



neon
Operated by Battelle

<i>Title:</i> NEON User Guide to Plant phenology observations (DP1.10055.001)	<i>Date:</i> 01/24/2023
<i>Author:</i> Katherine Jones	<i>Revision:</i> D

NEON USER GUIDE TO PLANT PHENOLOGY OBSERVATIONS (DP1.10055.001)

PREPARED BY	ORGANIZATION
Katherine Jones	FSU



<i>Title:</i> NEON User Guide to Plant phenology observations (DP1.10055.001)	<i>Date:</i> 01/24/2023
<i>Author:</i> Katherine Jones	<i>Revision:</i> D

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	07/19/2017	Initial Release
B	10/01/2020	Added information about about use of neonUtilities R package, updated taxonomy information, detailed sampling design changes to date
C	05/04/2022	Added language in section 4 Taxonomy addressing RTE species obfuscation in the data. Updated section 5.3 Data Revision with latest information regarding data release. Updated information regarding the geoNEON package
D	11/14/2022	Added sampling schedule table. Added language referencing the neonOS R package.



<i>Title:</i> NEON User Guide to Plant phenology observations (DP1.10055.001)	<i>Date:</i> 01/24/2023
<i>Author:</i> Katherine Jones	<i>Revision:</i> D

TABLE OF CONTENTS

1	DESCRIPTION	1
1.1	Purpose	1
1.2	Scope	1
2	RELATED DOCUMENTS AND ACRONYMS	2
2.1	Associated Documents	2
2.2	Acronyms	2
3	DATA PRODUCT DESCRIPTION	3
3.1	Spatial Sampling Design	4
3.2	Temporal Sampling Design	5
3.3	Sampling Design Change	7
3.3.1	Growth Form updates	7
3.3.2	Discontinued Sampling Locations	8
3.3.3	Addition of Reproductive Phenophases in Alaska	8
3.4	Variables Reported	8
3.5	Spatial Resolution and Extent	9
3.6	Temporal Resolution and Extent	11
3.7	Associated Data Streams	11
3.8	Product Instances	12
3.9	Data Relationships	12
4	TAXONOMY	13
5	DATA QUALITY	13
5.1	Data Entry Constraint and Validation	13
5.2	Automated Data Processing Steps	14
5.3	Data Revision	14
5.4	Quality Flagging	14
6	DATA PROCESSING STEPS	15



<i>Title:</i> NEON User Guide to Plant phenology observations (DP1.10055.001)	<i>Date:</i> 01/24/2023
<i>Author:</i> Katherine Jones	<i>Revision:</i> D

7 REFERENCES

15

LIST OF TABLES AND FIGURES

Table 1	Summary of Phase I/II transitions and sampling schedule	6
Table 2	Reference pointIDs relative to based on transect meter	10
Table 3	Offset direction based on transect meter and direction from transect values	10
Table 4	Descriptions of the dataQF codes for quality flagging	14
Figure 1	Placement of phenology transect relative to the NEON tower, airshed and tower plots	5



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 Plant phenology observations - Repeat phenophase status and intensity measurements collected from individuals along an 800-m transect or from within a fixed plot - 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 Plant Phenology Observations (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: Plant Phenology (AD[07]). The raw data that are processed in this document are detailed in the file, NEON Raw Data Ingest Workbook for TOS Plant Phenology Observations (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.000907	TOS Science Design for Plant Phenology
AD[07]	NEON.DOC.014040	TOS Protocol and Procedure: Plant Phenology
AD[02]	NEON.DOC.000987	TOS Protocol and Procedure: Measurement of Vegetation Structure
AD[09]	NEON.DOC.000912	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.004825	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[13]	Available on NEON data portal	NEON Ingest Conversion Language Function Library
AD[14]	Available on NEON data portal	NEON Ingest Conversion Language
AD[15]	Available with data download	Categorical Codes csv

2.2 Acronyms

Acronym	Definition
USA-NPN	USA National Phenology Network



3 DATA PRODUCT DESCRIPTION

The plant phenology observations product provides in-situ observations of the phenological status and intensity of tagged plants (or patches) during discrete observation events. NEON employs status-based monitoring, in which the phenological condition of an individual is reported any time that individual is observed. At every observation bout, records are generated for every phenophase that is occurring and for every phenophase not occurring. With this approach, events (such as leaf emergence in Mediterranean climates, or flowering in many desert species) that may occur multiple times during a single year, can be captured. Continuous reporting of phenophase status enables quantification of the duration of phenophases rather than just their date of onset while allowing the explicit quantification of uncertainties in phenophase transition dates that are introduced by monitoring in discrete temporal bouts.

Observation bouts are completed by NEON field technicians and occur at terrestrial sites according to site- and season-specific sampling intervals. Species and sampling schedules for plant phenological observations vary between sites. For an overview of the data collection strategy, see the NEON TOS Science Design - Plant Phenology AD[06], Elmendorf 2016, as well as the plant phenology protocol AD[07]).

Products derived from this sampling include the observed phenophase status (whether or not a phenophase is occurring) and the intensity of phenophases for which phenophase status = 'yes'. According to the USA National Phenology Network (USA-NPN), a phenophase is defined as 'an observable stage or phase in the annual life cycle of a plant or animal that can be defined by a start and end point'. Phenophase intensity, then, quantifies the degree to which a given phenophase is being expressed (continuous phenomenon, although recorded in bins). The phenophase categories employed by NEON are derived from the USA-NPN (Denny 2014; www.usanpn.org); although not all reproductive phenophases are monitored by NEON technicians due to labor constraints. Taxonomic identification and geolocation data (collected upon tagging of plants, though possibly updated through time), as well as an annual survey of plant size and condition are also provided.

Species for phenology observation are selected based on relative abundance in the Tower airshed. To determine relative abundance, taxonomic and vegetation structure surveys are conducted in all Tower base plots in advance of transect and plot establishment according to the Plant Diversity and Vegetation Structure protocols (AD[09] and AD[08] respectively). Phenology occurs in two distinct phases at each site. **Phase I** focuses on intraspecific variation of phenology within dominant species at the site; 30 individuals of only the 3 most common species at the site are selected for monitoring. **Phase II** addresses community phenology and highlights interspecific variation; in this phase up to 20 species are selected with fewer individuals selected for monitoring per species. For the list of species selected and the schedule for Phase I and Phase II observations at each NEON site, refer to Appendix C of TOS Protocol and Procedure: Plant Phenology (AD[07]).

The NEON phenology measurements track sensitive and easily observed indicators of biotic responses to climate variability by monitoring the timing and duration of phenological stages in plant communities. Plant phenology is affected by forces such as temperature, timing and duration of pest infestations and disease outbreaks, water fluxes, nutrient budgets, carbon dynamics, and food availability and has feedbacks to trophic interactions, carbon sequestration, community composition and ecosystem function.



3.1 Spatial Sampling Design

Plant phenology observations occur at all terrestrial NEON sites along an 800 meter square loop transect (primary). The primary transect is intentionally positioned within the primary airshed, to align with flux measurements collected by instrumentation mounted on the NEON tower and alongside Tower base plots to facilitate selection of individuals in both the plant phenology and vegetation structure protocols (Figure 1). At sites where the primary airshed is not located directly North of the tower, an additional 200 m x 200 m plot is established within view of a canopy level, tower-mounted, phenology camera (phenocam). Both the phenocam plot and the primary transect are oriented in the four cardinal directions and are situated within the tower airshed where possible (Figure 1). Selected individuals along the primary transect are located 1-10 m from the transect in order to minimize trampling and to allow observations without requiring field personnel to leave the transect. To the extent possible, given the heterogeneity in the plant community around the transect, selected individuals are distributed evenly around the transect. At sites where the loop transect is located directly north of the NEON Tower, no additional phenocam plot is established. See TOS Science Design for Plant Phenology (AD[06]), TOS Protocol and Procedure: Plant Phenology (AD[07]), TOS Science Design for Spatial Sampling (AD[02]) and TOS Science Design for Plant Diversity (AD[09]) for further details.

As much as possible, sampling occurs in the same locations over the lifetime of the Observatory. However, over time some sampling locations may become impossible to sample, due to disturbance or other local changes. When this occurs, the location and its location ID are retired. A location may also shift to slightly different coordinates. Refer to the locations endpoint of the NEON API for details about locations that have been moved or retired: <https://data.neonscience.org/data-api/endpoints/locations/>

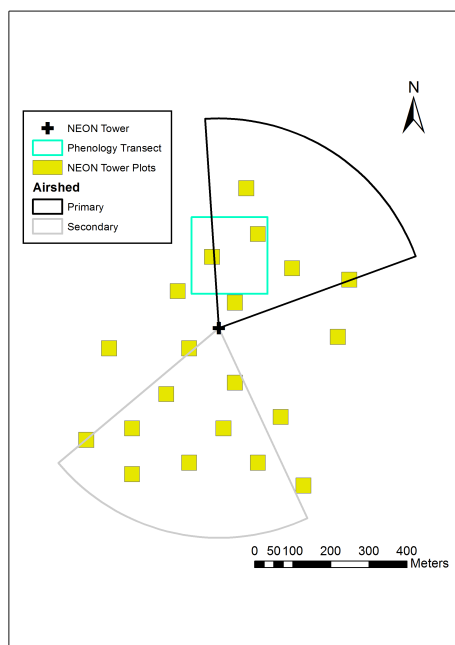


Figure 1: Placement of phenology transect relative to the NEON tower, airshed and tower plots

3.2 Temporal Sampling Design

All individuals/patches selected for phenology monitoring within a site are monitored with the same frequency; the primary transect and the phenocam plot are monitored on the same schedule. Sampling schedules are either 1) year round, 2) seasonal, or 3) hybrid (Table 1).

Year round sampling schedule. At sites with limited or no pronounced dormant season (e.g., tropical, southeast), or >1 growing season/year (e.g., desert southwest), observation bouts will occur throughout the entire year at a frequency of 1x/ week.

Seasonal sampling schedule. At sites with well-defined, discrete, growing seasons, the seasonal sampling frequency varies to capture rapid changes during phenological transition periods. The sampling season begins with at least one observation each year within 7 days prior to the onset of springtime activity (variably defined as breaking leaf buds, breaking needle buds, emerging needles, or initial growth; see column 2 in Table 1). Intensive sampling (2-3x/week) occurs in conjunction with the onset of springtime phenological activity and continues through the early spring development. Once >50% of leaf/canopy development has occurred sampling can be reduced to once a week until full canopy has developed. Post 95% canopy development, sampling is further reduced to once every other week until reproductive phenophases are complete. A second intensive stage (sampling twice a week) begins again in the fall to capture leaf senescence/coloring, then reduces to once a week through the end of the season.



Hybrid sampling schedule. In some sites, especially those in arid or semiarid environments, phenology may be driven by moisture rather than temperature. In this case, sampling dates and frequency may be more episodic but sampling has the same goal of maintaining high frequency during periods of rapid change, but are flexible to start and end sampling multiple times throughout the year. One example of a hybrid sampling schedule is starting with year round regular sampling (1x/week), increasing frequency during periods of growth, then dropping to <1/week during dormant periods. The varied intensity is intended to strategically use sampling periods in order to monitor phases of rapid phenological change, while minimizing labor/disturbance associated with frequent measurements during times of year of less rapid change or inactivity.

Phenological observations are recorded on all individuals each time monitoring is performed, unless the end-of-season phenophase/trigger has been reached for a particular individual, in which case, monitoring of that individual is not required for the remainder of the season. See TOS Protocol and Procedure: Plant Phenology (AD[07]) for additional details on phenophase events that trigger changes to observation frequency.

Table 1: Summary of Phase I/II transitions and sampling schedule

Domain ID	Site ID	Sampling locations	Phase I schedule	Phase I First Year	Phase II schedule	Phase II First Year
D01	BART	2	Seasonal	2014	Seasonal	2021
D01	HARV	2	Seasonal	2013	Seasonal	2021
D02	BLAN	2	Seasonal	2015	Year round	2021
D02	SCBI	2	Seasonal	2014	Year round	2021
D02	SERC	2	Seasonal	2015	Year round	2021
D03	DSNY	2	Year round	2013	Year round	2021
D03	JERC	2	Year round	2014	Year round	2021
D03	OSBS	2	Year round	2013	Year round	2021
D04	GUAN	2	Year round	2015	Year round	2021
D04	LAJA	1	Year round	2016	Year round	2021
D05	TREE	2	Seasonal	2015	Seasonal	2023
D05	UNDE	2	Seasonal	2014	Seasonal	2022
D06	KONZ	2	Seasonal	2015	Seasonal	2021
D06	UKFS	2	Seasonal	2015	Year round	2021
D07	GRSM	2	Seasonal	2015	Seasonal	2021
D07	MLBS	1	Seasonal	2017	Seasonal	2022
D07	ORNL	2	Seasonal	2014	Seasonal	2021
D08	DELA	2	Year round	2015	Year round	2021
D08	TALL	1	Year round	2014	Year round	2021
D08	LENO	2	Hybrid	2016	Year round	2021



Domain ID	Site ID	Sampling locations	Phase I schedule	Phase I First Year	Phase II schedule	Phase II Sampling First Year
D09	NOGP	2	Seasonal	2016	Seasonal	2022
D09	WOOD	2	Seasonal	2014	Seasonal	2022
D10	CPER	2	Year round	2013	Year round	2021
D10	RMNP	2	Seasonal	2017	Seasonal	2022
D11	CLBJ	2	Year round	2016	Year round	2022
D11	OAES	2	Year round	2015	Year round	2022
D12	YELL	2	Hybrid	2018	Year round	2023
D13	MOAB	2	Year round	2015	Year round	2022
D13	NIWO	2	Hybrid	2015	Year round	2021
D14	JORN	2	Year round	2015	Year round	2022
D14	SRER	2	Year round	2016	Year round	2021
D15	ONAQ	2	Year round	2014	Year round	2022
D16	ABBY	2	Seasonal	2016	Year round	2021
D16	WREF	2	Seasonal	2018	Year round	2023
D17	SJER	2	Year round	2017	Year round	2022
D17	SOAP	2	Year round	2018	Year round	2023
D17	TEAK	2	Hybrid	2019	Seasonal	2024
D18	BARR	2	Seasonal	2017	Seasonal	2022
D18	TOOL	2	Seasonal	2017	Seasonal	2022
D19	BONA	2	Seasonal	2017	Seasonal	2022
D19	DEJU	2	Seasonal	2016	Seasonal	2021
D19	HEAL	2	Seasonal	2015	Seasonal	2021
D20	PUUM	1	Year round	2018	Year round	2023

3.3 Sampling Design Change

3.3.1 Growth Form updates

Following one or more years of sampling for some species, the initially assigned growth form was updated to better reflect the phenology of the species. For these species, data collected prior to the update may include a slightly different set of phenophases:



Scientific Name	Initial Growth Form	Updated Growth Form	Update Date	Sites
<i>Yucca elata</i>	Evergreen broadleaf	Evergreen broadleaf - no leaves	2017-10	JORN
<i>Larrea tridentata</i>	Evergreen broadleaf	Evergreen broadleaf - no leaf buds	2017-12	SRER
<i>Ephedra viridis</i>	Evergreen broadleaf	Evergreen broadleaf - no leaves	2018-05	MOAB
<i>Acacia koa</i>	Evergreen broadleaf	Evergreen broadleaf - no leaf buds	2020-09	PUUM

3.3.2 Discontinued Sampling Locations

Domain	Named Location	Date
D05	STEI	2018-01
D06	KONA	2018-01
D09	DCFS	2018-01
D10	STEI	2018-01

3.3.3 Addition of Reproductive Phenophases in Alaska

In 2022, sites in Alaska began making observations of ‘Fruits’ and ‘Ripe fruits’ phenophases on berry species. Data for years prior to 2022 do not include these phenophases.

Domain	Named Location	Species
D18	TOOL	<i>Empetrum nigrum</i> , <i>Rubus chamaemorus</i> , <i>Vaccinium uliginosum</i> , <i>Vaccinium vitis-idaea</i>
D19	BONA	<i>Rosa acicularis</i> , <i>Rubus chamaemorus</i> , <i>Vaccinium uliginosum</i> , <i>Vaccinium vitis-idaea</i>
D19	DEJU	<i>Empetrum nigrum</i> , <i>Rosa acicularis</i> , <i>Vaccinium uliginosum</i> , <i>Vaccinium vitis-idaea</i>
D19	HEAL	<i>Empetrum nigrum</i> , <i>Rubus chamaemorus</i> , <i>Vaccinium uliginosum</i> , <i>Vaccinium vitis-idaea</i>

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 Plant Phenology Observations (AD[04]). All variables reported in the published data (L1 data) are also provided separately in the file, NEON Data Publication Workbook for TOS Plant Phenology Observations (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 GEOID09 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 spatial resolution at which phenology data will be tracked is per individual ($n \leq 100$ individuals per primary phenology plotID).

individualID (unique ID given to the tagged individual or patch) → **plotID** (unique ID given to transect or plot) → **siteID** (ID of NEON site) → **domainID** (ID of a NEON domain).

The basic spatial data included in the data downloaded include the latitude, longitude, and elevation of the centroid of the plot where sampling occurred + associated uncertainty due to GPS error. 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 individual, 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. The namedLocation field in the data is the named location of the transect or plot; more precise geographic data require the named location of the nearest pointID. Construct the named location of the pointID associated with each record in phe_perindividual by first looking up the pointID of the two pointID on either side of the transectMeter labeled referencePoint_A and referencePoint_B respectively (Table 2) then concatenating the fields for namedLocation and referencePoint as: namedLocation + '' + referencePoint, e.g. referencePoint 'W' of namedLocation 'HARV_053.phenology.phe' has a complete named location of 'HARV_053.phenology.phe.W'.



Table 2: Reference pointIDs relative to based on transect meter

phe_perindividual_pub:transectMeter	refPt_A	transectMeter _{refPt_A}	refPt_B	transectMeter _{refPt_B}
0 <transectMeter ≤ 100	SW	0	W	100
100 <transectMeter ≤ 200	W	100	NW	200
200 <transectMeter ≤ 300	NW	200	N	300
300 <transectMeter ≤ 400	N	300	NE	400
400 <transectMeter ≤ 500	NE	400	E	500
500 <transectMeter ≤ 600	E	500	SE	600
600 <transectMeter ≤ 700	SE	600	S	700
700 <transectMeter ≤ 800	S	700	SW	800

- Use the API (http://data.neonscience.org/api/v0/locations/HARV_053.phenology.phe.W) to query for elevation (“locationElevation”), easting (“locationUtmEasting”), northing (“locationUtmNorthing”), coordinateUncertainty (“Value for Coordinate uncertainty”), elevationUncertainty (“Value for Elevation uncertainty”), and utmZone (“locationUtmZone”) as inputs to the next step.
- Assign the offsetDirection based on the value in phe_perindividual_pub.transectMeter and phe_perindividual_pub.directionFromTransect according to guidelines in Table 3.

Table 3: Offset direction based on transect meter and direction from transect values

phe_perindividual_in:transectMeter	phe_perindividual_in: directionFromTransect	offsetDirection
0 <transectMeter ≤ 200	Left	West
	Right	East
200 <transectMeter ≤ 400	Left	North
	Right	South
400 <transectMeter ≤ 600	Left	East
	Right	West
600 <transectMeter ≤ 800	Left	South
	Right	North

- Determine the geolocation of the transectMeter for each individual in utms as the weighted mean of the northing (in utms) of reference_pointA and reference_pointB, and the weighted mean of the easting referencePoint_A and referencePoint_B, where the weighted mean is defined as:

$$\bar{x} = \sum_{i=1}^n w_i x_i \tag{1}$$



where x_i is the northing or easting of the reference points, and

$$w_b = \frac{\text{transectMeter}_i - \text{transectMeter}_{\text{refPt}_A}}{100} \quad (2)$$

where $\text{transectMeter}_{\text{refPt}_A}$ is the value from Table 2 and

$$w_a = 1 - w_b \quad (3)$$

5. Calculate the geolocation of the actual plant by adding or subtracting the ninetyDegreeDistance from the geolocation of the transectMeter, according to the offsetDirection specified in Table 3.

Example: transectMeter=85, ninetyDegreeDistance=4, directionFromTransect=right. An observer is assumed to be 85% of the way between the SW and W reference points of the plot. If the SW reference point is at (0,0) (easting and northing, in utms), and the W reference point is at (-5, 103), for example, the true position of the transectMeter is at easting = $(0.15 \times 0 + 0.85 \times -5) = -4.25$ and northing = $(0.15 \times 0 + 0.85 \times 103) = 87.55$ Since the observer is between 0-200m on the transect, s/he is traveling north, so a ninetyDegreeDistance of 4 to the right is equivalent to an offset of 4m E. The position of the plant in utms is (87.55, -0.25).

6. Increase coordinateUncertainty by an appropriate amount (suggested 5 m) to account for error introduced by navigating around the transect. Error may also be introduced due to tape stretching on initial establishment and deviations from the the transect in order to avoid obstacles.

3.6 Temporal Resolution and Extent

The finest resolution at which temporal data are reported is the **date** or **dayOfYear**, the status of each phenophase, for a tagged individual, on the day of observation.

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. Combined, data from each observation **date** contribute to a time course that may be used to assess phenophase transition dates, duration and change across years. Code to stack files across months is available here: <https://github.com/NEONScience/NEON-utilities>

3.7 Associated Data Streams

individualID is a linking variable that ties phenophase observations and associated metadata to the Vegetation structure (DP1.10098.001) and the Plant foliar traits (DP1.10026.001) data products. Users are advised to leverage data from all associated data products to provide context to phenology data.

3.8 Product Instances

There are approximately 50 plant phenology observation bouts per calendar year at every terrestrial site. There are no more than 100 sampling points (individuals or patches) located along the phenology transect and no more than 9 sampling points within the phenocam plot. The number of phenophases observed varies by growth form and ranges from 1 phenophase (cactus) to 7 phenophases (semi-evergreen broadleaf). In addition to the 50 phenophase observation bouts, is a bout to collect annual size and disease status measurements from each sampling point. Thus, no single site should ever exceed 38,259 data product instances in a given calendar year.

3.9 Data Relationships

The protocol dictates that each sampling location (individual or patch) is established once. Each location is tagged and assigned an identifier that is unique across the Observatory; phe_perindividual.individualID. Observations of phenophase status in phe_statusintensity and annual measurements of size and health conditions in phe_perindividualperyear are linked to location metadata contained in phe_perindividual via **individualID**. 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.

phe_perindividual - > One record expected per individualID for all time. Duplicate records may exist if location or taxon information is updated. If duplicate records are present in this table, users are advised to retain the record with the latest **editedDate**.

phe_perindividualperyear - > One record expected per individualID per year. Duplicate records for a given **individualID** may exist if measurements are collected on different days in which case users are advised to merge non-null values for a given **individualID**. If full duplicate records are present in this table, or duplicates for which there are conflicting values for a given field, users are advised to retain the record with the latest **editedDate**.

phe_statusintensity - > One record expected per individualID x phenophaseName combination for a given date. Duplicate records may exist if edits are made to a record (e.g. phenophaseStatus = 'uncertain', edited to either a 'yes' or 'no' based on review with lead technician). If duplicate records are present in this table, users are advised to retain the record with the latest **editedDate**.

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

The neonOS package provides functions to facilitate resolution of duplicate records and joining tables. The neonOS package is available from Comprehensive R Archive Network (CRAN; <https://cran.r-project.org/web/packages/neonOS/index.html>) and can be installed using the install.packages() function in R. For instructions on using neonOS to check for duplicates and join tables, see the Using the neonOS Package



to Check for Duplicates and Join tables tutorial on the NEON website: <https://www.neonscience.org/resources/learning-hub/tutorials/neonos-duplicates-joins>

4 TAXONOMY

NEON manages taxonomic entries by maintaining a master taxonomy list based on the community standard, if one exists. Through the master taxonomy list, synonyms submitted in the data are converted to the appropriate name in use by the standard. The master taxonomy for plants is the USDA PLANTS Database (USDA, NRCS. 2014. <https://plants.usda.gov>). Taxon ID codes used to identify taxonomic concepts in the NEON master taxonomy list are alpha-numeric codes, 4-6 characters in length based on the accepted scientific name. Each code is composed of the first two letters of the genus, followed by the first two letters of the species and first letter of the terminal infraspecific name (if applicable) then, if needed, a tiebreaking number to address duplicate codes. Genus and family symbols are the first five (genus) or six (family) letters of the name, plus tiebreaking number (if needed). Symbols were first used in the Soil Conservation Service's National List of Scientific Plant Names (NLSPN) and have been perpetuated in the PLANTS system. The portions of the PLANTS Database included in the NEON plant master taxonomy list includes native and naturalized plants present in NEON observatory sampling area including the Lower 48 U.S. States, Alaska, Hawaii, and Puerto Rico. NEON plans to keep the taxonomy updated in accordance with USDA PLANTS Database starting in 2020 and annually thereafter.

The master taxonomy list includes geographic range and nativity as described by the USDA PLANTS Database. A list for each NEON domain includes those species with ranges that overlap the domain as well as nativity designations - introduced or native - in that part of the range. Errors are generated if a species is reported at a location outside of its known range. If the record proves to be a reliable report, the master taxonomy table is updated to reflect the distribution change.

Prior to the 2022 data release, publication of species identifications were obfuscated to a higher taxonomic rank when the taxon was found to be listed as threatened, endangered, or sensitive at the state level where the observation was recorded. The state-level obfuscation routine was removed from the data publication process at all locations excluding sites located in D01 and D20, and data have been reprocessed to remove the obfuscation of state-listed taxa for all years. Federally listed threatened and endangered or sensitive species remain obfuscated at all sites and sensitive species remain redacted at National Park sites.

The full master taxonomy lists are available on the NEON Data Portal for browsing and download: <http://data.neonscience.org/static/taxon.html>.

5 DATA QUALITY

5.1 Data Entry Constraint and Validation

Many quality control measures are implemented at the point of data entry within a mobile data entry application or web user interface (UI). For example, data formats are constrained and data values



controlled through the provision of dropdown options, which reduces the number of processing steps necessary to prepare the raw data for publication. 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 Plant Phenology Observations, 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]).

5.2 Automated Data Processing Steps

Following data entry into a mobile application or web user interface, the steps used to process the data through to publication on the NEON Data Portal are detailed in the NEON Algorithm Theoretical Basis Document: OS Generic Transitions (AD[14]).

5.3 Data Revision

All data are provisional until a numbered version is released. Annually, NEON releases a static version of all or almost all data products, annotated with digital object identifiers (DOIs). The first data Release was made in 2021. During the provisional period, QA/QC is an active process, as opposed to a discrete activity performed once, and records are updated on a rolling basis as a result of scheduled tests or feedback from data users. The Issue Log section of the data product landing page contains a history of major known errors and revisions.

5.4 Quality Flagging

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

Table 4: Descriptions of the dataQF codes for quality flagging

fieldName	value	definition
dataQF	legacyData	Data recorded using a paper-based workflow that did not implement the full suite of quality control features associated with the interactive digital workflow

Records of land management activities, disturbances, and other incidents of ecological note that may have a potential impact are found in the Site Management and Event Reporting data product (DP1.10111.001)



6 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[12]).

7 REFERENCES

Denny, E. G., K. L. Gerst, A. J. Miller-Rushing, G. L. Tierney, T. M. Crimmins, C. A. Enquist, P. Guertin, A. H. Rosemartin, M. D. Schwartz, and K. A. Thomas. 2014. Standardized phenology monitoring methods to track plant and animal activity for science and resource management applications. *International Journal of Biometeorology*:1–11.

Elmendorf, S. C., K. D. Jones, B. I. Cook, J. M. Diez, C. A. Enquist, R. A. Hufft, M. O. Jones, S. J. Mazer, A. J. Miller-Rushing, D. J. Moore, and others. 2016. The plant phenology monitoring design for the national ecological observatory network. *Ecosphere* 7.

National Phenology Network. "Nature's Notebook." https://www.usanpn.org/natures_notebook.

USDA, Forest Service. 2011. Forest Inventory and Analysis Nation Core Field Guide. Volume I: Field Data Collection Procedures for Phase 2 Plots. Version 5.1

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