NEON USER GUIDE TO SITE MANAGEMENT AND EVENT REPORTING (DP1.10111.001)

PREPARED BY	ORGANIZATION
Rachel Swanson	Science
Tanya A. Vance	Science

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	07/30/2019	Initial Release
B	10/16/2020	Included general statement about usage of neonUtilities R package and statement about possible location changes. Updated section 3.5 and 3.8 to include information about recording events in one month intervals and taxonomy information.
C	04/08/2022	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
D	04/22/2025	Added sample code using geoNEON in the Appendix. Updated the url for spatial data in section 3.4. Added information about the new neonUtilities Python package.

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 description of the disturbance 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, Site management and event reporting - record of land management activities, disturbances, and other incidents of ecological note within a NEON site - 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 Site management and event reporting (DP1.10111.001) (AD[04]), provided in the download package for this data product.

This document describes the process for ingesting and performing automated quality assurance and control procedures on the data collected in the field pertaining to NEON Protocol and Procedure: Site management and disturbance data collection (AD[06]). The raw data that are processed in this document are detailed in the file, NEON Raw Data Ingest Workbook for Site management and event reporting (DP0.10111.001) (AD[03]), provided in the download package for this data product. Please note that raw data products (denoted by ‘DP0’) may not always have the same numbers (e.g., ‘10111’) as the corresponding L1 data product.

2 RELATED DOCUMENTS

2.1 Associated Documents

AD[01]	NEON.DOC.000001	NEON Observatory Design (NOD) Requirements
AD[02]	NEON.DOC.002652	NEON Data Products Catalog
AD[03]	Available with data download	Validation csv
AD[04]	Available with data download	Variables csv
AD[05]	NEON.DOC.001152	NEON Aquatic Sampling Strategy
AD[06]	NEON.DOC.003282	NEON Protocol and Procedure: Site management and disturbance data collection
AD[07]	NEON.DOC.000913	TOS Science Design for Spatial Sampling
AD[08]	NEON.DOC.000008	NEON Acronym List
AD[09]	NEON.DOC.000243	NEON Glossary of Terms
AD[10]	NEON.DOC.004825	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[11]	Available on NEON data portal	NEON Ingest Conversion Language Function Library
AD[12]	Available on NEON data portal	NEON Ingest Conversion Language
AD[13]	Available with data download	Categorical Codes csv

3 DATA PRODUCT DESCRIPTION

The Site management and event reporting data product (DP1.10111.001) provides records of management activities and stochastic disturbances within all NEON sites. Records are reported about activities and disturbances that have the potential to impact or affect data products across all NEON data-generating systems (TIS, TOS, AIS, AOS, and AOP) that are not collected as part of existing data product protocols.

Each NEON site experiences a range of planned and unplanned events at varying temporal and spatial scales. Without context, many of these activities and perturbations could be interpreted by data users as a response to the aforementioned forcings. For example, a site may have a history of applying herbicides to control an invasive plant and, during the course of the study, the landowner ceases these activities and the plant begins to recolonize the study area. Without the knowledge of the changes in land-management practices, the data user could make inferences that are incorrect or that do not account for the other variables that may be causing the observed variation. Equally important are the random events that may impact multiple data products across plots and aquatic reaches or within individual plots and transects. Knowledge of burns, wind damage, flooding, erosional processes, and the like are all important to the integrity and utility of NEON data products.

3.1 Spatial Sampling Design

Site management and event reporting occurs at all NEON terrestrial and aquatic sites (Figure 1). See NEON Aquatic Sampling Strategy (AD[05]), NEON Protocol and Procedure: Site management and disturbance data collection (AD[06]), and TOS Science Design for Spatial Sampling (AD[07]) 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: Generic NEON site depicting airborne, aquatic, and terrestrial data collection.

3.2 Temporal Sampling Design

Site management activity and disturbance reporting occurs at the time NEON staff are aware of the planned or unplanned event, either through direct observation during normal scheduled site visits or secondary reports from landowners, site hosts, and other reliable sources. For planned events, reporting is expected to occur only after the management activity has been started, as the data user is presumed to be not interested in planned activities that did not occur. For unplanned events, data recording takes place as soon as the disturbance is observed as well as when it is safe to assess the scale and intensity of the impact, preferably within 10 days of the observation. See NEON Aquatic Sampling Strategy (AD[05]) and the NEON Protocol and Procedure: Site management and disturbance data collection (AD[06]).

3.3 Variables Reported

All variables reported from the field (L0 data) are listed in the file, NEON Raw Data Ingest Workbook for Site management and event reporting (DP0.10111.001) (AD[03]). All variables reported in the published data (L1 data) are also provided separately in the file, NEON Data Publication Workbook for Site management and event reporting (DP1.10111.001) (AD[04]).

Where possible or relevant, 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 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.

3.4 Spatial Resolution and Extent

The finest resolution at which site management and event reporting data are reported is at the spatial scale at which the planned or unplanned event occurred: at the site, reach or airshed level, at a group of specified plots, or at the level of an individual sampling location. The reporting level is the lowest possible demarcation (such as to plot level, if possible, rather than just a site reporting). The site management and event reporting data capture the NEON assets that were impacted (e.g., plot numbers, airshed, stream reach) and where possible, quantify the extent of the manipulation or disturbance (e.g., square meters, hectares). Small spatio-temporal scale events typically occur at the plot level for TOS, the transect level for AOS, the sensor location for AIS, and the tower for TIS.

Overall,this results in a spatial hierarchy of:

locationID (finest spatial resolution, ID of location within site) → **siteID** (ID of NEON site) → **domainID** (ID of a NEON domain).

Shapefiles of all NEON Terrestrial Observation System sampling locations can be found on the NEON sci-

ence webpage at https://www.neonscience.org/data-samples/data/spatial-data-maps. If users are interested in the geospatial locations of the data relative to a global coordinate system, those can be retrieved using the NEON data API using the **namedLocation** and the following:

1. The getLocTOS() function from the geoNEON package, available here: https://github.com/NEONScience/NEON-geolocation
    - See the Appendix for sample code using the geoNEON::getloc() function
2. The NEON API: http://data.neonscience.org/api

3.5 Temporal Resolution and Extent

In order to match the structure of NEON data products and to ensure events are published as quickly as possible, site management and event reporting data are recorded in thirty day intervals. This means events longer than one month will have multiple records for one eventID. The eventID, which includes the site code, original start date, and event type, will remain the same for all related records. The original start date in the eventID compared to the start date for the record indicates if there are related records for previous months. The "ongoingEvent" field indicates if a record for the next subsequent month is expected.

Date uncertainty is captured in a record's "estimatedOrActualDate" field. If the date is "estimated" additional comments are recorded in the "dateRemarks" field. The start and end date range includes any date uncertainty.

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.

3.6 Associated Data Streams

3.7 Product Instances

The site management and event reporting data are collected at all 81 NEON Observatory field sites. These 81 sites are divided into 47 terrestrial sites and 34 aquatic sites, consisting of 24 wadeable streams, 3 non-wadeable streams (rivers), and 7 lakes. The number of records per site per year are in flux and dependent on the variable management activities and disturbances at each site.

3.8 Data Relationships

The protocol dictates that each site event is reported once, yielding a unique **eventID**. Since site management and event reporting data are published in one month intervals, events longer than one month will have multiple records for one eventID. In addition, site events with differing intensity levels at different locations (for example, fire) will also have matching eventIDs. 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.

sim_eventData.csv - > One record expected for each month per **eventID**. The eventID includes the initial startDate of the event.

Data presented in this data product could potentially relate to any other data product in our catalog-linked through the siteID or locationID and time range. Users should be aware that the dates may not match exactly.

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 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 Site management and event reporting (DP0.10111.001), 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[12]).

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.

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[11]).

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.

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. Data recorded before August 2019 are considered legacy data for Site Management and Event Reporting.

6 REFERENCES

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

7 APPENDIX

Below is sample code to extract spatial locations from SIM data using geoNEON::getLocTOS() in R.

```
# load libraries
library(devtools)
library(neonUtilities)
library(geoNEON)
library(tidyverse)

install_github('NEONScience/NEON-geolocation/geoNEON', dependencies=TRUE)

# download SIM data
simData <- loadByProduct(
  dpID='DP1.10111.001',
  check.size=F,
  site = "HARV",
  package='basic',
  # Optional API token for faster downloads. See the API token tutorial
  token = Sys.getenv('NEON_PAT'))

# Turn all tables in the list to dataframe (DF) in the global environment
# The name of table = name of DF.
list2env(simData, envir=.GlobalEnv)

# if warranted, filter to rows you are interested, for example eventType = droughtPerceived
HARV_drought <- sim_eventData %>%
  dplyr::filter(eventType == "droughtPerceived")

# or SIM events that impact a subset of plots.  For example, mammal grids.
#Note that "HARV," is in the search to include events that impacted the entire site
HARV_mam <- sim_eventData %>%
```

```
  dplyr::filter(grepl("mam|HARV,", locationID))

# Use the geoNEON function getLocTOS to get a nested DF for all locationIDs in sim_eventData
# For each data record, the locationID field will be populated with a data frame,
# containing the spatial data for each location noted in the record.
# For each location, the nested data frame contains the centroid point coordinates.
# See the "polygon" or "plot dimension" fields for necessary values to create
# the associated polygons.
# The location "AIRSHED" is not recorded in the NEON database of named locations,
# so no spatial data will be returned for it.
# See the "NEON TIS Airshed" layer in NEON's ArcGIS Online portal
# https://neon.maps.arcgis.com/home/index.html for a shapefile of all tower airsheds.
drought_spatialData <- getLocTOS(data = HARV_drought,
                                 dataProd = 'sim_eventData',
                                 token = Sys.getenv('NEON_PAT'))

# mammal example:
mam_spatialData  <- getLocTOS(data = HARV_mam,
                              dataProd = 'sim_eventData',
                              token = Sys.getenv('NEON_PAT'))

# To get the set of locations for one event record:
droughtStart <- drought_spatialData %>%
  dplyr::filter(startDate==as.POSIXct("2017-05-01", tz="GMT")) %>%
  dplyr::select(locationID) %>%
  .[[1]] %>%
  .[[1]]

# Use tidyr::unnest() to row-bind the list-column spatial values in the nested DF
# This creates a data frame with a row for each combination of location and event,
# so there will typically be many rows for each event.
drought_spatialData_all <- unnest(drought_spatialData,
                                  cols=locationID,
                                  names_sep="_")

# One possible approach to summarizing these data, using the mammal example:
# 1. Filter the data frame to just rows for mammal plots or the entire site
# 2. Group by eventID. Events are recorded in 30-day intervals throughout the
# duration of the event.
# 3. Use mutate() to calculate the earliest and latest date for each eventID.
# 4. De-duplicate using distinct()
# End result: a table in which each record represents the full duration of a given event,
# at a given location relevant to mammal sampling.
mam_spatialData_all <- unnest(mam_spatialData,
```

```
                            cols=locationID,
                            names_sep="_") %>%
dplyr::filter(locationID_subtype=="mammalGrid" | locationID_namedLocation=="HARV") %>%
dplyr::group_by(eventID) %>%
dplyr::mutate(eventStartDate = min(startDate),
              eventEndDate = max(endDate)) %>%
dplyr::ungroup() %>%
dplyr::distinct(namedLocation, eventID, .keep_all = TRUE)
```