NEON USER GUIDE TO HERBACEOUS CLIP HARVEST (DP1.10023.001)

PREPARED BY	ORGANIZATION
Katherine Jones	SCI
Courtney Meier	SCI
Dave Barnett	SCI

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	06/16/2017	Initial Release
B	11/15/2018	2018 sampling modifications described
C	10/14/2019	Additional known data quality issue described, BoutNumber described for Distributed plots
D	06/22/2020	Included general statement about usage of neonUtilities R package and statement about possible location changes
E	10/19/2020	Added Sampling Design Changes secion and content
F	05/10/2021	Adjusted mechanism for generating eventID and enhanced tracking of plots not sampled according to protocol guidelines
G	04/26/2022	Updated section 4.3 Data Revision with latest information regarding data release. Updated information regarding the geoNEON package.
H	03/10/2024	Described Optimization that resulted in reduced sampling at some sites, updated criteria for Distributed plot sampling, and new sample condidition field.
H.1	08/09/2024	Described data-driven Optimization that resulted in subsampling of standing dead and bryophyte material at select sites.
J	04/17/2025	Provided directions for calculating herbaceous productivity at sites with domestic grazing and other minor edits. Added information about the new neonUtilities Python package.
K	02/02/2026	Described 2025 nitrogen fixing functional group sorting change, 2026 Distributed plot sampling suspension, methods for tracking durability of plot sampling locations, and for calculating within-plot clip strip locations.

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 (DP1.10023.001) (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 (DP0.10023.001) (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 Data Products Catalog
AD[04]	Available with data download	Validation csv
AD[05]	Available with data download	Variables csv
AD[06]	NEON.DOC.000914	TOS Science Design for Plant Biomass and Productivity
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.004825	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[14]	Available on NEON data portal	NEON Ingest Conversion Language Function Library
AD[15]	Available on NEON data portal	NEON Ingest Conversion Language
AD[16]	Available with data download	Categorical Codes csv
AD[17]	NEON.DOC.005424	NEON Algorithm Theoretical Basis Document (ATBD) – OS Data Quality Control

2.2 Acronyms

Acronym	Definition
ANPP	Aboveground Net Primary Productivity
TOS	Terrestrial Observation System

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 a minimum of one 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
- Nitrogen-fixing forbs and woody species with diameter at decimeter height < 1 cm
- 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 and Productivity (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 at all terrestrial NEON sites with herbaceous productivity > 10% of the sum of herbaceous and woody productivity. Sampling occurs at a subset of Tower plots at some sites and at all Tower plots at others, and in a maximum of twenty randomly selected Distributed plots. The Distributed plots selected for the herbaceous clip harvest protocol are selected according to a spatially balanced, stratified-random design in order to create an unbiased estimate of herbaceous biomass at the site scale (see AD[08] for details). The protocol is implemented only at Distributed plots with $\geq$ 50% herbaceous plant cover at the scale of the plot (Figure 1).

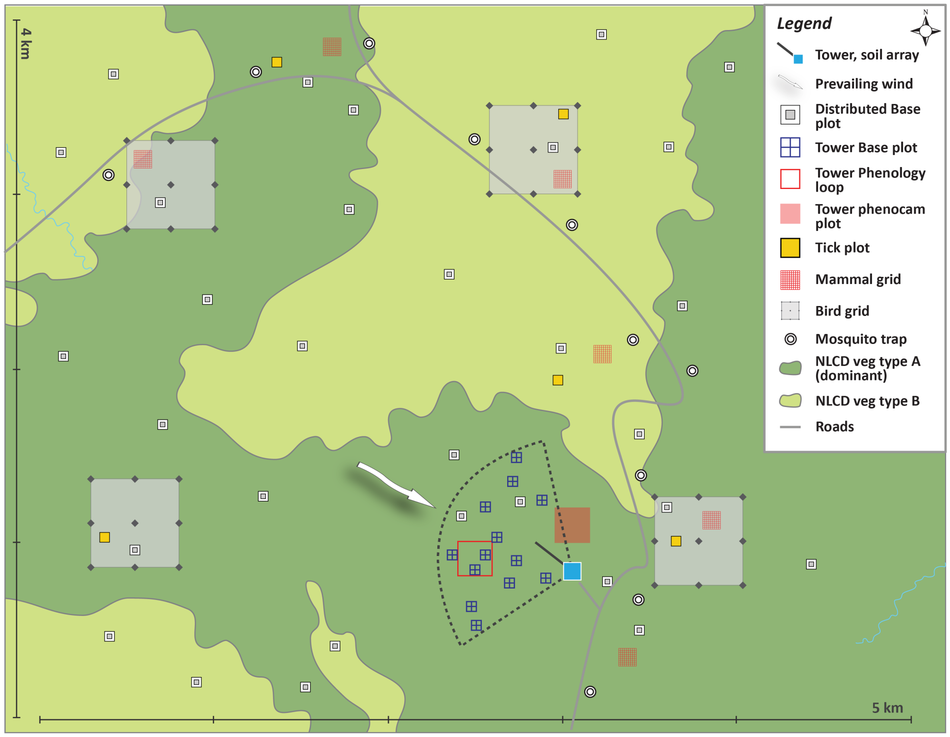

Figure 1: Generalized TOS sampling schematic, showing the placement of Distributed and Tower plots.

In forest, grassland, and rangeland 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 2 and Figure 3). 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 1m^2 and 10m^2 nested subplots (used for Plant Presence and Percent Cover or 'plant diversity' sampling) are never clipped. This ensures that generation of herbaceous plant productivity data minimally affects the quality of plant diversity data. See TOS Science Design for Plant Biomass and Productivity (AD[06]), TOS Protocol and Procedure: Measurement of Herbaceous Biomass (AD[07]), TOS Science Design for Spatial Sampling (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 (AD[12]). 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 until all of the possible 500+ clip cells in a plot have been sampled (or rejected) at which time sampling begins at the top of the list again. This would occur after more than five years at the most frequently clipped plots with grazing and will not be neccessary for the life of the Observatory at most plots.

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 consistent with Long-Term Agroecosystem Research (LTAR) protocols.

Sampling typically occurs in the same locations over the lifetime of the Observatory. However, sampling locations may become impractical to sample, due to disturbance or other local changes. When this occurs, the location and its location ID are retired or shifted to slightly different coordinates. Refer to the TOS plot location changes spreadsheet found in the “Terrestrial Observation System Sampling Locations” download on the spatial-data-maps page at neonscience.org (https://www.neonscience.org/data-samples/data/spatial-data-maps) for details about locations that have been retired or added since the operations phase started in 2019. The same download also includes the ’versionedPoints” and “versionedSubplots” files, which document shifts in coordinates.

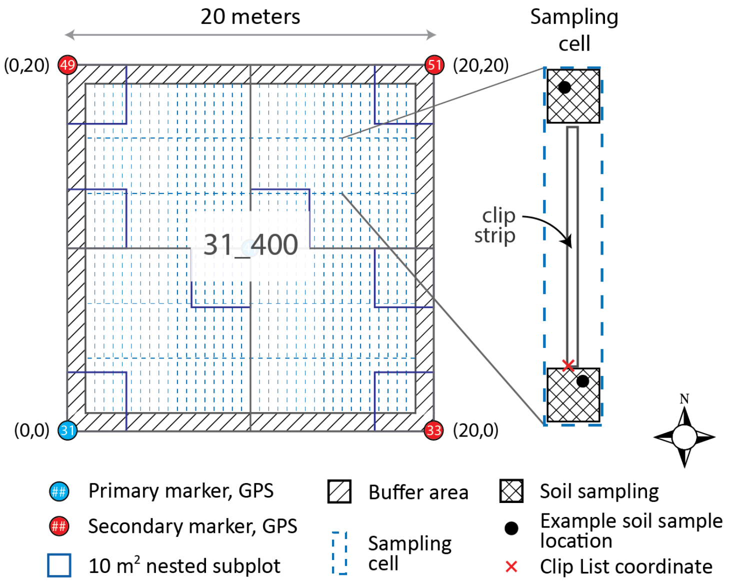

Figure 2: A 20m x 20m (400m^2) NEON plot showing potential clip harvest ”cells” (dashed blue lines). Cells overlapping 10m^2 nested subplots are not sampled for herbaceous biomass. When grazing management is absent, one randomly selected grid cell is chosen for harvest per sampling bout. At sites managed for grazing, two grid cells are sampled per bout. For each subplotID (grey numbers), the pointID at the SW corner is used to calculate the location in UTM of a given Clip List coordinate (red ‘x’).

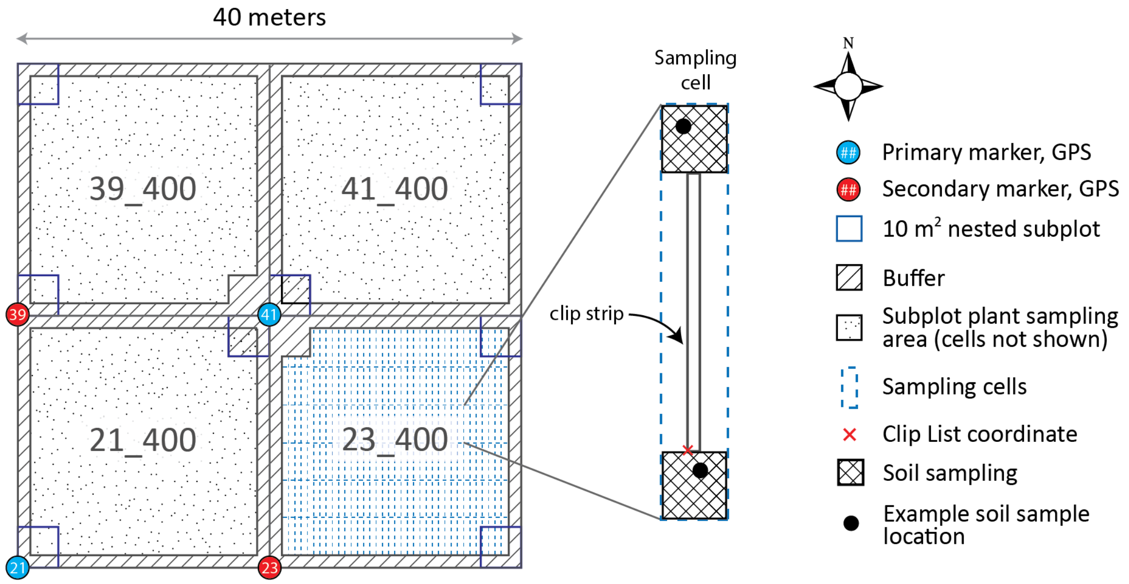

Figure 3: A 40m x 40m (1600 m^2) NEON Tower plot showing clip-harvest cells within a 400 m^2 subplot. Cells overlapping 10m^2 nested subplots are not sampled for herbaceous biomass. Clip harvest samples are generated from two randomly selected subplots per sampling bout, but for clarity, cells are not 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 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 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. Distributed plots are sampled at five year intervals at each site (see Design Changes for 2026 exception). In a site sampling year, Distributed plots are harvested once per growing season. At sites with agriculture, each plot is sampled once per year before the crop is harvested.

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 or 8 weeks depending on the compensatory response of species present.
- Sites with continuous growth and decomposition (e.g., tropical and sub-tropical grasslands) are harvested more frequently than sites with periods of senescence.
- Agriculture sites with numerous cash crop rotations are are clip-harvested prior to each crop harvest.

Within a sampling season, herbaceous clip harvests are scheduled at phenologically consistent times across sites to allow for data comparability across the Observatory and through time. To determine phenologically consistent sampling dates across widely different ecosystems, the TOS utilizes MODIS-EVI phenology data for the majority of sites (Didan 2021), and more specifically, per site averages from the most recent 10 years that are updated every 5 years. Ungrazed, wildland site herbaceous clip harvest sampling schedules correspond to the average date after which greenness begins to decrease, ensuring a high probability that all herbaceous biomass has been produced prior to clipping. For grazed sites, scheduled sampling corresponds to green-up dates, and it is assumed that cows are present from that point forward; clip harvest bouts continue until cows are removed and/or plants have senesced for the entirety of the 4 or 8 week grazing interval. At agriculture sites, MODIS is considered a rough guide; field staff monitor crops for senescence and begin clipping after the first leaves begin to dry or yellow, and work with sites to ensure clipping occurs before crops are harvested.

Clip harvest events and the resulting herbaceous biomass values can be grouped by bout according to a unique eventID. The eventID is comprised of HBP (the data product code), year, site code, and an integer - the International Organization for Standardization (ISO) week date in which the sampling bout began.

3.3 Sampling Design Changes

Several design changes have been implemented over the course of data collection. Such changes arise due to continual evaluation of the sampling design for best practices in collaboration with external technical working groups. Design changes also occur when optimization is necessary to ensure that allocation of sampling effort is poised to maximize returns to the scientific community.

- Prior to 2018: Distributed plots were sampled at three-year intervals. Subsequent sampling at five-year intervals coincides with other plant biomass and productivity data products at each site.
- Prior to 2018: All Tower plots were sampled during each sampling bout at each site.
- Prior to 2018: The Tower plot intra-annual sampling at grazed sites occurred every four weeks during the growing season.
- Prior to 2018: At select grazed sites, those bouts not associated with peak biomass during which samples are not sorted by functional group, the complete clip strip was collected and weighed. For the 2018 and subsequent sampling years a subsample of the plot-specific clip harvest is dried and weighed, but data-driven estimates of the entire sample mass are reported.
- Prior to 2019: Samples were sorted into functional groups that included the category 'Leguminous Forbs (LFB)'. Subsequent sampling places leguminous forbs and nitrogen fixers < 1 cm DBH into a 'N-fixer (NFX)' functional group category.
- 2020: The field **samplingImpractical** was added to the data to allow for the generation of a record when a plot could not be sampled or a sample could not be dried for a particular bout and year. If field sampling was not possible **samplingImpractical** is populated with a value other than 'OK' (e.g., 'location flooded') in the hbp_perbout table, no sample is collected, and no record is generated in the hbp_massdata table. If a sample is collected but not dried or weighed, **samplingImpractical** is populated with a value other than 'OK' in the hbp_massdata table and no mass data are available for that sample.

- 2020: The field **biophysicalCriteria** was added to the data to account for instances when sampling occurred but conditions were not optimal. If conditions were suboptimal - the sample was not collected during peak biomass - the field **biophysicalCriteria** is populated with a value other than 'OK - no known exceptions' (e.g., 'conditions not met: most plants not yet flowering') but the sample is collected and mass data are generated. From 2021 onward, bout numbers are assigned according to the ISO week that the bout began.
- Prior to 2021: The bout number for each clipping event within a site was assigned a number. Sampling at Tower plots began with bout = 1 and continued with ordered numeric increments for each sampling event at each site within a sampling year. Bout number = 0 was assigned to Distributed plots that are typically harvested only once per season for biomass stocks. This means that agriculture sites where different Distributed plots may be harvested at different times all have boutNumber = 0, regardless of clipDate. From 2021 onward, bout numbers are assigned according to the ISO week that the bout began.
- Prior to 2021: The eventID was comprised of HBP, year, site code, and the bout number.
- May of 2021: The data were structured such that records were created when sampling did not occur. Such instances include:

  -Distributed plots in forested NLCD types. The field **targetTaxaPresent** is set to 'No', and the **remarks** field is populated with 'forest cover type, sampling criteria not met.'

  -Distributed plots in non-forest NLCD types where the herbaceous cover < 50%. The field **targetTaxaPresent** is set to 'No', and the **remarks** field is populated with 'herbaceous cover < 50% of plot area, sampling criteria not met.'

  -Grazed sites when a plot was not subjected to grazing during the interval prior to the scheduled sampling event. The field **samplingImpractical** is set to 'exclosure cell not clipped, livestock absent' or 'ambient cell not clipped, livestock absent' for the clip cell under the exclosure and exposed to grazing respectively.
- 2022: Prior to the 2023 sampling season, harvesting and sorting all bryophyte material from the clipped biomass resulted in prolonged sampling collection events at a subset of sites due to high density of bryophyte material. To reduce the time associated with sorting bryophytes from the current-year herbaceous material, a 25% subsample is sorted at HEAL, BONA, and DEJU, and a 50% subsample is sorted at BARR and TOOL. Mass ratios are then calculated to deliver end users calculated estimates of dried, current-year mass as if the entire clip strip was processed and weighed.
- 2023:

  -Prior to 2024, Distributed plots were sampled if the National Land Cover Data class was not of a forest type. Beginning in 2023, Distributed plots are sampled if the aerial cover of the plot is $\geq$ 50% herbaceous cover regardless of NLCD cover type.

  -Herbaceous clip harvest sampling was suspended in Tower plots at 11 forested sites (BART, LENO, SERC, HARV, ORNL, PUUM, UNDE, DELA, GUAN, SCBI, and WREF) where herbaceous aboveground productivity was determined to be < 10% of the sum of aboveground herbaceous and woody productivity, based on an analysis of data collected through 2021. The NEON Science Design stipulates that only vegetation components that contribute > 10% cover or productivity should be sampled.

- 2023: Prior to the 2023 sampling season, harvesting and sorting all old standing dead (OSD) material from the current-year herbaceous masses resulted in prolonged sample collection and processing events due to the high density of herbaceous material at a subset of sites. To reduce the time associated with sorting OSD from target material, a 25% subsample is sorted at D09: DCFS, NOGP, WOOD; D11: CLBJ, OAES; and D17: SJER. Mass ratios are then calculated to deliver data that reflect estimates of dried, current-year masses as if the entire clip strip was processed and weighed.
- 2024:

  -The field **eventID** was changed such that “Tower” or “Distributed” was appended to better enable correct handling of samples taken at different temporal frequencies.

  -The field **sampleCondition** was added to the hbp_massdata table to enable a description sample condition relevant to the sample quality and data analysis.

  -The subplotID structure was changed to always contain the point that anchors the southwest corner as well as the subplot area in meters squared (see Figure 1 and 2). The ‘subplotID’ in all past data have been edited to be consistent with the new subplotID structure. Additionally, the subplotID centroid spatial data are now available via the NEON API.
- 2025: Until the start of the 2025 collection, site-specific species lists guided field sorting of fresh mass to functional group, including nitrogen fixing species. A 2024 publication (Kates et al. 2024) established a definitive standard for which species have the capacity to fix nitrogen. Starting in 2025 NEON is sorting clipped biomass into the nitrogen-fixing functional group according to this new standard.
- 2026: Data collection for the Herbaceous clip harvest at Distributed plots was suspended in 2026, leading to a 6-year interval between sampling events rather than the expected 5-year interval. The annual sampling at Tower plots is unaffected by this change and will be carried out in 2026 as usual. Distributed plot sampling originally scheduled for 2026 will be implemented in 2027, and the schedule originally intended for 2027 will be implemented in 2028 and so forth, leading to a consistent 6-year interval between sampling events across all sites in the schedule until 2032.

3.4 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 (DP0.10023.001) (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 (DP1.10023.001) (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 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 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, there are two gridded clip harvest locations within two different subplotIDs within each plotID, and at grazed sites (**exclosure** = 'Y' and 'N') there are two clip harvest locations per subplotID (Figure 3):

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

Tower plots and Distributed plot that are 20 m x 20 m contain a single clip harvest location in one subplotID within each plotID:

Tower and Distributed plots (400m^2): clipID → subplotID → plotID → habitat (NLCD veg class) → siteID → domainID

The naming convention for subplots within Base plots consists of the identity of the plot point in the southwest corner of the subplot and the scale or size of the subplot. For example, subplot '21_400' is located with point 21 in the southwest corner and is 400m^2 (20m x 20m, Figure 2) and subplot '41_100' is located such that point 41 is in the southwest corner and is 100m^2 (10m x 10m, Figure 2).

Subplots within Base plots in data releases prior to the 2024 data release – Release 2024 – follow a slightly different naming convention. Previously, subplots of 100m^2 or 400m^2 were identified only by the identity of the point in the southwest corner of the subplot (e.g., '21' or '41'). The difference is that the updated subplotID contains the scale or area of the subplot in the string (e.g., '21_400' or '41_100').

The basic spatial data included in the data downloaded include spatial location (northing and easting) and elevation of the *centroid* of the plot where sampling occurred, plus associated uncertainty due to GPS error and plot width. Shapefiles of all NEON Terrestrial Observation System sampling locations can be found on the NEON science webpage at https://www.neonscience.org/data-samples/data/spatial-data-maps.

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

- Use the getLocTOS function from the geoNEON package, available here: https://github.com/NEONScience/NEON-geolocation
- Or follow these steps to perform the same calculation:
  1. Find the namedLocation field in the data; this is the named location of the plot. For example, 'HARV_052.basePlot.hbp' is a complete named location for plotID 'HARV_052'. Determining the geographic location of the clip harvest requires the named location of the pointID in the corner of the subplot from which the sample was taken. For example, point 41 from the southwest corner of subplot 41_400. Construct the named location of the pointID of each record in hbp_perbout by concatenating the fields for namedLocation and

pointID as: namedLocation + '.' + pointID, e.g. pointtID of subplot '41_400' 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/data-api; e.g. http://data.neonscience.org/api/v0/locations/HARV_026.basePlot.bet.E?history=true) to query for elevation ("locationElevation"), easting("locationUtmEasting"), northing("locationUtmNorthing"), coordinateUncertainty ("Value for Coordinate uncertainty"), elevationUncertainty ("Value for Elevation uncertainty"), and utmZone ("locationUtmZone"). If the location data include more than one record, use the coordinates corresponding to the sampling date.
3. Identify the clipCellNumber of each clip strip (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. The calculated locations for Northing and Easting of the clip cell correspond to the SW corner of the clip strip, not the centroid.

$$yCoordNorthing = pointIDLocationUtmNorthing + offsetNorthing$$

$$xCoordEasting = pointIDLocationUtmEasting + offsetEasting$$

4. Increase coordinateUncertainty by an appropriate amount to account for error introduced by navigating within plots. Technicians may shift elevated clip strip 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.

3.6 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.

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

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 and download data via the api is available here: https://github.com/NEONScience/NEON-utilities.

3.7 Associated Data Streams

Additional information about woody vegetation present in the plot, if relevant, is available in the NEON Vegetation Structure data product (DP1.10098.001). 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 (DP1.10033.001).

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)

3.8 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 maximum of 960 samples generated in Tower plots. In years scheduled 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.

When sampling does not occur, records are generated to account for expected plots. Obstacles to sampling such as snow or management activities are reflected in the **samplingImpractical** field. Likewise, **samplingImpractical** describes instances at grazed sites when a Tower plot was not subjected to grazing during the interval prior to the sampling event and when the event is not the peak biomass sampling bout. Sampling does not occur at Distributed plots where herbaceous cover is less than 50 percent. These instances are reflected in the **remarks** field and **targetTaxaPresent** is set to 'No.'

3.9 Data Relationships

The unique **sampleID** is generated from the fields **clipID** and the **collectDate**. Each sampleID created in the field sampling activity, hbp_perbout, results in 1-5 child records, subsampleIDs, in the hbp_massdata table. 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.

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

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

sampleID and the associated barcode (**sampleCode**) can be used to link the tables. The **sampleID** and **sampleCode** are generated for each collection event, and **subsampleIDs** and **subsampleCodes** are generated for each functional group within a sample in the hbp_massdata table. Physical samples are discarded following measurement.

Primary key for hbp_perbout → **plotID**, **sampleID**

Primary key for hbp_massdata → **subsampleID**, **qaDryMass**

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 DATA QUALITY

4.1 Data Entry Constraint and Validation

Many quality control measures are implemented at the point of data entry within a mobile data entry application or web user interface (UI). For example, data formats are constrained and data values controlled through the provision of dropdown options, which reduces the number of processing steps necessary to prepare the raw data for publication. 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 (DP0.10023.001), 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[15]).

Data collected prior to 2017 were processed using a paper-based workflow that did not implement the full suite of quality control features associated with the interactive digital workflow.

A schematic of the data entry application design is depicted in Figure 4.

4.2 Automated Data Processing Steps

Following data entry into a mobile application or web user interface, the steps used to process the data through to publication on the NEON Data Portal are detailed in the NEON Algorithm Theoretical Basis Document: OS Generic Transitions (AD[13]).

Published data are reviewed for completeness, timeliness, and validity using an internal set of tests and metrics, as detailed in the NEON Algorithm Theoretical Basis Document: OS Data Quality Control (AD[17]). These quality tests are used to guide process improvements, audits of analytical facilities, and data updates, but do not generate quality flags in published data.

4.3 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.

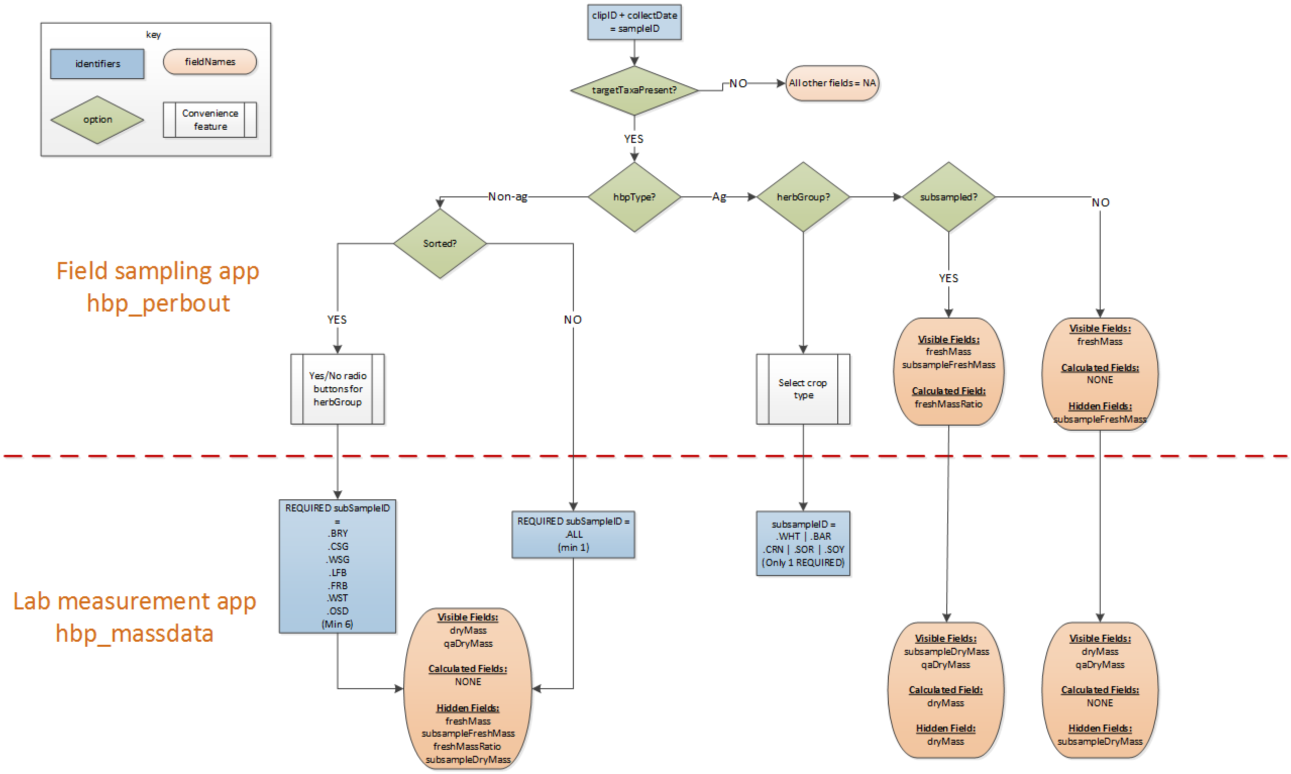

Figure 4: Schematic of the applications used by field technicians to enter herbaceous biomass data

4.4 Quality Flagging

The **dataQF** field in each data record is a quality flag for known errors applying to the record. Please see 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

Additional known quality issues that are not flagged: For all records published prior to 2019-08-27, the **measuredBy** and **recordedBy** fields were populated incorrectly and may or may not be unique.

5 Calculating Productivity at Grazed Sites

At sites with domestic cattle grazing in the area where Tower plots are established, data are reported from a clip cell exposed to cattle grazing - ambient - and from an exclosure-protected clip cell - exclosure

- in each subplot sampled to enable accounting for consumption (Milchunas et al. 1994). Multiple bouts of sampling, separated by 4 or 8 week grazing intervals, are conducted to enable accounting for compensatory growth. Exclosures are established within plots and each subplot sampled prior to the arrival of cattle and sampling bouts are repeated until cattle are removed and/or plants senesced for the entirety of the previous grazing interval. Generally, primary productivity is calculated by determining consumption across all bouts and adding the standing biomass left following the removal of grazers or senescence has occurred. Clipped biomass is dried and reported in grams in the field **dryMass** in the hbp_massdata table.

Because the subplot size is large (400m^2), the clip strips are small (0.2m^2), the exclosed clip cell and the ambient clip cells are randomly placed in the subplot (i.e., not paired), and the heterogeneity of standing herbaceous mass within the subplot, it is considered best practice to calculate primary productivity of herbaceous vegetation in grazed sites across all plots, at the scale of the site, not at plot scales.

Equations and notation to estimate total productivity follows:

5.1 Grazing Consumption per Bout

For each sampling bout i, grazing consumption at the site scale is calculated as the difference between the **mean biomass** of clip strips protected from grazing (‘exclosure’) and those exposed to grazing (‘ambient’), averaged across all sampled plots at the site during that bout:

$$C_i = \overline{M}_{\text{exclosure},i} - \overline{M}_{\text{ambient},i}$$

Where:

- C_i is the site-level grazing consumption during the i-th bout (g).
- $\overline{M}_{\text{exclosure},i}$ is the mean mass (g) in the exclosure clip strips, averaged across all plots during bout i.
- $\overline{M}_{\text{ambient},i}$ is the mean mass (g) in the ambient clip strips, averaged across all plots during bout i.

The total site-level grazing consumption over all N bouts is then:

$$\text{Total Consumption} = \sum_{i=1}^{N} C_i = \sum_{i=1}^{N} \left(\overline{M}_{\text{exclosure},i} - \overline{M}_{\text{ambient},i}\right)$$

5.2 Remaining Standing Biomass (Post-Grazing or Post-Senescence)

Once grazing ceases or plants senesce for the entirety of the previous grazing interval, the remaining standing biomass is the mean reported mass from the final ambient clip cells, calculated across all sampled plots during the last bout:

$$M_{\text{final, ambient}} = \text{Mean dry mass of ambient clip cells from the last bout or senescence}$$

5.3 Total Herbaceous Productivity

The total herbaceous productivity is the sum of the total grazing consumption and the remaining standing mass:

$$P_{\text{total}} = \text{Total Consumption} + M_{\text{final, ambient}}$$

Where:

- P_{total} is the total herbaceous productivity (g) over the growing season.
- Total Consumption is the sum of mass consumed by grazers across all bouts, based on site-level mean mass.
- $M_{\text{final, ambient}}$ is the mean remaining mass in ambient clip strips across all plots after grazing has ceased or senescence.

5.4 Example

Consider a scenario with three sampling bouts:

Bout	$\overline{M}_{\text{exclosure},i}$ (g)	$\overline{M}_{\text{ambient},i}$ (g)	Consumption (C_i) (g)
Bout 1	120	100	20
Bout 2	170	150	20
Bout 3	200	180	20

Note that the data reflect means across all plots at the site for each bout.

- **Total Consumption**:

$$\text{Total Consumption} = 20 + 20 + 20 = 60\,\text{g}$$

Assuming the site-level mean of final standing ambient clip mass after the last bout is 250 g:

- **Remaining mass**:

$$M_{\text{final, ambient}} = 250\,\text{g}$$

- **Total productivity**:

$$P_{\text{total}} = 60\,\text{g} + 250\,\text{g} = 310\,\text{g}$$

6 References

Didan, K. MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V061. 2021. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MOD13Q1.061

Kates, H.R., O'Meara, B.C., LaFrance, R. et al. Shifts in evolutionary lability underlie independent gains and losses of root-nodule symbiosis in a single clade of plants. Nat Commun 15, 4262 (2024). https://doi.org/10.1038/s41467-024-48036-3

Milchunas, D.G., Lauenroth, W.K., & Burke, I.C. 1994. Grasses and Grazers: The Influence of Grazing on Vegetation Dynamics. Ecology, 75(6): 1512-1525. https://doi.org/10.2307/1942062