen, and macroalgae sampling (AD[06]). The raw data that are processed in this document are detailed in the file, NEON Raw Data Validation for Aquatic Plant Clip Harvest (DP0.20066.001) (AD[03]), provided in the download package for this data product. Please note that raw data products (denoted by 'DPO') 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.003039               | AOS Protocol and Procedure: Aquatic plant, bryophyte, lichen, and macroalgae 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] | Nicl Language.pdf             | 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        |
|---------|-------------------|
| AFDM    | ash-free dry mass |

### 3 DATA PRODUCT DESCRIPTION

Aquatic plant, bryophyte, and macroalgae clip harvest data include taxonomy and biomass data, and provide information related to the NEON Grand Challenge areas of Biodiversity, Biogeochemistry, and Invasive Species, as well as additional data about the macrophyte community in lakes, non-wadeable streams (rivers), and wadeable streams. These data can be used to assess the health of the aquatic ecosystem over time. Aquatic vegetation will be sampled via clip harvest once per year at each NEON aquatic site (AD[05]), and presence/absence data will be collected three times per year. 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: Aquatic plant, bryophyte, lichen, and macroalgae sampling (AD[06]).

#### 3.1 Spatial Sampling Design

Clip harvest collections are made at 10 locations per site in lakes and rivers. These locations are randomized GPS points (Figure 1) within the zone of plant colonization at the site (i.e., within the depth ranges that plants are known to colonize based on light attenuation in the lake or river).

At wadeable stream sites, each clip harvest (Figure 2) is collected within 4 m of the nearest point transect (NEON.DOM.SITE.DP1.20072) at a total of 10 locations: 5 in the dominant habitat type, 5 in the second-most dominant habitat type. Transects should be established in alternating habitat types along the reach (e.g., transect 1 in a pool, transect 2 in a riffle, transect 3 in a pool). Over time, the habitats present at these transects may change, however transects will not be relocated unless they are no longer able to be sampled. If it is not possible to separate each transect by a different habitat type, transects may be spaced at least 10 m apart within the same habitat type/unit. Transects are established during the first sampling bout by domain technicians and coordinates are collected using a high-precision GPS unit (e.g., Trimble with a tornado antenna). Plot markers are established on one bank, with the transect extending perpendicular to stream flow across the wetted channel. Transects remain in place until significant morphological changes occur in the stream (e.g., bed-moving spate) that causes the transect to be unusable. After such geomorphological changes, a transect may be re-established in a new location. See AOS Protocol and Procedure: Aquatic plant, bryophyte, lichen, and macroalgae sampling (AD[06]) for additional details on strategy and SOPs.

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/>

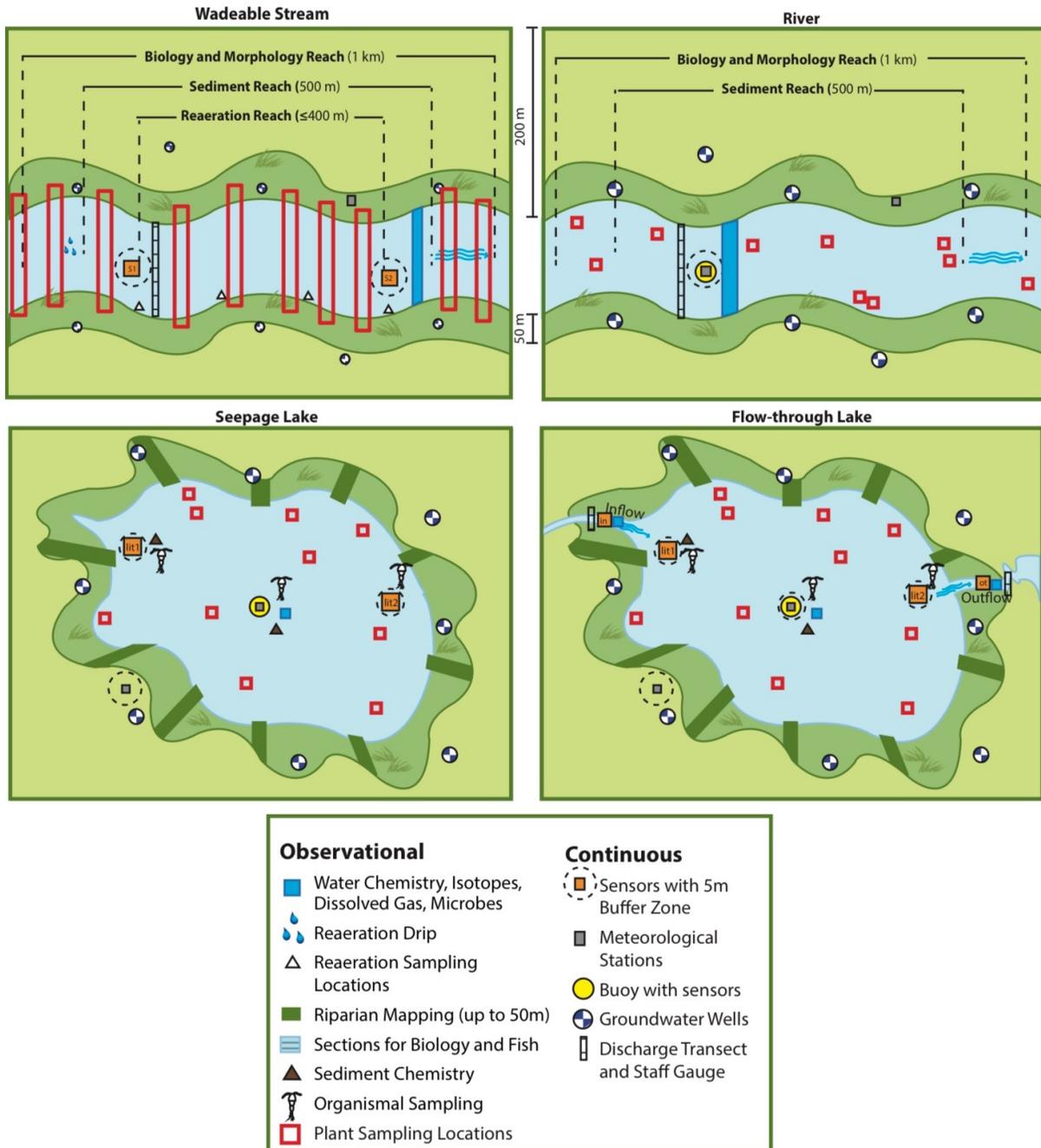


Figure 1: Aquatic plant sampling locations in all types of aquatic sites (rivers, wadeable streams, and lakes).

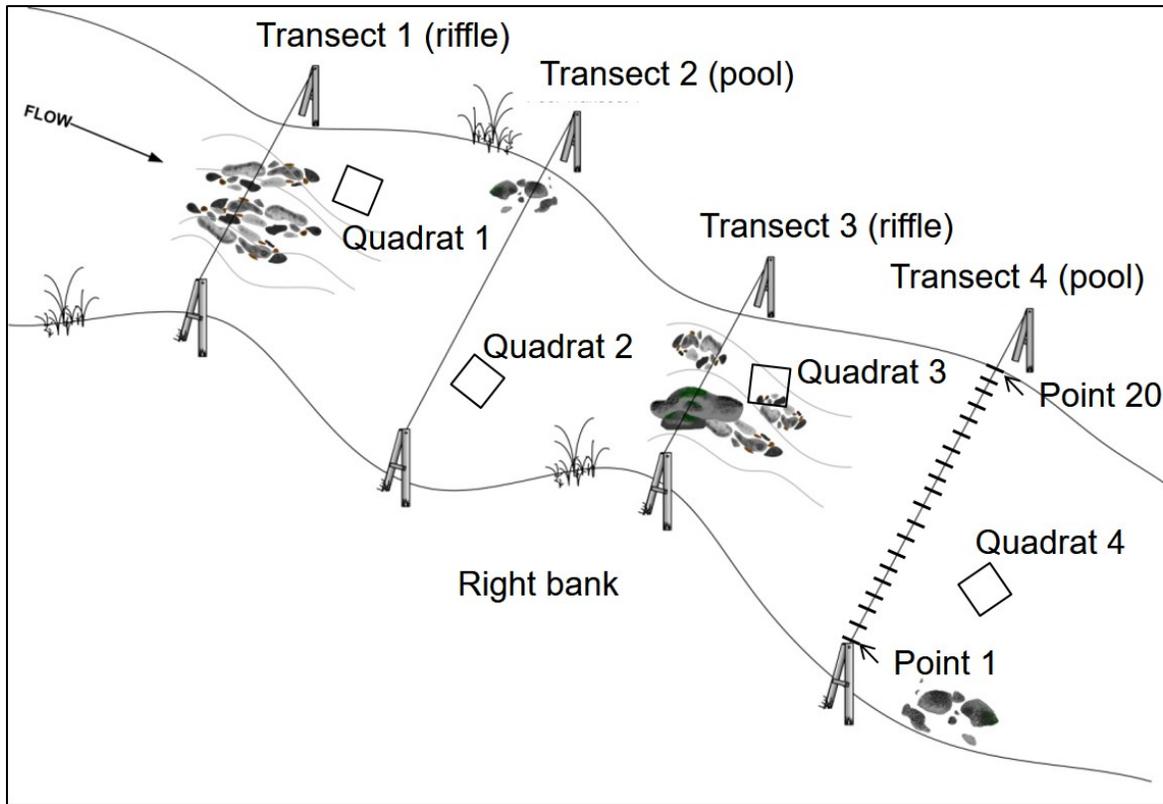


Figure 2: Aquatic plant sampling locations in wadeable streams, showing quadrat locations in proximity to established transects

### 3.2 Temporal Sampling Design

Aquatic plant clip harvest data are collected during biological and sediment chemistry sampling bout 2 (mid-summer). Data collection for this data product occurs within one day per bout at a given site. See NEON Aquatic Sampling Strategy (AD[05]), AOS Protocol and Procedure: Aquatic plant, bryophyte, lichen, and macroalgae sampling (AD[06]) for additional details.

Additional presence/absence plant point data are collected in this data product in lakes and rivers during sampling bouts 1 (spring) and 3 (autumn).

### 3.3 Sampling Design Changes

Prior to January 1, 2017, clip harvest samples and data were collected 3 times per year during all biological and sediment chemistry sampling bouts (bouts 1, 2, and 3). Sample collection was stopped during bouts 1 and 3 to avoid over-harvesting of aquatic vegetation.

### 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 Aquatic Plant Clip Harvest (DP0.20066.001) (AD[03]). All variables reported in the published data (L1 data) are also provided separately in the file, NEON Data Variables for Aquatic Plant Clip Harvest (DP1.20066.001) (AD[04]).

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 Aquatic Observation System (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.

### 3.5 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>

### 3.6 Spatial Resolution and Extent

The finest resolution at which spatial data are reported is near a single transect (wadeable streams) or at a randomized point (lakes and rivers) within a site. For example, data may be collected at a specific depth in the water column of a lake.

Lake and river spatial hierarchy = locationID (ID of sampling point where collection occurred) -> siteID (ID of NEON site) -> domainID (ID of a NEON domain)

Wadeable stream spatial hierarchy = locationID (ID of sampling quadrat near a named transect within site) -> siteID (ID of NEON site) -> domainID (ID of a NEON domain)

The basic spatial data included in the data downloaded include the targeted and sampled latitude and longitude of the randomized point in a lake or river. To obtain details about the fixed transect locations sampled in streams, including latitude, longitude and elevation of the named location, refer to the locations endpoint of the NEON API: <https://data.neonscience.org/data-api/endpoints/locations/>. Details about locations that have been updated or retired can also be access via the locations endpoint of the NEON API.

### 3.7 Associated Data Streams

In wadeable streams, data for the Aquatic plant bryophyte macroalgae point counts in wadeable streams product (DP1.20072.001) are collected at the same time as clip harvest data. The two data products may be linked through their shared **eventID** (or a combination of **siteID** and **collectDate**). In addition, morphospecies names and identifications for all aquatic plants products (including this one) are provided in the `apc_morphospecies` table, provided in the expanded package with downloads of either product and can be linked through **siteID**, **morphospeciesCreatedDate**, and **morphospeciesID**. Samples for the Aquatic plant bryophyte chemical properties (DP1.20063.001) data product are also generated from clip harvest subsamples, and may be linked through **sampleID**.

Data and maps produced by the Bathymetric and morphological maps data product (DP4.00132.001) may also be relevant to clip harvest data.

### 3.8 Product Instances

Derived products from aquatic plant, bryophyte, and macroalgae clip harvest are calculated per taxon, per point, per site/bout. The total number of bouts per site is 1 per year (3 bouts per year for data collected prior to January 1, 2017), with 10 clip harvest collections/observations per site per bout. The number of specimens produced depends on the density of vegetation at the site. Physical samples exist only if a given location contains plants, i.e. where **targetTaxaPresent** = "Y".

### 3.9 Data Relationships

A record in `apl_biomass`, `apl_taxonomyRaw`, or `apl_processedTaxonomy` must have a corresponding record in `apl_clipHarvest` describing field collection conditions, location, and metadata during sample collection. If **fieldID** is empty in `apl_clipHarvest`, there will be no additional records in the `apl_biomass`, `apl_taxonomyRaw`, or `apl_processedTaxonomy` 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.

`apl_clipHarvest.csv` -> One record is created for each clip harvest observation. If a sample is collected, a **fieldID** is created in this table and linked in subsequent data tables (`biomass.csv`). This table may also indicate the presence or absence of vegetation at 10 randomly selected points in lakes and rivers during Bouts 1 and 3, with no additional plant taxonomic records or samples generated after this field data table (e.g., no vegetation is collected for biomass analysis).

`apl_biomass.csv` -> One record (**sampleID**) is created for each subsample of a plant or bryophyte taxon or mixed macroalgae processed for dry mass and ash-free dry mass (AFDM) in the NEON domain lab. There may be multiple records per field sample created in `apl_clipHarvest`, depending on the taxonomic diversity of the collected sample. Data from this table are linked to the `clipHarvest` data table through the **fieldIDs** in each table.

`apl_taxonomyProcessed.csv` -> One record is created for each taxonomic group identified in a sample created in `apl_biomass`. Taxonomic identifications are made to the lowest practical taxonomist level.

The taxonomic nomenclature in this file has been standardized and desynonymized according to NEON’s master taxonomy for plants or NEON’s master taxonomy for algae. Data are linked to the biomass table through the **sampleIDs** in each table.

apl\_taxonomyRaw.csv -> One record is created for each taxonomic group identified in a sample created in apl\_biomass. Taxonomic identifications are made to the lowest practical taxonomist level. The taxonomic nomenclature in this file reflects the verbatim identifications provided by the external taxonomist and may contain synonyms. Data are linked to the biomass table through the **sampleIDs** in each table.

apc\_voucher.csv -> Plant voucher data may be collected at any time during the field season except during normal clip harvest or point count collection events, when field technicians see an appropriate specimen in the field. Data apply to the NEON aquatic site, and may be linked through **taxonID** and **siteID**. A **sampleID** is created during voucher field collection that can be linked to the external taxonomy data (apc\_voucherTaxonomyRaw and apc\_voucherTaxonomyProcessed).

apc\_voucherTaxonomyProcessed.csv -> One record is created for each physically collected **sampleID** that is identified to the lowest practical taxonomist level. The processed taxonomy table represents taxonomy from the domain support facility or the external lab that has been checked and standardized against the NEON taxonomy table for plants or algae. Data are linked to apc\_voucher through **sampleID**.

apc\_voucherTaxonomyRaw -> One record is created for each physically collected **sampleID** that is identified to the lowest practical taxonomist level. The raw taxonomy table represents taxonomy from the domain support facility or the external lab, and is not checked against the NEON taxonomy table for plants or algae. Data are linked to apc\_voucher through **sampleID**.

apc\_morphospecies.csv -> The morphospecies table is used both for this data product, and the point count data product (DP1.20072.001). Morphospecies are assigned when the specimen cannot be identified by the NEON field technician. Technicians may revisit this table to update **taxonID** data at a later time if they are able to positively identify the specimen. New morphospecies names are assigned per site each year. Not all morphospecies may be resolved. Join with apl\_biomass or apl\_taxonomyProcessed via the **morphospeciesID** and year of collection.

apl\_identificationHistory.csv -> One or more records expected per identificationHistoryID. Records are only created when data corrections to taxonomic identifications are made. If errors in identification are detected through QAQC processes after data publication, then corrected taxonomy will be provided in the apl\_taxonomyRaw and apl\_taxonomyProcessed tables. The apl\_identificationHistory table is populated with all prior names used for specimen(s) in the data product. When data are populated in the apl\_identificationHistory table, **identificationHistoryID** is used as a linking variable between the apl\_identificationHistory table and all other tables where updates were made.

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>.

## 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.

Prior to the 2022 data release, publication of species identifications were obfuscated to a higher taxonomic rank when the taxon was found to be listed as threatened, endangered, or sensitive at the state level where the observation was recorded. The state-level obfuscation routine was removed from the data publication process at all aquatic locations excluding sites located in D01, and data have been reprocessed to remove the obfuscation of state-listed taxa for all years. Federally listed threatened and endangered or sensitive species remain obfuscated at all sites and sensitive species remain redacted at National Park sites.

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

### 4.1 Plant Taxonomy

The master taxonomy for plants is the USDA PLANTS Database (USDA, NRCS. 2014. <https://plants.usda.gov>). Taxon ID codes used to identify taxonomic concepts in the NEON master taxonomy list are alphanumeric codes, 4-6 characters in length based on the accepted scientific name. Each code is composed of the first two letters of the genus, followed by the first two letters of the species and first letter of the terminal infraspecific name (if applicable) then, if needed, a tiebreaking number to address duplicate codes. Genus and family symbols are the first five (genus) or six (family) letters of the name, plus tiebreaking number (if needed). Symbols were first used in the Soil Conservation Service's National List of Scientific Plant Names (NLSPN) and have been perpetuated in the PLANTS system. The portions of the PLANTS Database included in the NEON plant master taxonomy list includes native and naturalized plants present in NEON observatory sampling area including the Lower 48 U.S. States, Alaska, Hawaii, and Puerto Rico. NEON plans to keep the taxonomy updated in accordance with USDA PLANTS Database starting in 2020 and annually thereafter.

The master taxonomy list for plants includes geographic range and nativity as described by the USDA PLANTS Database. A list for each NEON domain includes those species with ranges that overlap the domain as well as nativity designations - introduced or native - in that part of the range. Errors are generated if a species is reported at a location outside of its known range. If the record proves to be a reliable report, the master taxonomy table is updated to reflect the distribution change.

### 4.2 Algae Taxonomy

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 for algae 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.

### 4.3 Plant Morphospecies

Plants are identified by field ecologists to the lowest taxonomic level that they are able. An identification qualifier may be used to indicate if a field ecologist is uncertain with a particular level of identification. A morphospeciesID will be assigned in the field if the taxonomist cannot identify a specimen. Morphospecies names and descriptions are good for 1 year at a site, after that year, if the morphospecies has not been resolved, field ecologists will re-assign a new morphospeciesID. Taxa with morphospeciesIDs are sent to expert taxonomists for vascular and bryophyte identification, and can be linked to field data. NEON does not automatically resolve the scientificNames in the field data once expert taxonomy are returned. Field tables need to be joined with taxonomy tables by the data user using the morphospeciesID and year of collection to resolve unknown taxa. Any specimens sent to an expert taxonomist are later sent to the NEON Biorepository for curation and storage.

### 4.4 Identification History

Beginning in 2023, the `apl_identificationHistory` table was added to track any changes to taxonomic identifications that have been published in NEON data. Such taxonomic revisions may be necessary when errors are found in QAQC checks, or when evidence from genetic analysis of samples or re-analysis of archived samples indicate a revision is necessary. Requests for taxonomic changes are reviewed by NEON science staff. Proposed changes are evaluated based on evidence in the form of photographs, existing samples, genetic data, consultation with taxonomic experts, or range maps. Upon approval, the existing record in the `apl_taxonomyRaw` and `apl_taxonomyProcessed` tables, the `apc_voucherTaxonomyRaw` and `apc_voucherTaxonomyProcessed` tables, or `apc_voucher` table are updated with the new taxonomic information and a unique identifier is added to the **identificationHistoryID** field. A record with the same `identificationHistoryID` is created in the `apl_identificationHistory` table where the previous taxonomic information is archived along with the date the change was made.

## 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 data entry workflow for collecting aquatic plant clip harvest field data is diagrammed in Figure 3, and the lab workflow is diagrammed in Figure 4.

An additional set of constraints are implemented during the process of ingest into the NEON database. The product-specific data constraint and validation requirements built into data entry applications and database ingest are described in the document NEON Raw Data Validation for Aquatic Plant Clip Harvest (DPO.20066.001), 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[10]).

### 5.3 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.

### 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.

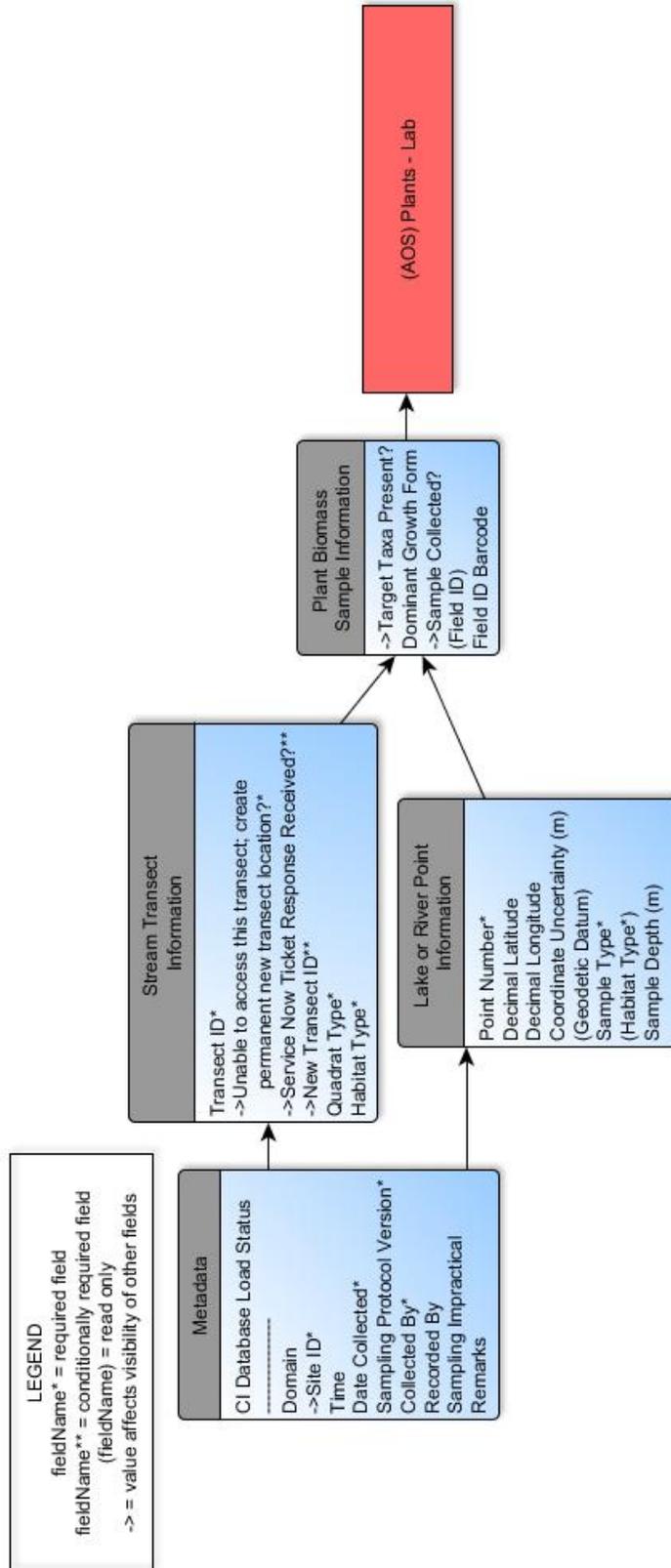


Figure 3: Schematic of the applications used by field technicians to enter aquatic plant clip harvest field data

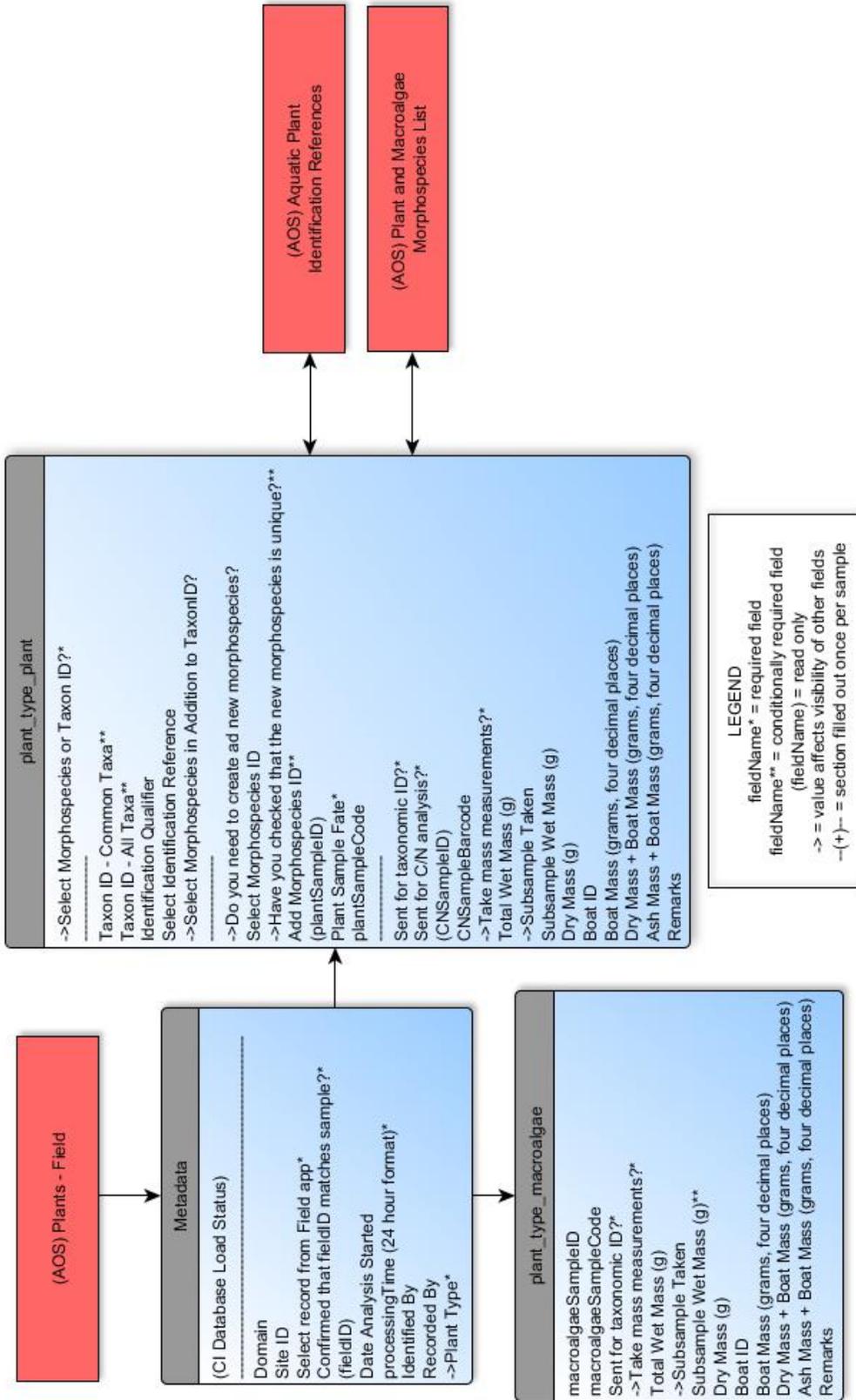


Figure 4: Schematic of the applications used by field technicians to measure aquatic plant biomass data and subsamples in the domain lab

Table 1: 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                 |
| dataQF    | implausible sampling location data | Spatial data associated with the randomized point in a lake or river is inaccurate (ie. outside of the reach boundary) and the true sampling location could not be determined |

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 aquatic plant clip harvest specimens conform to the current data quality standards used by practitioners. Ten percent of all samples are quality checked for taxonomic difference between two external taxonomists. These records are indicated by the field **qcTaxonomyStatus**, which indicates whether the specimen has been quality checked. Details on the external facility quality control can be found in the external lab SOP.

If errors in identification are detected through QAQC processes after data publication, then corrected taxonomy is provided in the `apl_taxonomyProcessed` and `apl_taxonomyRaw` tables for clip harvest data or the `apc_voucherTaxonomyProcessed`, `apc_voucherTaxonomyRaw`, and `apc_voucher` tables for voucher specimens, and previous taxonomic information is preserved in the `apl_identificationHistory` table (see Sections 3.9 Data Relationships and 4.4 Identification History above for more details).

## 6 REFERENCES

USDA, NRCS. 2014. The PLANTS Database (<http://plants.usda.gov>, 25 August 2014). National Plant Data Team, Greensboro, NC 27401-4901 USA.