

<i>Title:</i> NEON User Guide to Plant foliar physical and chemical properties (NEON.DP1.10026) and Plant foliar stable isotopes (NEON.DP1.10053)	<i>Date:</i> 12/07/2017
<i>Author:</i> Samantha Weintraub	<i>Revision:</i> A

# NEON USER GUIDE TO PLANT FOLIAR PHYSICAL AND CHEMICAL PROPERTIES (NEON.DP1.10026) AND PLANT FOLIAR STABLE ISOTOPES (NEON.DP1.10053)

<b>PREPARED BY</b>	<b>ORGANIZATION</b>	<b>DATE</b>
Samantha Weintraub	FSU	12/07/2017

<i>Title:</i> NEON User Guide to Plant foliar physical and chemical properties (NEON.DP1.10026) and Plant foliar stable isotopes (NEON.DP1.10053)	<i>Date:</i> 12/07/2017
<i>Author:</i> Samantha Weintraub	<i>Revision:</i> A

## CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	12/04/2017	Initial Release

## 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 . . . . .	3
<b>3</b>	<b>DATA PRODUCT DESCRIPTION</b>	<b>4</b>
3.1	Spatial Sampling Design . . . . .	4
3.2	Temporal Sampling Design . . . . .	6
3.3	Theory of Measurements . . . . .	7
3.4	Variables Reported . . . . .	7
3.5	Spatial Resolution and Extent . . . . .	8
3.6	Temporal Resolution and Extent . . . . .	10
3.7	Associated Data Streams . . . . .	10
3.8	Product Instances . . . . .	10
3.9	Data Relationships . . . . .	11
3.10	Special Considerations . . . . .	12
<b>4</b>	<b>DATA QUALITY</b>	<b>12</b>
4.1	Data Entry Constraint and Validation . . . . .	12
4.2	Automated Data Processing Steps . . . . .	13
4.3	Data Revision . . . . .	13
4.4	Quality Flagging . . . . .	13
4.5	Analytical Facility Data Quality . . . . .	13
<b>5</b>	<b>REFERENCES</b>	<b>15</b>

## LIST OF TABLES AND FIGURES

Figure 1	Representation of a NEON site with select Tower and Distributed plots shown . . . . .	5
Figure 2	Representation of NEON base plots used for foliar sampling . . . . .	6
Figure 3	Schematic of the application used by field technicians to enter foliar sampling field data . . . . .	16

## 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 measured height of a sun-lit foliage sample, 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 two related L1 data products. The first is Plant foliar physical and chemical properties (NEON.DP1.10026), which includes field metadata, leaf mass per area calculations, and carbon (C), nitrogen (N), lignin, chlorophyll, and major/minor/trace element concentrations in sun-lit foliage samples. The second is Plant Foliar stable isotopes (NEON.DP1.10053), which includes C and N stable isotope values in those same sun-lit foliage samples. This user guide also provides details relevant to the publication of the data products via the NEON data portal, with additional detail available in the files NEON Data Variables for Plant foliar physical and chemical properties (NEON.DP1.10026) (AD[07]) and NEON Data Variables for Plant foliar stable isotopes (NEON.DP1.10053) (AD[08]), 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: Canopy Foliage Sampling (AD[11]). The raw data that are processed in this document are detailed in the files NEON Raw Data Validation for Plant foliar physical and chemical properties, Level 0 (NEON.DP0.10026) (AD[04]), NEON Raw Data Validation for Carbon and nitrogen concentrations and stable isotopes in plants and soil (NEON.DP0.10103) (AD[05]), and NEON Raw Data Validation for Plant lignin concentrations (NEON.DP0.10031) (AD[06]), 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.DP0.10026.001_dataValidation.csv	NEON Raw Data Validation for Plant foliar physical and chemical properties, Level 0 (NEON.DP0.10026)
AD[05]	NEON.DP0.10103.001_dataValidation.csv	NEON Raw Data Validation for Carbon and nitrogen concentrations and stable isotopes in plants and soil (NEON.DP0.10103)
AD[06]	NEON.DP0.10031.001_dataValidation.csv	NEON Raw Data Validation for Plant lignin concentrations (NEON.DP0.10031)
AD[07]	NEON.DP1.10026.001_variables.csv	NEON Data Variables for Plant foliar physical and chemical properties (NEON.DP1.10026)
AD[08]	NEON.DP1.10053.001_variables.csv	NEON Data Variables for Plant foliar stable isotopes (NEON.DP1.10053)
AD[09]	NEON.DOC.000906	TOS Science Design for Terrestrial Biogeochemistry
AD[10]	NEON.DOC.001024	TOS Protocol and Procedure: Canopy Foliage Sampling
AD[11]	NEON.DOC.000987	TOS Protocol and Procedure: Measurement of Vegetation Structure
AD[12]	NEON.DOC.000008	NEON Acronym List
AD[13]	NEON.DOC.000243	NEON Glossary of Terms
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
$\delta^{13}\text{C}$	delta 13C, the stable carbon isotope ratio ( $^{13}\text{C}:^{12}\text{C}$ ) in a sample compared to a reference material, reported in parts per thousand
$\delta^{15}\text{N}$	delta 15N, the stable nitrogen isotope ratio ( $^{15}\text{N}:^{14}\text{N}$ ) in a sample compared to a reference material, reported in parts per thousand
AOP	Airborne Observation Platform
C	Carbon
LMA	Leaf mass per area
MODIS	Moderate Resolution Imaging Spectroradiometer
N	Nitrogen

### 3 DATA PRODUCT DESCRIPTION

The Plant foliar physical and chemical properties (NEON.DP1.10026) and Plant foliar stable isotopes (NEON.DP1.10053) data products provide physical (leaf mass per area), chemical (C, N, lignin, chlorophyll, and major/minor/trace element), and stable isotope ( $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$ ) data for plant foliage samples collected using TOS Protocol and Procedure: Canopy Foliage Sampling (AD[10]). Plant foliar sampling and analyses implement the guidelines and requirements described in TOS Science Design for Terrestrial Biogeochemistry (AD[09]). All physical, chemical and isotopic data are reported at the spatial resolution of a woody individual or a bulk herbaceous clip strip located within a NEON plot. The temporal resolution is that of a single collection date.

Whenever possible, foliar data are collected in conjunction with overflights of the NEON Airborne Observation Platform (AOP), which conducts remote sensing of ecosystem chemical and physical characteristics using hyperspectral and LiDAR measurements. Ground-based foliar data can be used to ground-truth and validate AOP measurements, and may be of utility to the community in refining algorithms to map canopy foliar traits. Additionally, foliar data will inform species and site-level estimates of foliar physio-chemical properties and how those change over time, which will have value independent of remote sensing observations.

During canopy foliage sampling, only sun-lit vegetation is collected and analyzed. A variety of techniques are used to accomplish this, from hand clippers and pole pruners in low-stature sites to line launchers and slingshots in high-stature sites. Sampling method is recorded in the field data table.

Plant foliar physical, chemical and isotopic data can shed light onto key ecological processes including net primary productivity, nutrient uptake and allocation, rates of decomposition and herbivory, and responses to environment stress such as drought at the plot, site, and continental scales. They also provide essential data for understanding change in vegetation and canopy dynamics over time.

#### 3.1 Spatial Sampling Design

Sun-lit plant foliage is sampled at all terrestrial NEON sites. For sites with predominantly forest/shrubland cover, foliar sampling is conducted at 14 plots. For sites that are predominantly grassland, foliage is sampled from 20 plots. At all sites, four of the sampled plots are located within the NEON tower airshed (Tower plots) and 10-16 plots are distributed across the landscape (Distributed plots) in representative vegetation cover classes (Figure 2).

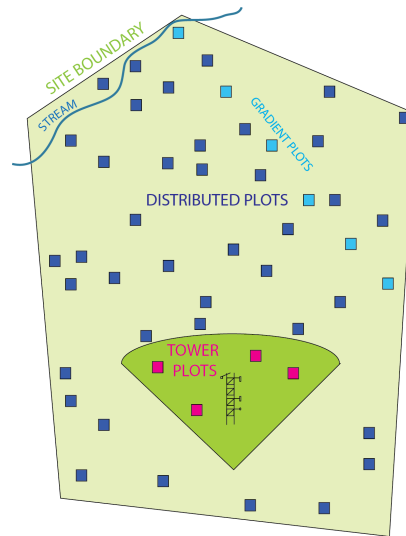


Figure 1: Representation of a NEON site with select Tower and Distributed plots shown

In plots dominated by woody individuals (> 75% cover trees and shrubs), the top three most abundant *canopy* species in that plot are identified, then one individual from each is sampled. If a site has low diversity, then less than three unique species per plot may be collected. If a site has high diversity and no clear dominants, a species that is a member of the site-level overstory, but has yet to be sampled with sufficient replication (target  $n = 3$ ), is selected. The aim of this approach is to provide data for the major canopy species at the site and landscape scale; however, it will not enable sampling of rare taxa.

Sun-lit leaves from three georeferenced individuals per plot are sampled. Georeferencing is accomplished using the Mapping and Tagging procedure outlined in TOS Protocol and Procedure: Measurement of Vegetation Structure (AD[11]), and many trees sampled for foliar properties will also have structure data available from that protocol. Foliar sampling can occur anywhere within a NEON plot except for the plant diversity nested subplots (Figure 2).



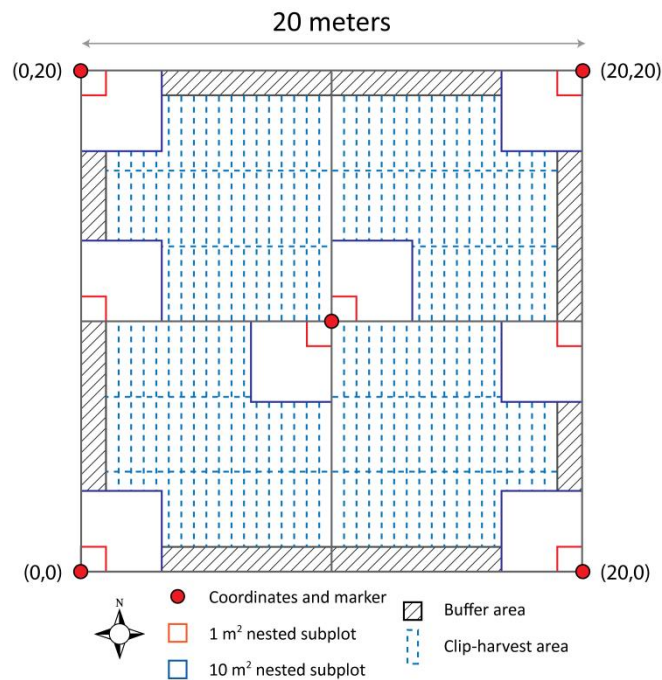


Figure 2: Representation of NEON base plots used for foliar sampling

In plots dominated by herbaceous vegetation (>75% non-woody cover), all aboveground material in one or two (depending on plot size) randomly selected clip strips, which are nested within 0.5 x 3 m clip harvest areas or cells (Figure 2), are harvested, homogenized, and subsampled for physical, chemical, and isotopic analyses. Clip strips are usually 0.1 X 2 m, unless the site has patchy or scarce vegetation, in which case a 0.5 x 2 m area can be sampled. Clip strip dimensions are recorded in the data. Clip strip material is not sorted prior to subsampling and analysis.

In mixed plots where woody individuals and herbaceous plants both account for > 25% aerial cover, a combination of the two sampling approaches is employed, with one clip strip per 20 x 20 m plot or subplot and up to three georeferenced woody individuals (if present) collected. In this mixed cover case, clip strip selection is not entirely random as only those strips in sun-lit plot areas are considered for harvest.

### 3.2 Temporal Sampling Design

Foliar properties are measured at each site once every five years. During 'on' years, samples are collected during the period of historic peak greenness at the site, according to Moderate Resolution Imaging Spectroradiometer (MODIS) data. Whenever possible, sampling occurs within two weeks of the AOP overflight, which is also timed to coincide with peak greenness.

Upon collection of sun-lit foliage, chlorophyll subsamples are immediately packaged in tinfoil, flash-frozen using dry ice, then maintained in the dark and shipped frozen to an external facility for pigment extraction within two weeks of collection. Additional processing of samples, including scans for leaf mass per area (LMA) measurement

and initiation of sample drying for chemical and isotopic analyses, occurs within 1-5 days of collection. During this holding time, samples are held at 4°C in sealed plastic bags.

### 3.3 Theory of Measurements

LMA, a measure of plant investment in storage versus light capture, is estimated using the standard flatbed scanning method (Pérez-Harguindeguy et al. 2013). Briefly, a representative subsample of foliage (one to many leaves or needles) is scanned, then the scanned area is calculated using ImageJ (Schneider et al. 2012) and normalized by the dry mass of foliage scanned. See AD[10] for more details on the procedure.

Chlorophyll a, b, and carotenoids are measured using 100 % methanol extractions, with absorbance read on a spectrophotometer and concentrations estimated using the coefficients reported in Lichtenthaler (1987). Concentrations of lignin are determined using the acid detergent lignin method and defined operationally as the acid-insoluble foliar residue. Concentration estimates of both lignin and cellulose are provided from this procedure. Major, minor and trace elements are measured using nitric acid-peroxide digestions and inductively coupled plasma mass spectrometry (ICP-MS) analyses.

Concentrations and stable isotope ratios of carbon and nitrogen are measured simultaneously using elemental analysis coupled to isotope ratio mass spectrometry (EA-IRMS). Isotopes are measured as the abundance ratio of a heavy, rare isotope (H) to a light, more common isotope (L), normalized by those same ratios in a standard reference material.

$$\delta = [(R_{sample}/R_{standard} - 1)] \times 1000$$

where R = H/L. For all NEON stable isotopic data,  $\delta^{15}\text{N}$  values are expressed on the atmospheric  $\text{N}_2$  scale and  $\delta^{13}\text{C}$  values are expressed on the Vienna Pee Dee Belemite scale.

For data collected in 2017 and beyond, standard operating procedures for laboratories performing chemical and stable isotope analyses can be found on the NEON Data Portal (<http://data.neonscience.org/home>), in the Rousources > Data Documentation > External Lab Protocols section.

### 3.4 Variables Reported

All variables reported from the field or laboratory technician (L0 data) are listed in the files NEON Raw Data Validation for Plant foliar physical and chemical properties, Level 0 (NEON.DP0.10026) (AD[04]), NEON Raw Data Validation for Carbon and nitrogen concentrations and stable isotopes in plants and soil (NEON.DP0.10103) (AD[05]), and NEON Raw Data Validation for Plant lignin concentrations (NEON.DP0.10031) (AD[06]). All variables reported in the published data (L1 data) are also provided separately in the files NEON Data Variables for Plant foliar physical and chemical properties (NEON.DP1.10026) (AD[07]) and NEON Data Variables for Plant foliar stable isotopes (NEON.DP1.10053) (AD[08]).

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>;

<p>Title: NEON User Guide to Plant foliar physical and chemical properties (NEON.DP1.10026) and Plant foliar stable isotopes (NEON.DP1.10053)</p>	<p>Date: 12/07/2017</p>
<p>Author: Samantha Weintraub</p>	<p>Revision: A</p>

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.

### 3.5 Spatial Resolution and Extent

The finest resolution at which spatial data are reported is a sample, taken from a woody individual or herbaceous clip strip, within a NEON plot.

**sampleID** (unique ID given to the sample) → **plotID** (ID of plot within site) → **siteID** (ID of NEON site) → **domainID** (ID of a NEON domain)

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>.

In order to link ground-based foliar data to remote sensing observations, users will most likely wish to derive more precise estimates of sample locations.

When **sampleType** = *woody individual*, first download the Woody vegetation structure (NEON.DP1.10098) data for the relevant sites. Then proceed to either of these two options:

- Filter to only the **individualIDs** that appear in the foliar data, then 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. Add the variables **pointID**, **stemDistance**, and **stemAzimuth** from `vst_mappingandtagging` to each **individualID** that appears in `cfc_fieldData` using a `match` or `join` function.
  2. Construct a `pointID` named location for each record in `cfc_fieldData` by concatenating the fields for `namedLocation` and `pointID` as follows: `namedLocation + '.' + pointID`, e.g. `pointID '41'` of `namedLocation 'HARV_052.basePlot.cfc'` has a `pointID` named location of `'HARV_052.basePlot.cfc.41'`.
  3. Use the API (<http://data.neonscience.org/api>; e.g. [http://data.neonscience.org/api/v0/locations/HARV\\_052.basePlot.cfc.41](http://data.neonscience.org/api/v0/locations/HARV_052.basePlot.cfc.41)) to query for easting (“`locationUtmEasting`”), northing (“`locationUtmNorthing`”), coordinateUncertainty (“`Value for Coordinate uncertainty`”), and utmZone (“`locationUtmZone`”) for each `pointID` named location and use as inputs to the next step.
  4. Calculate absolute position in UTM of each woody individual stem using **stemAzimuth**, **stemDistance** and the easting and northing values derived in step 3 above, using equations (1) and (2):

$$Easting = easting.pointID + d * \sin \theta \quad (1)$$

and

$$Northing = northing.pointID + d * \cos \theta \quad (2)$$

where,

$$\theta = \frac{stemAzimuth * \pi}{180} \quad (3)$$

$$easting.pointID = \quad (4)$$

the easting value of the pointID named location,

$$northing.pointID = \quad (5)$$

the northing value of the pointID named location,

$$d = \quad (6)$$

stemDistance

5. Increase **coordinateUncertainty** by an appropriate amount (suggested 1 m) to account for error introduced by navigating around the plot. Keep in mind that geolocations are of the stem.

When **sampleType** = *herbaceous clip strip*, there are two options to derive a more precise location:

- 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. Construct the named location of the subplot of each record in `cfc_fieldData` by concatenating the fields for `namedLocation` and `subplotID` as: `namedLocation + '.' + subplotID`, e.g. subplotID '41' of namedLocation 'WOOD\_047.basePlot.cfc' has a subplotID named location of 'WOOD\_047.basePlot.cfc.41'.
  2. Use the API (<http://data.neonscience.org/api>; e.g. [http://data.neonscience.org/api/v0/locations/WOOD\\_047.basePlot.cfc.41](http://data.neonscience.org/api/v0/locations/WOOD_047.basePlot.cfc.41)) to query for easting ("locationUtmEasting"), northing ("locationUtmNorthing"), coordinateUncertainty ("Value for Coordinate uncertainty"), and utmZone ("locationUtmZone") for each subplot named location as inputs to the next step.

- Use the clip cell lookup table, available here: [http://data.neonscience.org/api/v0/documents/clipCellNumber\\_lookup](http://data.neonscience.org/api/v0/documents/clipCellNumber_lookup) (clicking on link will initiate download), to find the offsets for each clipCellNumber and subplot (note that **subplot** in cfc\_fieldData\_in is the same as pointID in the offset table).
- Use these offsets to adjust the easting and northing values downloaded in step 2, using equations (7) and (8):

$$Easting = easting.subplotID + eastingOffset \quad (7)$$

and

$$Northing = northing.subplotID + northingOffset \quad (8)$$

- Increase coordinateUncertainty by an appropriate amount to account for error introduced by navigating within plots (suggested 0.5 m). Keep in mind that the calculated value is the center of the clip strip.

### 3.6 Temporal Resolution and Extent

The finest resolution at which temporal data are reported is the **collectDate**.

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>

### 3.7 Associated Data Streams

As above, users will likely wish to download the data from Woody plant vegetation structure (NEON.DP1.10098) in order to derive more precise locations of woody individual samples, joining on **individualID** in the vst\_mappingandtagging table. Moreover, many of the trees sampled for foliar properties will also have structure data available in that data product.

### 3.8 Product Instances

There is one foliar sampling event per year at 8-10 sites, with 20-60 samples collected per site depending on plot sizes and dominant vegetation. This will yield 160-600 unique foliar samples per year observatory-wide. Each sample is then subsampled for 6 unique physico-chemical measurements, yielding 960-3,600 unique data product instances in a given calendar year.

### 3.9 Data Relationships

TOS Protocol and Procedure: Canopy Foliage Sampling dictates that each woody individual or herbaceous clip strip collected yields a unique **sampleID** in the `cfc_fieldData` table in Plant foliar physical and chemical properties. A record from `cfc_fieldData` is expected to have one child record in `cfc_LMA` and several child records in `cfc_chemistrySubsampling`, provided in that same data product. Each record from `cfc_chemistrySubsampling` is expected to have one or two (if analytical replicates were conducted) child records in `cfc_chlorophyll`, `cfc_carbonNitrogen`, `cfc_elements` and `cfc_lignin` (Plant foliar physical and chemical properties) and `cfc_stableIsotopes` (Plant foliar stable isotopes). It is expected that child sample identifiers will appear a maximum of twice per chemistry table, but 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.*

*Plant foliar physical and chemical properties:*

`cfc_fieldData.csv` -> One record expected per **sampleID**, generates a single **chlorophyllSampleID** used to measure chlorophyll and carotenoids

`cfc_LMA.csv` -> One record expected per **sampleID**, generates a single **lmaSampleID** used to measure LMA

`cfc_chemistrySubsampling.csv` -> One record expected per **sampleID**, generates a **cnSampleID** used to measure carbon and nitrogen concentrations and stable isotopes and a **ligninSampleID** used to measure lignin and elements

`cfc_chlorophyll.csv` -> One record expected per **chlorophyllSampleID** x **analyticalRepNumber** combination, associated with chlorophyll and carotenoid data

`cfc_carbonNitrogen.csv` -> One record expected per **cnSampleID** x **analyticalRepNumber** combination, associated with percent carbon and nitrogen data

`cfc_elements.csv` -> One record expected per **ligninSampleID** x **analyticalRepNumber** combination, associated with major, minor and trace element data

`cfc_lignin.csv` -> One record expected per **ligninSampleID** x **analyticalRepNumber** combination, associated with lignin data

`cfc_chlorophyllParameters.csv` -> One record expected per **analyte** x **chlCarotEquationInput** x **chlCarotExtinctionCoefficient** x **laboratoryName** combination, parameters used to calculate concentrations from absorbance

`cfc_chlorophyllSummary.csv` -> One record expected per **analyte** x **sampleType** x **laboratoryName** x **labSpecificStartDate** combination, used to associate sample data with relevant uncertainty values.

`bgc_CNiso_externalSummary.csv` -> One record expected per **analyte** x **sampleType** x **laboratoryName** x **labSpecificStartDate** combination, used to associate sample data with relevant uncertainty values.

`cfc_elementsSummary.csv` -> One record expected per **analyte** x **sampleType** x **laboratoryName** x **labSpecificStartDate** combination, used to associate sample data with relevant uncertainty values.

`lig_externalSummary.csv` -> One record expected per **analyte** x **sampleType** x **laboratoryName** x **labSpecificStartDate** combination, used to associate sample data with relevant uncertainty values.

<i>Title:</i> NEON User Guide to Plant foliar physical and chemical properties (NEON.DP1.10026) and Plant foliar stable isotopes (NEON.DP1.10053)	<i>Date:</i> 12/07/2017
<i>Author:</i> Samantha Weintraub	<i>Revision:</i> A

### *Plant foliar stable isotopes:*

cfc\_foliarStableIsotopes.csv -> One record expected per **cnSampleID** x **analyticalRepNumber** combination, associated with  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  data

bgc\_CNiso\_externalSummary.csv -> One record expected per **analyte** x **sampleType** x **laboratoryName** x **lab-SpecificStartDate** combination, used to associate sample data with relevant uncertainty values.

## 3.10 Special Considerations

In the course of data collection for these two data products, threatened and endangered (T&E) species may occasionally be sampled. To avoid publishing locations of sensitive species, identifications of T&E species are ‘fuzzed’, i.e., reported at a higher taxonomic rank than the raw data. Moreover, as sample identifiers contain taxonomic information (to aid technicians with field and laboratory workflows), all sample identifiers are obfuscated upon data product publication to the NEON data portal. Sample identifiers should thus be treated like barcodes, containing no human-readable information about the samples but being unique tags that allow for consistent sample linkage across tables.

Several variables used to track external laboratory procedures, such as raw wavelengths from pigment extractions and digest concentrations from elemental analyses, are available in the expanded package of the Plant foliar physical and chemical properties data product. These variables can be used to reconstruct the final, reported values for those users who are interested.

## 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. A schematic of the data entry application design for field collection of foliar samples is depicted in Figure 3. 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 Plant foliar physical and chemical properties, Level 0 (NEON.DP0.10026), 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. There is also a field named ‘entryValidationRulesParser’, which described the validation rules for external labs that submit spreadsheets to the NEON database. 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]]).

Note that field 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. Moreover, external laboratory data were also not subject to same full suite of quality controls.

## 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[14]).

## 4.3 Data Revision

All data are provisional until a numbered version is released; the first release of a static version of NEON data, annotated with a globally unique identifier, is planned to take place in 2020. 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 Change Log section of the data product readme, provided with every data download, contains a history of major known errors and revisions.

## 4.4 Quality Flagging

The **dataQF** field in each record is a quality flag for known issues applying to the record, added by NEON Science upon data review.

Chlorophyll and carotenoid data from 2016 was collected using legacy methods that allowed for a very long sample hold time (several months). From 2017 onwards, samples are analyzed within two weeks of collection. Accordingly, pigment data from 2016 has been flagged as described below.

fieldName	value	definition
dataQF	deprecatedMethod	Data generated using deprecated/legacy workflows; overly long hold time for pigment samples

## 4.5 Analytical Facility Data Quality

All analytical labs that generate plant foliar chemical and stable isotope data include standards or secondary reference materials run as unknowns alongside NEON samples in order to gauge run acceptability. Labs communicate batch-level issues with accuracy of check-standards or secondary reference materials, as well as record-level issues with samples or measurements, using the suite of quality flags described below. In general, an entry of 0 in a quality flag field means there is no issue to report.

In addition, long-term analytical precision and accuracy of check-standard or secondary reference material analyses are reported for each lab to allow users to interpret and analyze foliar chemistry and stable isotope data in the context of their uncertainty ranges. The data tables `cfc_chlorophyllSummary`, `bgc_CNiso_externalSummary`, `cfc_elementsSummary`, and `lig_externalSummary`, which are available in the data product expanded packages, contain the long-term precision and accuracy of lab analyses.

For further information about individual laboratory QA procedures, refer to the lab-specific SOPs found on the NEON Data Portal (<http://data.neonscience.org/home>) in the Resources > Data Documentation > External Lab Protocols section.



#### cfc\_chlorophyll

fieldName	value	definition
relativeAccuracyScale	A	Observed value of run QA material(s) within 0-5 % of known value
relativeAccuracyScale	B	Observed value of run QA material(s) within 5-10 % of known value
relativeAccuracyScale	C	Observed value of run QA material(s) within 10-20 % of known value
relativeAccuracyScale	D	Observed value of run QA material(s) greater than 20 % deviation from known value
chlMeasurementQF	0	No measurement issues to report
chlMeasurementQF	1	Absorbance value below the optimal range
chlMeasurementQF	2	Absorbance value above the optimal range
chlHandlingQF	0	No handling issues to report
chlHandlingQF	1	Sample exceeded standard holding time
chlHandlingQF	2	Sample exceeded standard extraction time
chlHandlingQF	3	Sample exceeded standard holding and extraction times

#### cfc\_carbonNitrogen

fieldName	value	definition
cnPercentQF	0	Indicator for both percentNitrogen and percentCarbon in range
cnPercentQF	1	Indicator for percentNitrogen out of range
cnPercentQF	2	Indicator for percentCarbon out of range
cnPercentQF	3	Indicator for both percentNitrogen and percentCarbon out of range

#### cfc\_elements and cfc\_lignin

fieldName	value	definition
accuracyQF	0	Observed values of run QA material fall within acceptance criteria
accuracyQF	1	Observed values of run QA material do not fall within acceptance criteria but samples cannot be re-run
measurementQF	0	No measurement issues to report
measurementQF	1	Sample values suspect and should be used with caution, with reason explained in remarks

#### cfc\_foliarStableIsotopes

fieldName	value	definition
cnIsotopeQF	0	Indicator for both d15N and d13C in range
cnIsotopeQF	1	Indicator for d15N out of range
cnIsotopeQF	2	Indicator for d13C out of range
cnIsotopeQF	3	Indicator for both d15N and d13C out of range

## 5 REFERENCES

Pérez-Harguindeguy, N., S. Díaz, E. Garnier, S. Lavorel, H. Poorter, et al. (2013). New handbook for standardised measurement of plant functional traits worldwide. *Australian Journal of Botany* 61: 167-234.

Schneider, C, W. Rasband, and K. Eliceiri. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature methods* 9: 671-675.

Lichtenthaler, H.K. (1987) Chlorophylls and Carotenoids: Pigments of Photosynthetic Biomembranes. *Methods in Enzymology* 148: 350-382.

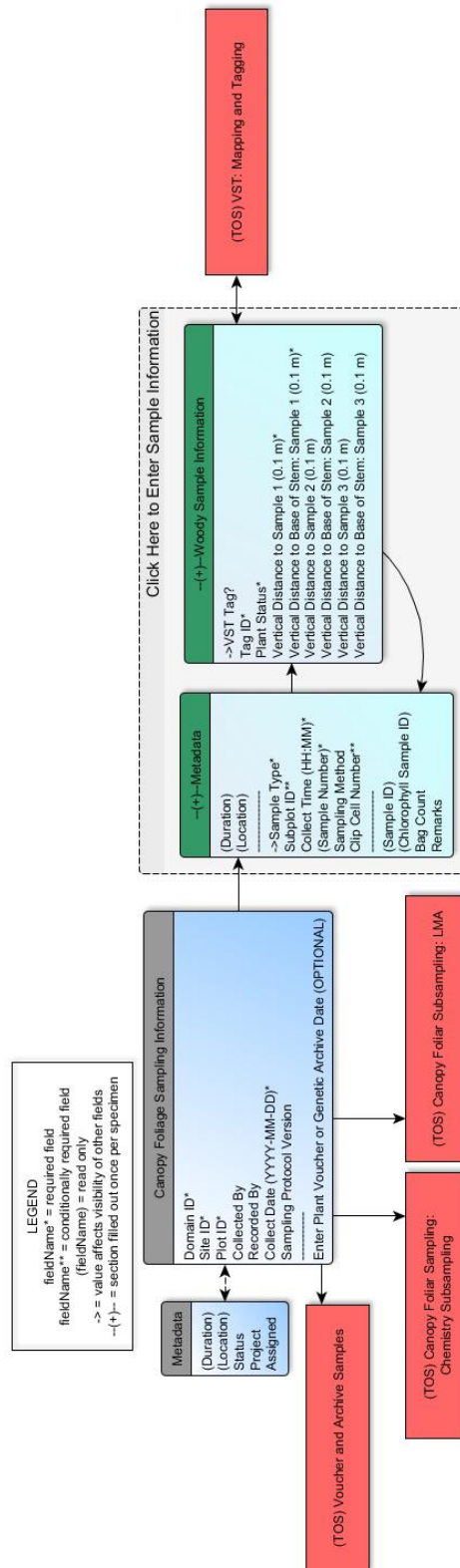


Figure 3: Schematic of the application used by field technicians to enter foliar sampling field data