NEON USER GUIDE TO AQUATIC PLANT BRYOPHYTE MACROALGAE CLIP HARVEST (NEON.DP1.20066)

PREPARED BY	ORGANIZATION	DATE
Stephanie Parker	AQU	07/28/2017
Caren Scott	AQU	07/28/2017

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	07/03/2017	Initial Release

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

1 DESCRIPTION

1.1 Purpose

This document provides an overview of the data included in this NEON Level 1 data product, the quality controlled product generated from raw Level 0 data, and associated metadata. In the NEON data products framework, the raw data collected in the field, for example, 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 (NEON.DP1.20066) (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 (NEON.DP0.20066) (AD[03]), provided in the download package for this data product. Please note that raw data products (denoted by 'DP0') may not always have the same numbers (e.g., '10033') as the corresponding L1 data product.

2 RELATED DOCUMENTS AND ACRONYMS

2.1 Associated Documents

AD[01]	NEON.DOC.000001	NEON Observatory Design (NOD) Requirements
AD[02]	NEON.DOC.002652	NEON Level 1, Level 2 and Level 3 Data Products Catalog
AD[03]	NEON.DP0.20066.001 _dataValidation.csv	NEON Raw Data Validation for Aquatic Plant Clip Harvest (NEON.DP0.20066)
AD[04]	NEON.DP1.20066.001 _variables.csv	NEON Data Variables for Aquatic Plant Clip Harvest (NEON.DP1.20066)
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]	OS_Generic_Transitions .pdf	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[10]	Nicl Language.pdf	NEON's Ingest Conversion Language (NICL) specifications

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 non-wadeable streams (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). If it is not possible to separate each transect by a different habitat type, transects may be spaced at least 10m 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.

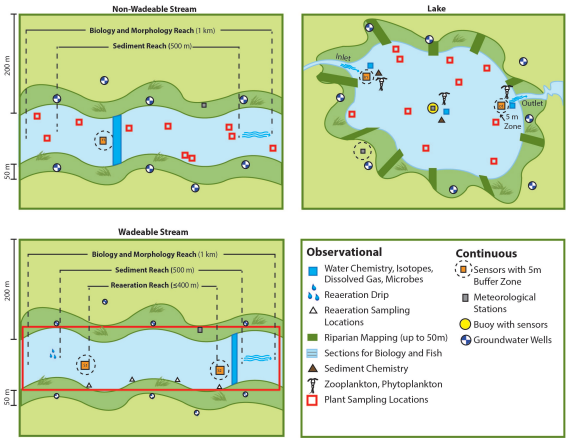

Figure 1: Aquatic plant sampling locations in all types of aquatic sites (non-wadeable streams/rivers, wadeable streams, and lakes).

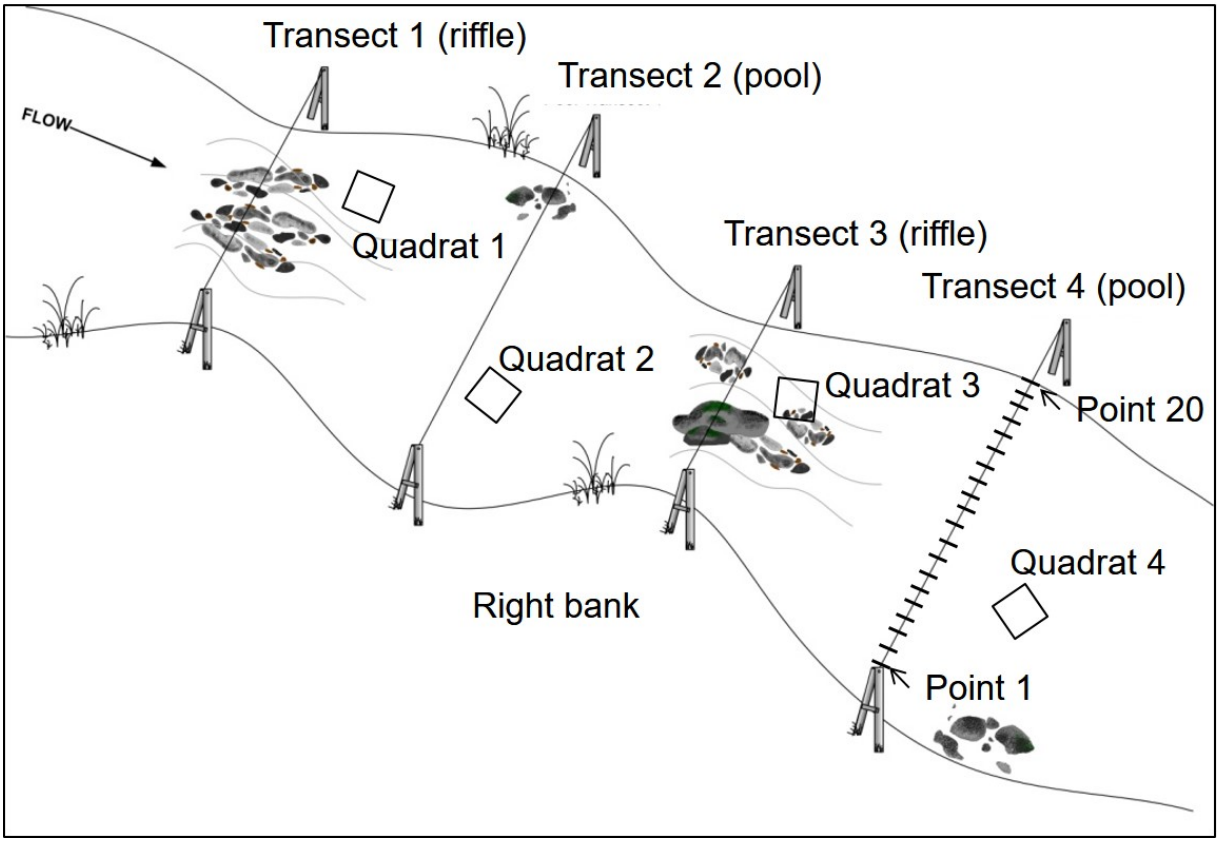

Figure 2: Aquatic plant sampling locations in wadeable streams, showing quadrat locations in proximity to established transects

3.2 Temporal Sampling Design

Aquatic plant clip harvest data are collected during biological and sediment chemistry sampling bout 2 (mid-summer). Prior to January 1, 2017, data were collected 3 times per year during all biological and sediment chemistry sampling bouts (bouts 1, 2, and 3). Data collection for this data product occurs 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.

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 Clip Harvest (NEON.DP0.20066) (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 (NEON.DP1.20066) (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 Terrestrial Observation System (TOS) 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 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

3.5 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.

Lakes and non-wadeable streams (rivers) spatial hierarchy = locationID (ID of sampling point where collection occurred) -> siteID (ID of NEON site) -> domainID (ID of a NEON domain)

Wadeable streams spatal 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 latitude, longitude, and elevation of the named location at the aquatic site (e.g., the permanent transect location in a wadeable stream) or the latitude and longitude of the randomized point in a lake or river.

3.6 Associated Data Streams

In wadeable streams, data for the Aquatic plant bryophyte macroalgae point counts in wadeable streams product (NEON.DP1.20072) 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 (NEON.DP1.20063) 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 (NEON.DP4.00132) may also be relevant to clip harvest data.

3.7 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 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".

3.8 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 **fieldID**s 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 **sampleID**s 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 **sampleID**s in each table.

apc_morphospecies.csv - > The morphospecies table is used both for this data product, and the point count data product (NEON.DP1.20072). 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.

3.9 Special Considerations

4 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 (NEON.DP0.20066), 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 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]).

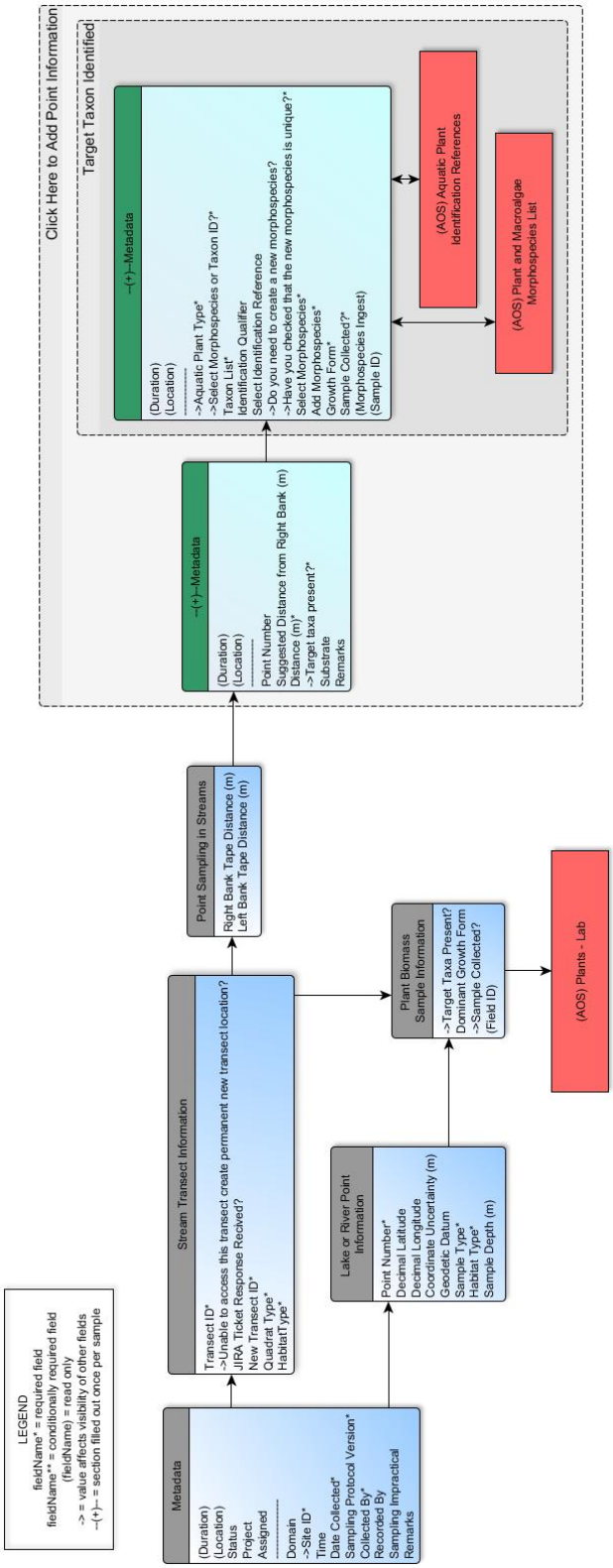

Figure 3: Schematic of the applications used by field technicians to enter aquatic point count and clip harvest field data

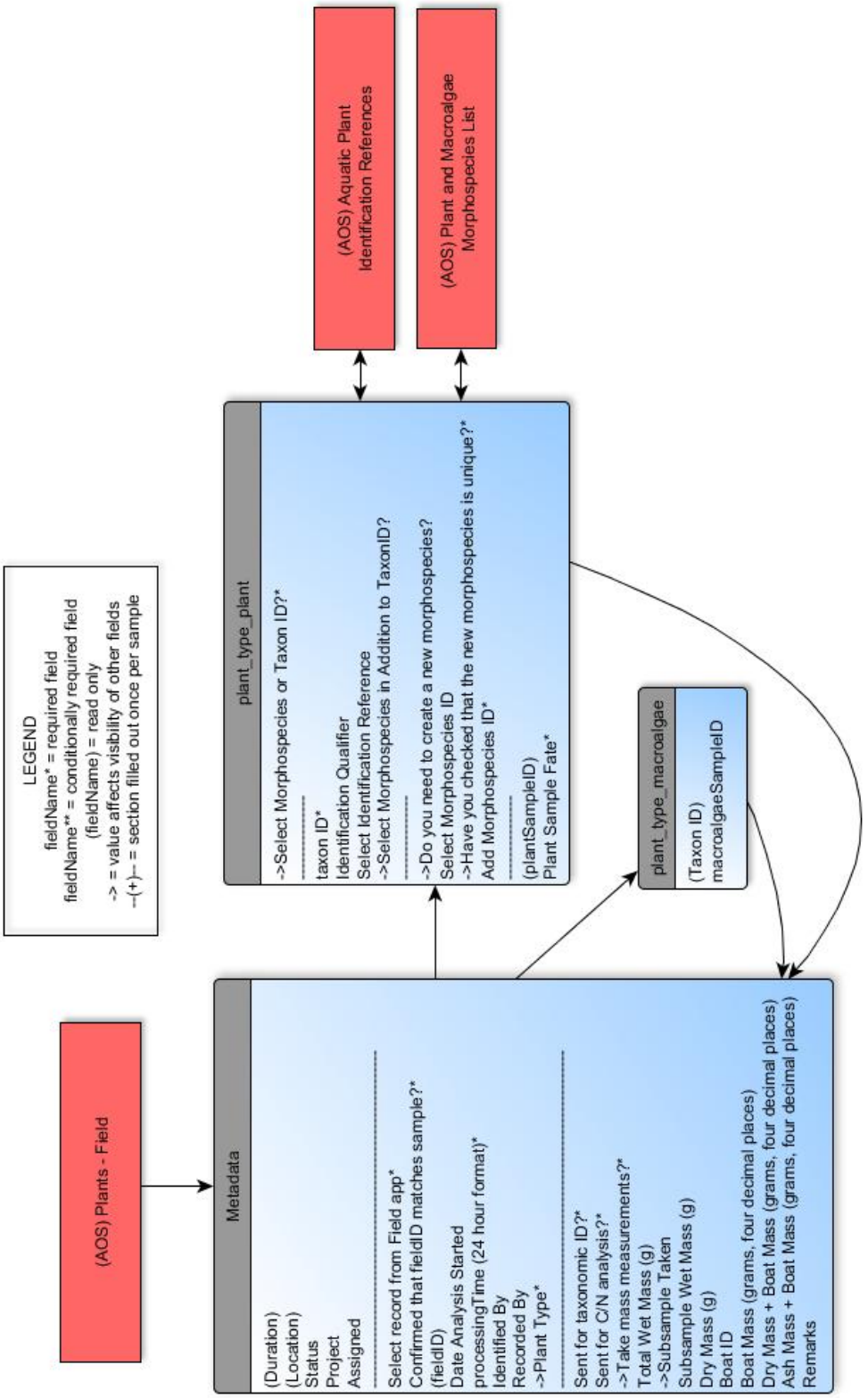

Figure 4: Schematic of the applications used by field technicians to measure aquatic plant biomass data in the domain lab