



<i>Title:</i> NEON User Guide to Aquatic plant, bryophyte, lichen, and macroalgae point counts in wadeable streams (DP1.20072.001)	<i>Date:</i> 04/20/2023
<i>Author:</i> Stephanie Parker	<i>Revision:</i> C

# NEON USER GUIDE TO AQUATIC PLANT, BRYOPHYTE, LICHEN, AND MACROALGAE POINT COUNTS IN WADEABLE STREAMS (DP1.20072.001)

<b>PREPARED BY</b>	<b>ORGANIZATION</b>
Stephanie Parker	AQU
Caren Scott	AQU



<i>Title:</i> NEON User Guide to Aquatic plant, bryophyte, lichen, and macroalgae point counts in wadeable streams (DP1.20072.001)	<i>Date:</i> 04/20/2023
<i>Author:</i> Stephanie Parker	<i>Revision:</i> C

## CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	07/03/2017	Initial Release
B	05/25/2020	Included general statement about usage of neonUtilities R package and statement about possible location changes. Updated taxonomy information. Add morphospecies clarification and updated spatial figure to indicate changes to lake littoral and inflow/outflow sensor locations.
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	04/20/2023	Added lake and river sites to voucher and morphospecies tables.



## 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 . . . . .	4
3.3	Variables Reported . . . . .	4
3.4	Temporal Resolution and Extent . . . . .	5
3.5	Spatial Resolution and Extent . . . . .	5
3.6	Associated Data Streams . . . . .	5
3.7	Product Instances . . . . .	6
3.8	Data Relationships . . . . .	6
3.9	Special Considerations . . . . .	7
<b>4</b>	<b>TAXONOMY</b>	<b>7</b>
4.1	Plant Taxonomy . . . . .	8
4.2	Algae Taxonomy . . . . .	8
<b>5</b>	<b>DATA QUALITY</b>	<b>9</b>
5.1	Data Entry Constraint and Validation . . . . .	9
5.2	Automated Data Processing Steps . . . . .	9
5.3	Data Revision . . . . .	13
5.4	Quality Flagging . . . . .	13
5.5	Analytical Facility Data Quality . . . . .	13
<b>6</b>	<b>REFERENCES</b>	<b>13</b>



<i>Title:</i> NEON User Guide to Aquatic plant, bryophyte, lichen, and macroalgae point counts in wadeable streams (DP1.20072.001)	<i>Date:</i> 04/20/2023
<i>Author:</i> Stephanie Parker	<i>Revision:</i> C

## LIST OF TABLES AND FIGURES

Table 1	Descriptions of the dataQF codes for quality flagging . . . . .	13
Figure 1	Aquatic plant sampling locations in wadeable streams, showing quadrat locations in proximity to established transects . . . . .	4
Figure 2	Schematic of the field application used by field technicians to enter wadeable stream point count field data . . . . .	10
Figure 3	Schematic of the morphospecies creation app used by field technicians to assign morphospecies to unknown aquatic plant taxa . . . . .	11
Figure 4	Schematic of the applications used by field technicians to opportunistically collect voucher specimens for taxonomist identification . . . . .	12

## 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 point count 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, lichen, and macroalgae point counts in 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 Aquatic plant, bryophyte, lichen, and macroalgae point counts in wadeable streams (DP1.20072.001) (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, bryophyte, lichen, and macroalgae point counts in wadeable streams (DPO.20072.001) (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]	Available on NEON data portal	NEON Ingest Conversion Language Function Library
AD[11]	Available on NEON data portal	NEON Ingest Conversion Language
AD[12]	Available with data download	Categorical Codes csv

### 2.2 Acronyms

Acronym	Definition
USDA	US Department of Agriculture
USEPA	US Environmental Protection Agency
USGS	US Geological Survey

### 3 DATA PRODUCT DESCRIPTION

Point counts of aquatic plants, bryophytes, lichens, and macroalgae include taxonomic identification and relative abundance of plants. These data address the NEON Grand Challenge areas of Biodiversity, Biogeochemistry, and Invasive Species, as well provide important data on the macrophyte community composition in wadeable streams, which may be used to assess the status of the aquatic ecosystem. Aquatic vegetation is sampled three times per year at each NEON aquatic site (AD[05]). Sampling dates are based on a combination of variables, including hydrology in streams, accumulated degree days (temperature), and riparian greenness (phenology). For additional information see the sampling design NEON Aquatic Sampling Strategy (AD[05]) and protocol AOS Protocol and Procedure: Aquatic plant, bryophyte, lichen, and macroalgae sampling (AD[06]).

Point transect data are only collected at wadeable streams, however, this data product includes voucher and morphospecies records for lake and river sites.

#### 3.1 Spatial Sampling Design

Point count data are collected at 10 permanently-marked locations per wadeable stream site. Each transect (Figure 1) is established during initial sampling at the site. Five transects are established in the dominant habitat type, and five 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. Only if the geomorphology of the site changes significantly will a transect re-established. 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/>

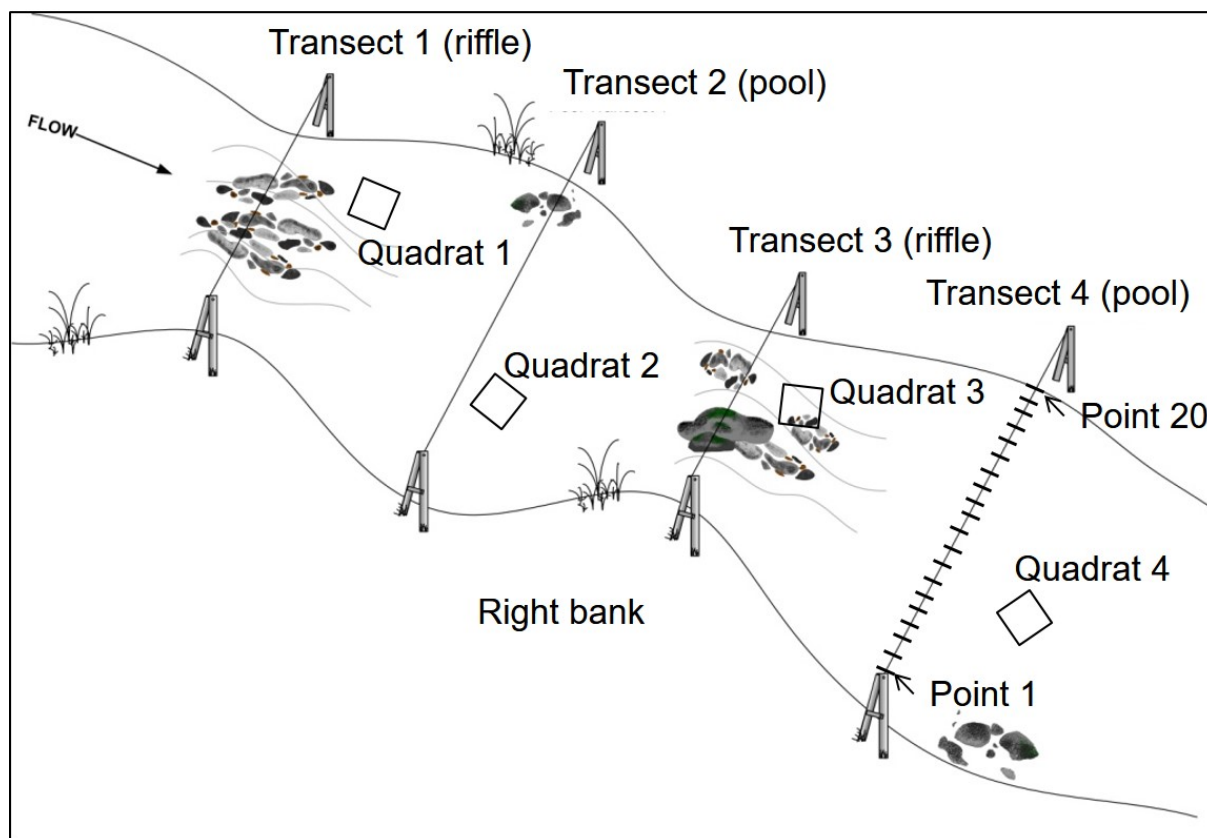


Figure 1: Aquatic plant sampling locations in wadeable streams, showing quadrat locations in proximity to established transects

### 3.2 Temporal Sampling Design

Aquatic point count data are collected during all three biological and sediment chemistry sampling bouts (roughly spring, summer, and fall). Data collection for this data product occurs within one day during a sampling bout at a given site. The number of specimens produced depends on the density of vegetation at the site.

### 3.3 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, bryophyte, lichen, and macroalgae point counts in wadeable streams (DPO.20072.001) (AD[03]). All variables reported in the published data (L1 data) are also provided separately in the file, NEON Data Variables for Aquatic plant, bryophyte, lichen, and macroalgae point counts in wadeable streams (DP1.20072.001) (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 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.

### 3.4 Temporal Resolution and Extent

The finest resolution at which temporal data are reported is at **collectDate**, the date and time of day when the sample was collected in the field.

The finest temporal resolution that Aquatic Plant Bryophyte Macroalgae Point Count data will be tracked is at the level of an individual observation per sampling bout. The total number of bouts per year is 3 per aquatic site. Each sampling bout occurs within one day at a given site. The number of observations produced depends on the density of vegetation at the site. Data collected for Aquatic plant bryophyte macroalgae clip harvest (NEON.DOM.SITE.DP0.20066) are collected at the same time.

### 3.5 Spatial Resolution and Extent

The finest resolution at which spatial data are reported is near a single transect in wadeable streams).

locationID (ID of 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 latitude, longitude, and elevation of the named location at the aquatic site (e.g., the permanent transect location in a wadeable stream).

### 3.6 Associated Data Streams

Data for the Aquatic plant bryophyte macroalgae clip harvest product (DP1.20066.001) are collected alongside point count data during bout 2. 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**.

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

### 3.7 Product Instances

There are a maximum of 3 point count bouts per year per site, with counts occurring at each of 10 transects. The number of points observed during each sampling event depends on the width of the stream.

### 3.8 Data Relationships

A record in `apc_perTaxon`, `apc_taxonomyRaw`, or `apc_taxonomyProcessed` must have a corresponding record in `apc_pointTransect`. Data in `apc_pointTransect` are field data that apply to the transect as a whole. A record in `apc_pointTransect` exists for each point visited along a stream transect, whether a plant taxon was present or not; `apc_perTaxon` records are only created for points where plants are found. `apc_taxonomy` records exist only where physical samples have been collected as part of `apc_perTaxon` sampling (non-null **sampleID**). 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.

`apc_pointTransect.csv` -> Data are recorded for all points surveyed along the wadeable stream transect in this table.

`apc_perTaxon.csv` -> If **targetTaxaPresent** = "Yes" in `apc_pointTransect`, a record will be entered here. All vegetation recorded at the point are recorded here, typically that is for one taxon, but may be for multiple taxa if there is vertical stacking of plants at the point. Data are linked to `apl_pointTransect` through **namedLocation** and **pointID**.

`apc_taxonomyProcessed.csv` -> One record is created for each physically collected **sampleID** in `apc_perTaxon` 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_perTaxon` through **sampleID**.

`apc_taxonomyRaw` -> One record is created for each **sampleID** in `apc_perTaxon` that is identified to the lowest practical taxonomist level. The raw taxonomy table represents taxonomy from the domain support facility or the external lab. The taxonomic nomenclature in this file reflects the verbatim identifications provided by the external taxonomist and may contain synonyms. Data are linked to `apc_perTaxon` through **sampleID**.

`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`). Vouchers from lake and river sites are also included in this table.

`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**. Voucher data from lake and river sites are also included in this table.

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**. Voucher data from lake and river sites are also included in this table.

apc\_morphospecies.csv -> The morphospecies table is used both for this data product, and the clip harvest data product (DP1.20066.001). 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 apc\_perTaxon or apc\_taxonomyProcessed via the **morphospeciesID** and year of collection. Morphospecies records from lake and river sites are also included in this table.

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

### 3.9 Special Considerations

Percent cover may be calculated using the equation from Bowden et al. 2006:

$$percentCover_i = \frac{N_i}{N_t} \times 100 \quad (1)$$

Where  $N_i$  is the number of observed points in a transect that match class type “i” (i.e., a particular **taxonID**) and  $N_t$  is the total number of points observed in the transect. This calculation can generate percent cover values >100% if there is vertical stacking of plants.

## 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](http://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.

**##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.

## 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 2, 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, bryophyte, lichen, and macroalgae point counts in wadeable streams (DP0.20072.001), 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 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]).

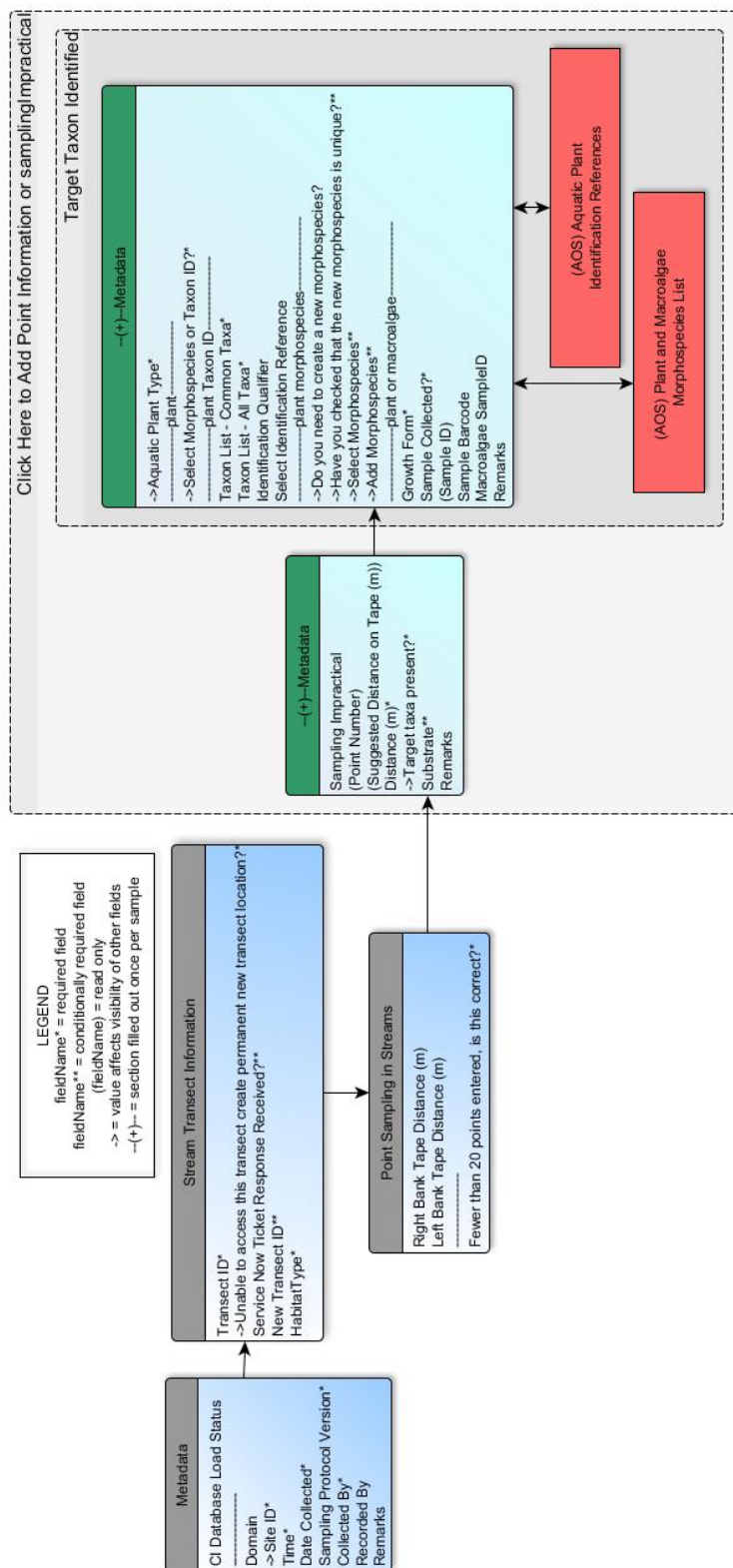


Figure 2: Schematic of the field application used by field technicians to enter wadeable stream point count field data

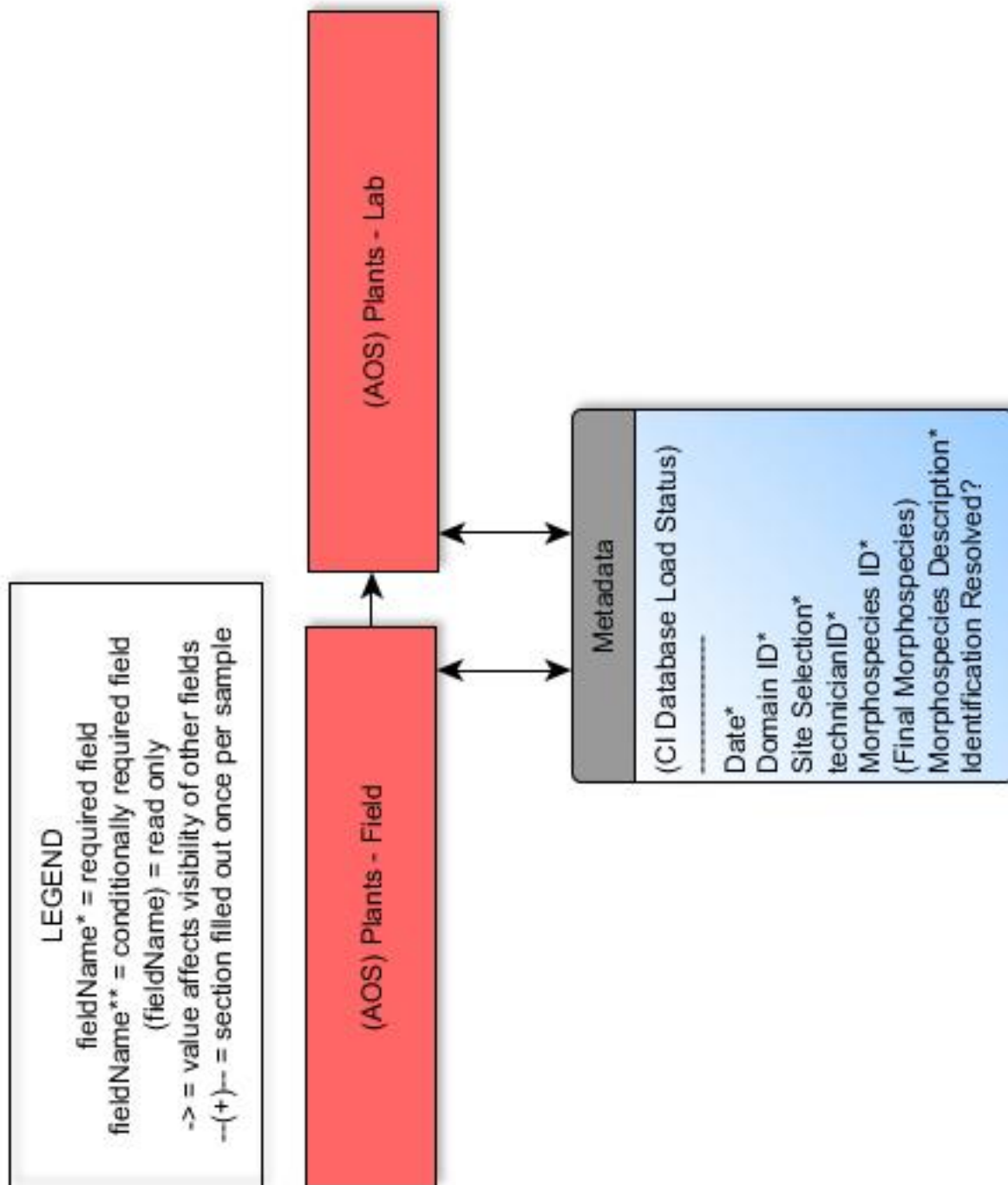


Figure 3: Schematic of the morphospecies creation app used by field technicians to assign morphospecies to unknown aquatic plant taxa



**LEGEND**

fieldName\* = required field  
 fieldName\*\* = conditionally required field  
 (fieldName) = read only  
 -> = value affects visibility of other fields

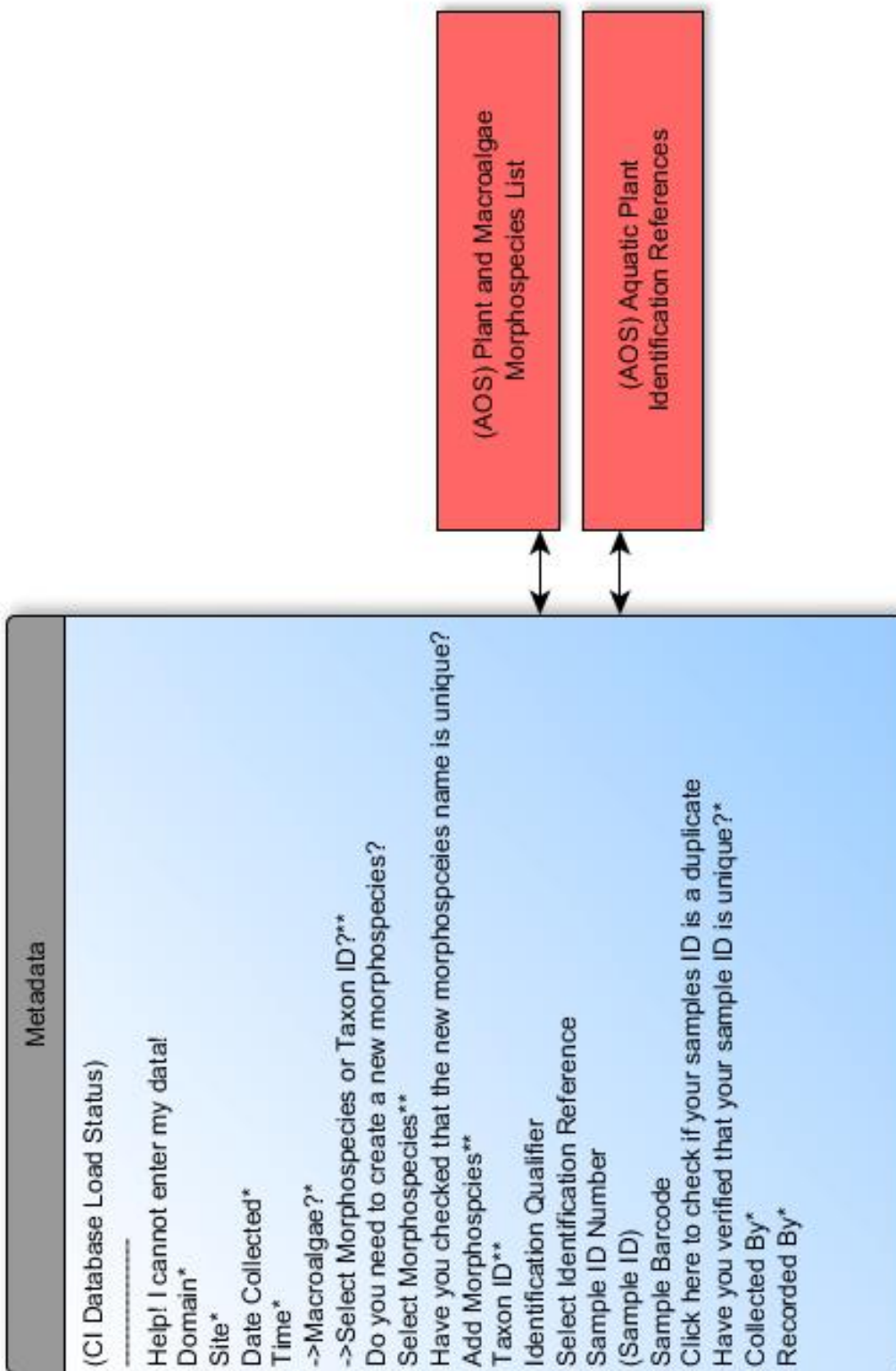


Figure 4: Schematic of the applications used by field technicians to opportunistically collect voucher specimens for taxonomist identification



### 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.

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

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 point count 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.

## 6 REFERENCES

Bowden, W.B., J.M. Glime, and T. Riis. 2006. Macrophytes and Bryophytes. Pages 381-414 in F.R. Hauer and G.A. Lamberti, editors. *Methods in Stream Ecology*, Second Edition. Elsevier, Inc., Boston, Massachusetts, USA.

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