NEON USER GUIDE TO HERBACEOUS CLIP HARVEST (NEON.DP1.10023)

PREPARED BY	ORGANIZATION	DATE
Katherine Jones	FSU	05/17/2017
Courtney Meier	FSU	05/17/2017

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	fillIn: mm/dd/yyyy	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 dry weights of litter functional groups 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 Herbaceous clip harvest - the dry weight of annual production of herbaceous plants clipped from randomly placed clip cells in terrestrial plots - 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 Publication Workbook for TOS Herbaceous Biomass: QA/QC of Raw Field and Lab Data (NEON.DP1.10023) (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 TOS Protocol and Procedure: Measurement of Herbaceous Biomass (AD[07]). The raw data that are processed in this document are detailed in the file, NEON Raw Data Ingest Workbook for TOS Herbaceous Biomass (NEON.DP0.10023) (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.000913	TOS Science Design for Spatial Sampling
AD[03]	NEON.DOC.002652	NEON Level 1, Level 2 and Level 3 Data Products Catalog
AD[04]	NEON.DP0.001920.001 _dataValidation.csv	NEON Raw Data Ingest Workbook for TOS Herbaceous Biomass (NEON.DP0.10023)
AD[05]	NEON.DP1.001931.001 _variables.csv	NEON Data Publication Workbook for TOS Herbaceous Biomass: QA/QC of Raw Field and Lab Data (NEON.DP1.10023)
AD[06]	NEON.DOC.000914	TOS Science Design for Plant Biomass, Productivity, and Leaf Area Index
AD[07]	NEON.DOC.014037	TOS Protocol and Procedure: Measurement of Herbaceous Biomass
AD[08]	NEON.DOC.000913	TOS Science Design for Spatial Sampling
AD[09]	NEON.DOC.000913	TOS Science Design for Plant Diversity
AD[10]	NEON.DOC.000008	NEON Acronym List
AD[11]	NEON.DOC.000243	NEON Glossary of Terms
AD[12]	NEON.DOC.014037	TOS Protocol and Procedure: Measurement of Herbaceous Biomass
AD[13]	NEON.DOC.001709	TOS Protocol and Procedure: Measurement of Bryophyte Productivity
AD[14]	OS_Generic_Transitions .pdf	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[15]		NEON's Ingest Conversion Language (NICL) specifications

2.2   Acronyms

Acronym	Definition
ANPP	Annual Net Primary Productivity

3   DATA PRODUCT DESCRIPTION

The herbaceous clip harvest data products provide mass data for herbaceous plant functional groups from individual sampling bouts. All masses reported following processing are reported at the spatial resolution of one clip strip and the temporal resolution of a single collection event.

Mass data for each collection event are reported to an accuracy of 0.01 grams. Masses < 0.01g are also reported and indicate presence of an identifiable sample present at masses below accuracy of the scales used to weigh samples.

For one sampling bout per growing season, the aboveground dry mass production associated with key herbaceous functional groups will be provided. These functional groups are defined as:

- Bryophytes (for which annual production is discernable)
- C3 cool-season graminoids
- C4 warm-season graminoids
- Leguminous forbs
- Non-leguminous forbs
- Woody-stemmed shrubs with diameter at decimeter height < 1 cm

All data products include spatial and temporal sampling details, as well as required metadata. For additional information on the sampling design and the associated protocol, see TOS Science Design for Plant Biomass, Productivity, and Leaf Area Index (AD[06]), and the TOS Field and Lab Protocol for Herbaceous Plant Biomass (AD[07]). Following data collection in the field and the lab, several standard QA/QC procedures are implemented, as described in Section 5 of this document, in order to produce quality-checked Level 1 herbaceous clip harvest data products.

Herbaceous clip harvest data may be used to estimate annual Aboveground Net Primary Productivity (ANPP) and aboveground biomass at plot, site, and continental scales. They also provide essential data for understanding vegetative carbon fluxes over time.

3.1   Spatial Sampling Design

The Herbaceous Clip Harvest sampling is implemented in Tower plots all terrestrial NEON sites and in Distributed plots at terrestrial NEON sites with non-forest, herbaceous dominated, National Land Cover Database (NLCD) Classification. Sampling occurs at all Tower plots, and in a maximum of twenty randomly selected Distributed plots, with $\geq$ 50% herbaceous plant cover, per site. In Distributed plots, the maximum potential sampling effort (n=20 plots) corresponds to sampling in ½ to ⅔ of the established Distributed plots per site. The Distributed plots at which the herbaceous clip harvest protocol is performed are chosen randomly in order to create an unbiased estimate of herbaceous biomass at the site scale (see AD[08] for details).

In non-agricultural systems, for both plot types, the clip harvest is performed in 0.1 m x 2 m clip strips that exist within 0.5 m x 3 m gridded clip "cells" (Figure 1 and Figure 2). For typical sampling bouts, one gridded clip-area "cell" is randomly chosen per 400 m^2 plot or subplot. However, at sites managed for grazing, an additional grid cell is placed within a grazing exclosure, resulting in two clip-strips harvested per sampling bout. Grid cells that overlap with 25m^2 nested subplots (used for seedling and sapling counts) may be used for clip harvests, but those that overlap with 1m^2 and 10m^2 nested subplots (used for diversity sampling) are never clipped. This ensures

that generation of herbaceous plant productivity data minimally affects the quality of plant biodiversity data. See TOS Science Design for Plant Biomass, Productivity, and Leaf Area Index (AD[06]), TOS Protocol and Procedure: Measurement of Herbaceous Biomass (AD[07]), TOS Science Design for Plant Biomass, Productivity, and Leaf Area Index (AD[08]) and TOS Science Design for Plant Diversity (AD[09]) for further details.

Placement of clip strips is random and utilizes the randomized list of grid cell locations being utilized for herbaceous clip harvest and bryophyte sampling (AD[12], AD[13]). This is an ordered list of all available sampling locations within a given plot, from which a **clipCell** is removed after being utilized for sampling in any one of the three sampling protocols. Removal precludes consideration of a **clipCell** for future sampling under all protocols.

Clip strip dimensions and orientation at agricultural sites differs from clip strips at non-agricultural sites. In all agricultural systems, clip strips are oriented perpendicular to crop rows. Plots planted with barley and wheat utilize clip stips that are 2.0 m x 0.5 m; plots planted with corn, sorghum and soybeans clip biomass within 1.5 m x 0.65 m clip strips.
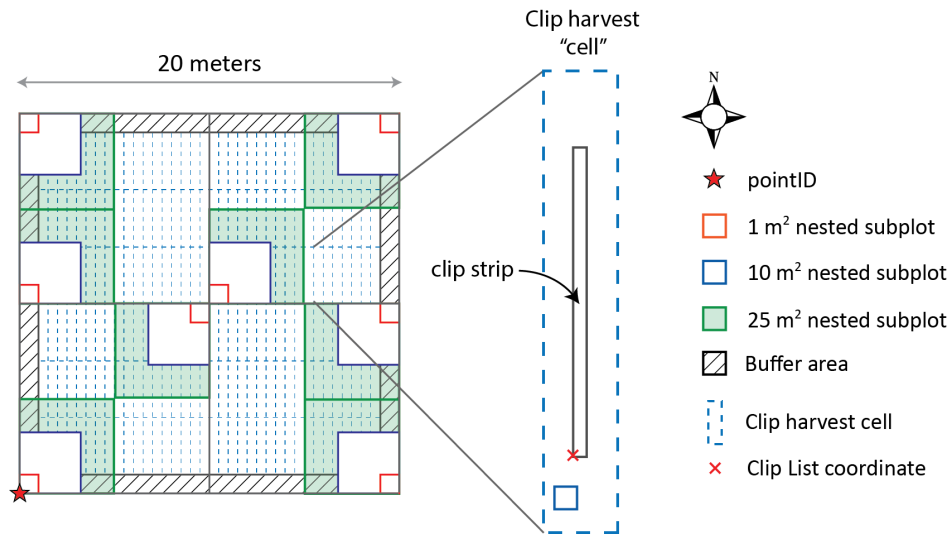
Figure 1: A 20m x 20m (400m^2) NEON plot showing potential clip harvest "cells" (dashed blue lines). Cells overlapping 10m^2 nested subplots are not sampled for herbaceous biomass. When grazing management is absent, one randomly selected grid cell is chosen for harvest per sampling bout. At sites managed for grazing, two grid cells are sampled per bout. The pointID (red star) is used to calculate the location in UTM of a given Clip List coordinate (red 'x').
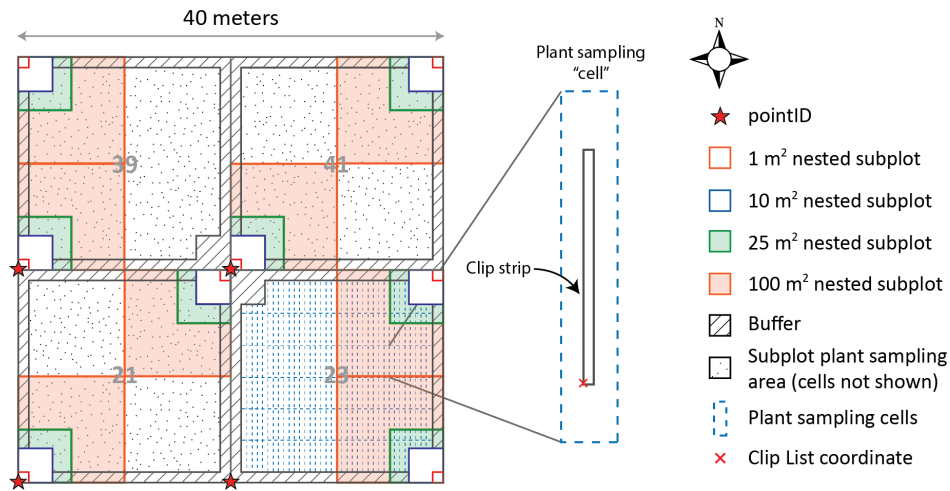
Figure 2: A 40m x 40m (1600 m^2) NEON Tower plot showing clip-harvest cells within a 400 m^2 subplot. Cells overlapping 10m^2 nested subplots are not sampled for herbaceous biomass. Clip harvest samples are generated from two randomly selected subplots per sampling bout, but for clarity, cells are not shown in 3 of 4 subplots. At sites managed for grazing, two grid cells per selected subplot are sampled per bout. For each subplotID (grey numbers), the pointID at the SW corner (red stars) is used to calculate the location in UTM of a given Clip List coordinate (red 'x').

3.2   Temporal Sampling Design

NEON field and lab technicians will generate the herbaceous clip harvest data product at least once per growing season for each site. However, the temporal sampling resolution depends on plot type: While Distributed plots located throughout the site are only ever harvested once per growing season, Tower plots may be sampled more than once per growing season if certain criteria are met. The default sampling frequency for Tower plots is once per year, but the following factors may result in additional sampling bouts per growing season, and hence additional instances of the data product, and are assessed on a per site basis:

- Sites with distinct early-season (C3 plants) and late-season biomass peaks (C4 plants) are clip-harvested twice per growing season.
- Sites actively managed for grazing are clip-harvested every 4 weeks.
- Sites with continuous growth and decomposition (e.g., tropical and sub-tropical grasslands) are harvested more frequently than sites with periods of senescence.

Date values are assigned a boutNumber in the field, and the boutNumber, year of sampling, and siteID in which the clip harvest occurred are used to create a unique eventID associated with each herbaceous dry weight mass value. Harvested herbaceous biomass is sorted to functional group for a minimum of one sampling bout per growing season.

3.3 Variables Reported

All variables reported from the field or laboratory technician (L0 data) are listed in the file, NEON Raw Data Ingest Workbook for TOS Herbaceous Biomass (NEON.DP0.10023) (AD[04]). All variables reported in the published data (L1 data) are also provided separately in the file, NEON Data Publication Workbook for TOS Herbaceous Biomass: QA/QC of Raw Field and Lab Data (NEON.DP1.10023) (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 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 Spatial Resolution and Extent

The overall spatial hierarchy of the herbaceous clip harvest data product is subtly different for Tower plots compared to Distributed plots. For Tower plots that are 40 m x 40 m and larger, the gridded clip harvest locations exist within subplotIDs within each plotID (Figure 2):

Tower plots (≥ 1600m^2): clipID → subplotID → plotID → habitat (NLCD veg class) → siteID → domainID

All other plots:clipID → plotID → habitat → siteID → domainID

The basic spatial data included in the data download include the latitude, longitude, and elevation of the centroid of the plot where sampling occurred + associated uncertainty due to GPS error and plot width. Shapefiles of all NEON Terrestrial Observation System sampling locations can be found in the Document Library: http://data.neonscience.org/documents.

To derive a more precise estimate of the location of each trap, there are two options:

- Use the def.calc.geo.os function from the geoNEON package, available here: https://github.com/NEONScience/NEON-geolocation

- Or follow these steps to perform the same calculation:

1. The namedLocation field in the data is the named location of the plot; more precise geographic data require the named location of the subplot. Construct the named location of the subplot of each record in hbp_perbout by concatenating the fields for namedLocation and subplotID as: namedLocation + '.' + subplotID, e.g. subplotID '41' of namedLocation 'HARV_052.basePlot.hbp' has a complete named location of 'HARV_052.basePlot.hbp.41'.

2. Use the API (http://data.neonscience.org/api; e.g. http://data.neonscience.org/api/v0/locations/HARV_052.basePlot.hbp.41) to query for elevation("locationElevation"), easting("locationUtmEasting"), nor-

thing("locationUtmNorthing"), coordinateUncertainty ("Value for Coordinate uncertainty"), elevationUncertainty ("Value for Elevation uncertainty"), and utmZone ("locationUtmZone") as inputs to the next step.

3. Calculate the clipCellNumber of each trap (last 3 digits of clipID).

4. Use the clip cell lookup table, available here: http://data.neonscience.org/api/v0/documents/ clipCellNumber_lookup (clicking on link will initiate download), to find the offsets for each clipCellNumber and subplot (=pointID), and use the offsets to adjust the UTM values downloaded in step 2.

5. Increase coordinateUncertainty by an appropriate amount to account for error introduced by navigating within plots. Technicians may shift elevated trap up to 1 meter to either the north or south within a selected clip cell to avoid obstacles. Additional error may be introduced due to tape stretching to navigate to locations within plots.

3.5   Temporal Resolution and Extent

The finest temporal resolution at which instances of the herbaceous clip harvest data product are tracked is the date of harvest, **collectDate**, for a given plot or subplot.

<div align="center">

collectDate (date an individual strip was clipped) → **annual production**

</div>

The NEON Data Portal currently 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: (LINK TBD)

3.6   Associated Data Streams

Additional information about woody vegetation present in the plot, if relevant, is available in the NEON Woody Plant Vegetation Structure data product (NEON.DP1.10098). Users are advised to leverage data from vegetation structure to provide context to clip harvest data. Additional information about litter production in the plot is available in the NEON Litterfall and Fine Woody Debris data product (NEON.DP1.10033).
–>

The herbaceous clip harvest data product is used as an input to the herbaceous production data product, the latter of which is detailed in NEON ATBD: TOS Herbaceous Biomass – Aboveground Productivity Calculations (AD[12]). Separate, additional herbaceous clip harvest sampling events are carried out to generate samples analyzed for isotopes and chemistry. The results of these analyses can be found in NEON's plant foliar stable isotopes (NEON.DOM.SITE.DP1.10053) and plant foliar chemical properties (NEON.DOM.SITE.DP1.10051) data products.

3.7   Product Instances

There are a maximum of 2 herbaceous clip harvest bouts per year at ungrazed sites, with samples collected from no more than 40 clip strips in Tower plots and no more than 20 clip strips in Distributed plots, for a total of 60 or

fewer clip strips per site. Each clip yields 1-5 herb group samples. No ungrazed site should ever exceed 500 data product instances in a single calendar year.

At grazed sites, Tower plots are sampled no more than 8 times per year. There are 2 clip strips per plot/subplot resulting in up to 80 clip strips in Tower plots. Only 1 bout per year is sorted to herb group resulting a maximimum of 960 samples generated in Tower plots. In years scheuduled for sampling in Distributed plots, 20 clip strips yielding 1-5 mass samples each are also clipped. No grazed site should ever exceed 1160 data product instances in a single calendar year.

3.8   Data Relationships

A selected clipID is, ideally, only ever harvested once. The unique sampleID is generated from the clipID and the collectDate. Each sampleID created in the field sampling activity, hbp_perbout, results in 1-5 child records, sub-sampleIDs, in the hbp_massdata table. Duplicates and/or missing data may exist where protocol and/or data entry abberations have occurred; users should check data carefully for anomalies before joining tables.

hbp_perbout.csv → One record expected per clipID per collectDate, producing only one instance of sampleID for all time in the hbp_perbout table

hbp_massdata.csv → One record expected per sampleID per herbGroup, generates a single subsampleID. Duplicate instances of a single subampleID are expected to exist where qaDryMass = 'Y'; these are samples that are reweighed for quality assurance purposes.

sampleIDs and **sampleBarcodes** will be generated for each collection event and functional group within a sample. The physical sample will be discarded following measurement.

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. 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 Ingest Workbook for TOS Herbaceous Biomass (NEON.DP0.10023), 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[15]).

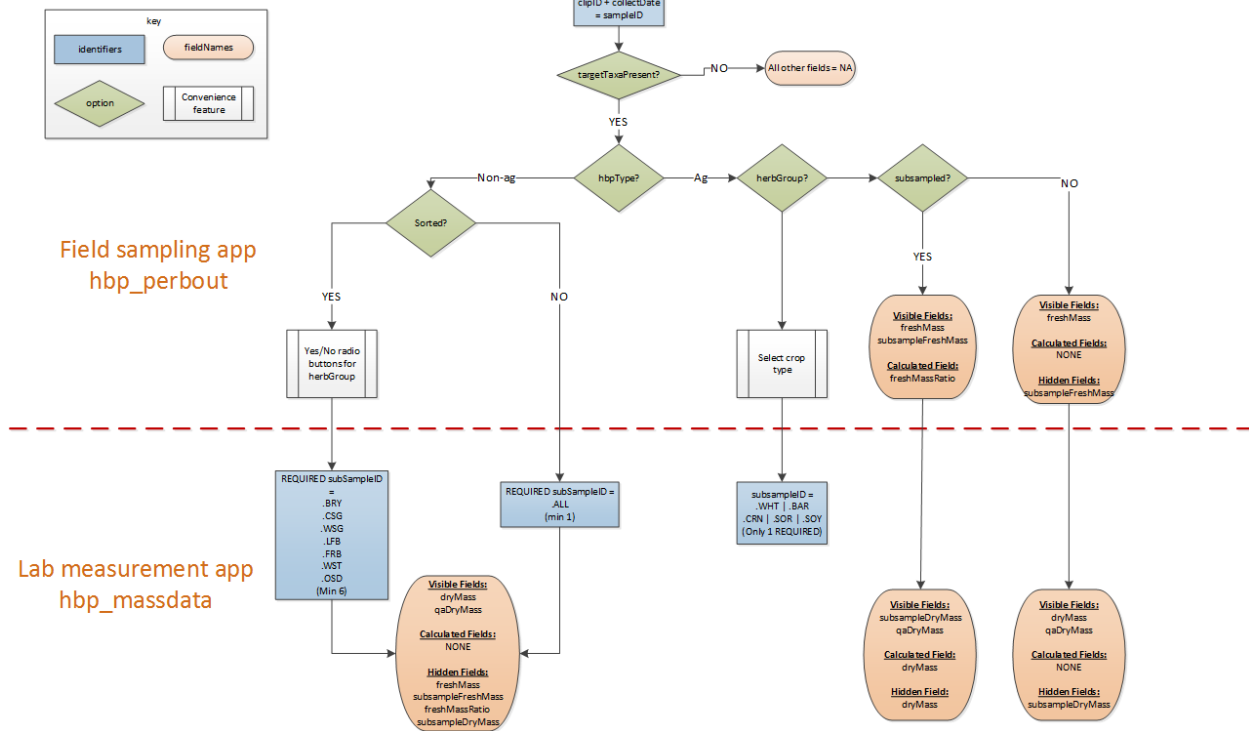A schematic of the data entry application design is depicted in Figure 3.

Figure 3: Schematic of the applications used by field technicians to enter litterfall and fine woody debris data

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[14]).

6 REFERENCES