NEON USER GUIDE TO SMALL MAMMAL BOX TRAPPING (NEON.DP1.10072)

PREPARED BY	ORGANIZATION	DATE
Tanya Chesney	DPS	03/09/2020
Kate Thibault	FSU	03/09/2020

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	07/19/2017	Initial Release
A	03/05/2020	Added description of annual rotation of pathogen and diversity grids at HEAL, TOOL and BARR

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 lifestage of an individual at the time the occurrence was recorded, 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, Small mammal box trapping - individual and trap-level data collected using box traps designed to capture small mammals - 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 Small mammal box trapping (NEON.DP1.10072.001) (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 TOS Protocol and Procedure: Small Mammal Sampling (AD[06]). The raw data that are processed in this document are detailed in the file, NEON Raw Data Ingest Workbook for TOS Small Mammal Abundance and Diversity (NEON.DP0.10001) (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., '10033') 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 Level 1, Level 2 and Level 3 Data Products Catalog
AD[03]	NEON.DP0.10001.001 _dataValidation.csv	NEON Raw Data Ingest Workbook for TOS Small Mammal Abundance and Diversity (NEON.DP0.10001)
AD[04]	NEON.DP1.10072.001 _variables.csv	NEON Data Publication Workbook for Small mammal box trapping (NEON.DP1.10072.001)
AD[05]	NEON.DOC.000915	TOS Science Design for Small Mammal Abundance and Diversity
AD[06]	NEON.DOC.000481	TOS Protocol and Procedure: Small Mammal Sampling
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]	OS_Generic_Transitions .pdf	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[11]		NEON's Ingest Conversion Language (NICL) specifications

3 DATA PRODUCT DESCRIPTION

The small mammal box trapping data product (NEON.DP1.10072) is the mark-recapture, demographic, and size data collected by box trapping for small mammals (i.e., rodents (Rodentia) < 600 grams). Here small mammals are defined based on a combination of taxonomic, behavioral, dietary, and size constraints. This definition includes any rodent that is (1) nonvolant; (2) nocturnally active; (3) forages predominantly aboveground; and (4) is greater than 5 grams, but less than approximately 500-600 g. In North America, this includes cricetids, heteromyids, small sciurids, and introduced murids. It does not include shrews, large squirrels, rabbits, or weasels, despite the fact that individuals of these species may be incidentally captured. Small mammals are widespread, sensitive to local environmental changes, and known to carry and transmit zoonotic agents. Species-specific demography and population sizes, prevalence of pathogens important to public health, species richness, and relative abundances can be monitored simultaneously and ultimately linked to land use and climate changes, and therefore provide useful metrics of responses in biodiversity to these and other drivers (Kao et al. 2012). Moreover, small mammals as primary and secondary consumers interact significantly with plants and ground invertebrates (other NEON sentinel taxa), and generally represent size classes, life histories, and home range sizes that are distinct from the other taxa sampled by NEON (Kao et al. 2012).

Small mammals are sampled at regular intervals by NEON field technicians at core and relocatable sites using standard mark-recapture techniques. Mammals are trapped using box traps (models LFA, XLK, H.B. Sherman Traps, Inc., Tallahassee, FL, USA) and, at sites in Puerto Rico, larger wire traps suitable for catching *Rattus spp.* (model 201, Tomahawk Live Trap, Hazlehurst, WI, USA). Box traps are arrayed in 10 x 10 grids at most sites, with one trap per station and 10 m spacing between traps. Due to extremely high capture rates, grids at Santa Rita Experimental Range (D14, Arizona) are 7 x 7 arrays. Where used, wire traps are used only in alternate bouts of trapping and placed at every other trap station in the 10 x 10 grid, such that a total of 50 wire traps are set.

Products resulting from this sampling include the species identification and unique identifier for each individual captured, as well as a suite of standard size measurements and reproductive condition data. Sample identifiers for any blood, ear, hair, whisker, fecal, and/or voucher samples collected are also provided. For additional details, see AD[05] and AD[06].

3.1 Spatial Sampling Design

Small mammal box trapping is executed at all NEON terrestrial sites, except for Hawaii. Box traps are arrayed in three to eight (depending on the size of the site) 10 x 10 grids, and are collocated with Distributed Base Plots (at which plant, ground beetle and soil sampling may occur), where possible. Specifically, collocated grids are placed 150m +/- 50m in a random direction from the center of the Base Plot. Each grid consists of 100 traps, with 10m spacing between traps (Figure 1). Grids are distributed within the sites proportional to the permitted sampling area contained within each National Land Cover Database (NLCD) class, with the restriction that >50% of the grid must fall within the target NLCD class. Grids may intersect dirt roads <10m in width but are situated such that all grids are >25m from paved roads and buildings and trapping points do not fall within streams, lakes, or ponds. Mammal grids must be separated by a minimum distance of 135m. Due to the equipment and time required to complete sampling, the approximate center (trap location E5) of these grids is not more than 300m from roads that can be accessed by NEON technicians. When fewer than 6 Distributed Base Plots are within 300m of roads, the Mammal Grid centers are placed at a random azimuth and specified distance (150m +/- 50m) from the next available randomized sample locations that are within 300m of roads. See TOS Science Design for Small Mammal

Abundance and Diversity (AD[05]), TOS Protocol and Procedure: Small Mammal Sampling (AD[06]), and TOS Science Design for Spatial Sampling (AD[07]) for further details.

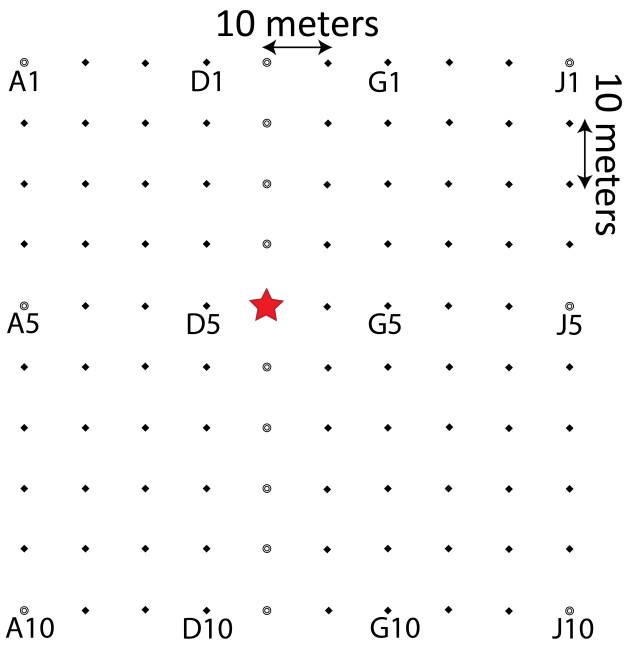

Figure 1: Design of the small mammal trapping grid, consisting of 100 box traps separated by 10 m. Trap coordinates are designated by a unique letter and number combination (e.g., A1, G10). The plot-level coordinates provided in the data product are from trap E5, designated by the red star.

3.2 Temporal Sampling Design

Small mammal sampling occurs in bouts, with a bout comprised of three consecutive (or nearly) nights of trapping on pathogen grids and one night of trapping on the diversity grids. Designation of grids as pathogen or diversity remains consistent through time except at TOOL, BARR, and HEAL, where grid designation (pathogen or diversity) alternates yearly to prevent permafrost damage. Sampling for this product is based on the lunar calendar, with timing of sampling constrained to occur within 10 days before or after the new moon. The number of bouts per year is determined by site type. 6 bouts of sampling per year are conducted at core sites; 4 bouts of sampling per year are conducted at relocatable sites. Sampling bouts occur during the snow-free season and are typically centered around periods of warm temperatures and peak greenness. At sites in the lower latitudes, trapping can occur any time of year. Any expected deviations from this schedule are indicated in the site-specific protocol appendices. See TOS Science Design for Small Mammal Abundance and Diversity (AD[05]) and the TOS Protocol and Procedure: Small Mammal Sampling (AD[06]).

3.3 Variables Reported

All variables reported from the field or laboratory technician (L0 data) are listed in the file, NEON Raw Data Ingest Workbook for TOS Small Mammal Abundance and Diversity (NEON.DP0.10001) (AD[03]). All variables reported in the published data (L1 data) are also provided separately in the file, NEON Data Publication Workbook for Small

mammal box trapping (NEON.DP1.10072.001) (AD[04]).

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.

3.4 Spatial Resolution and Extent

The finest spatial resolution at which small mammal data will be tracked is trap location (i.e., trapCoordinate). One trapping grid (plotID) consists of 100 trap locations (see Figure 1 above); the geographic coordinates for these locations are produced using a Geographic Information System prior to the initiation of sampling.

trapCoordinate (point location of a trap within a plot) → **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 downloaded include the latitude, longitude, and elevation of the approximate centroid of the plot where sampling occurred (i.e., trap E5) + associated uncertainty due to GPS error and plot width. Shapefiles of all NEON Terrestrial Observation System sampling locations can be found in the Document Library: http://data.neonscience.org/documents.

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

- Use the def.calc.geo.os function from the geoNEON package, available here: https://github.com/NEONScience/NEON-geolocation
- Or follow these steps to perform the same calculation:

    1. The namedLocation field in the data is the named location of the plot; more precise geographic data require the named location of the traps. Construct the named location of the trap of each record in mam_pertrapnight by concatenating the fields for namedLocation and trapCoordinate as: namedLocation + '.' + trapCoordinate, e.g. trapCoordinate 'A1' of namedLocation 'HARV_001.mammalGrid.mam' has a complete named location of 'HARV_001.mammalGrid.mam.A1'.
    2. Use the API (http://data.neonscience.org/data-api; e.g. http://data.neonscience.org/api/v0/locations/HARV_001.mammalGrid.mam.A1) 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.
    3. Increase coordinateUncertainty by an appropriate amount to account for error introduced by navigating within traps. NEON field technicians use a combination of high-resolution GPS units, recreational GPS units, and measuring tape to demarcate the sampling grids, with an uncertainty associated with

each trap location of approximately 2 meters. Technicians then place traps within a 1-m radius of these locations during each sampling bout.

3.5 Temporal Resolution and Extent

The finest temporal resolution at which small mammal data will be tracked is the collectDate - which is the date when traps are collected (traps are always set the evening prior).

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

bloodSampleID is a linking variable that ties specific samples and associated metadata to the Rodent-borne pathogen status data product (NEON.DP1.10064).

earSampleID is a linking variable that ties specific samples and associated metadata to the Small mammal sequences DNA barcode data product (NEON.DP1.10076).

3.7 Product Instances

There are a maximum of 6 sampling bouts per year and 8 trapping grids at each core site, while, at relocatable sites, there are a maximum of 4 sampling bouts per year and 6 trapping grids. Each night of trapping (3 per bout for the 3 pathogen grids, 1 per bout for diversity grids) typically includes 100 traps set per night. A record is provided in the pertrapnight table for each trap set. Therefore, a maximum of 8,400 records is expected for each year at any given core site and 4,800 records per year at relocatable sites.

3.8 Data Relationships

The protocol dictates that each grid (plotID) is sampled at most once per collectDate (local time, one record expected in mam_perplotnight per plotID per collectDate). A record from mam_perplotnight may have 100+ child records in mam_pertrapnight (1 per trap expected; >100 expected where a single trap contains multiple captures or there is uncertainty in a trap coordinate associated with a capture). mam_perplotnight and mam_pertrapnight tables may be joined on nightuid, an auto-generated identifier for each unique combination of plotID and collectDate. A record from mam_pertrapnight may have zero (if no blood sampled and/or blood not sent for pathogen analyses) or one child record(s) in the Rodent-borne pathogen status data product (DP1.10064) rpt_bloodtesting table. mam_pertrapnight and rpt_bloodtesting may be joined on bloodSampleID. Similarly, a record from mam_pertrapnight may have zero (if no ear punch collected and/or ear punch not sent for barcoding analyses) or one child record(s) in the Small mammal sequences DNA barcode data product (NEON.DP1.10076) mam_BOLDvoucherInfo table. Duplicates and/or missing data may exist where protocol and/or data entry abberations have occurred; users should check data carefully for anomalies before joining tables.

mam_perplotnight.csv - > One record expected per plotID per collectDate (day of year, local time, generates a single nightUID)

mam_pertrapnight.csv - > One (or more in the case of multiple captures at a single trap) records expected per trapCoordinate per plotID per collectDate (or per nightUID). Generates zero or one (depending on whether pathogens are being sampled.) bloodSampleIDs used in Rodent-borne pathogen status data product (DP1.10064).

bloodSampleIDs, **fecalSampleIDs**, **earSampleIDs** and **hairSampleIDs** will be generated according to resource availability and sampling schedules. Up to 20 blood samples are collected per plotID per collectDate from cricetids, murids, and dipodids. From 2017 onward, only a subset of collected blood samples will be analyzed for pathogens (up to 140 per site per year). Up to 10 fecal and ear samples are collected per species (taxonID) per plotID per collectDate. Hair and whisker samples are only collected from individuals belonging to the dominant genus at each site. While fecal and hair/whisker samples may be collected multiple times from a particular individual, if recaptured, ear punches are only collected once per marked individual. A subset of ear punches, up to 95 per domain per year, are consumed for DNA barcoding analyses. Where not consumed in analyses, these samples are available by request.

voucherSampleIDs will be generated for incidental mortalities that result from trapping. No intentional vouchering is happening on the mark-capture, long-term sampling grids. When they result from trapping on the long-term sampling grids, these vouchers are recorded in the mam_pertrapnight table.

mam_voucher.csv (expanded package only) - > This table includes data for vouchers that are found outside of the trapping effort or are generated through trapping exercises on training grids.

4 DATA QUALITY

4.1 Data Entry Constraint and Validation

Many quality control measures are implemented at the point of data entry within a mobile data entry application or web user interface (UI). For example, data formats are constrained and data values controlled through the provision of dropdown options, which reduces the number of processing steps necessary to prepare the raw data for publication. An additional set of constraints are implemented during the process of ingest into the NEON database. The product-specific data constraint and validation requirements built into data entry applications and database ingest are described in the document NEON Raw Data Ingest Workbook for TOS Small Mammal Abundance and Diversity (NEON.DP0.10001), 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[11]).

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

A schematic of the data entry application design is depicted in Figure 2.

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]).

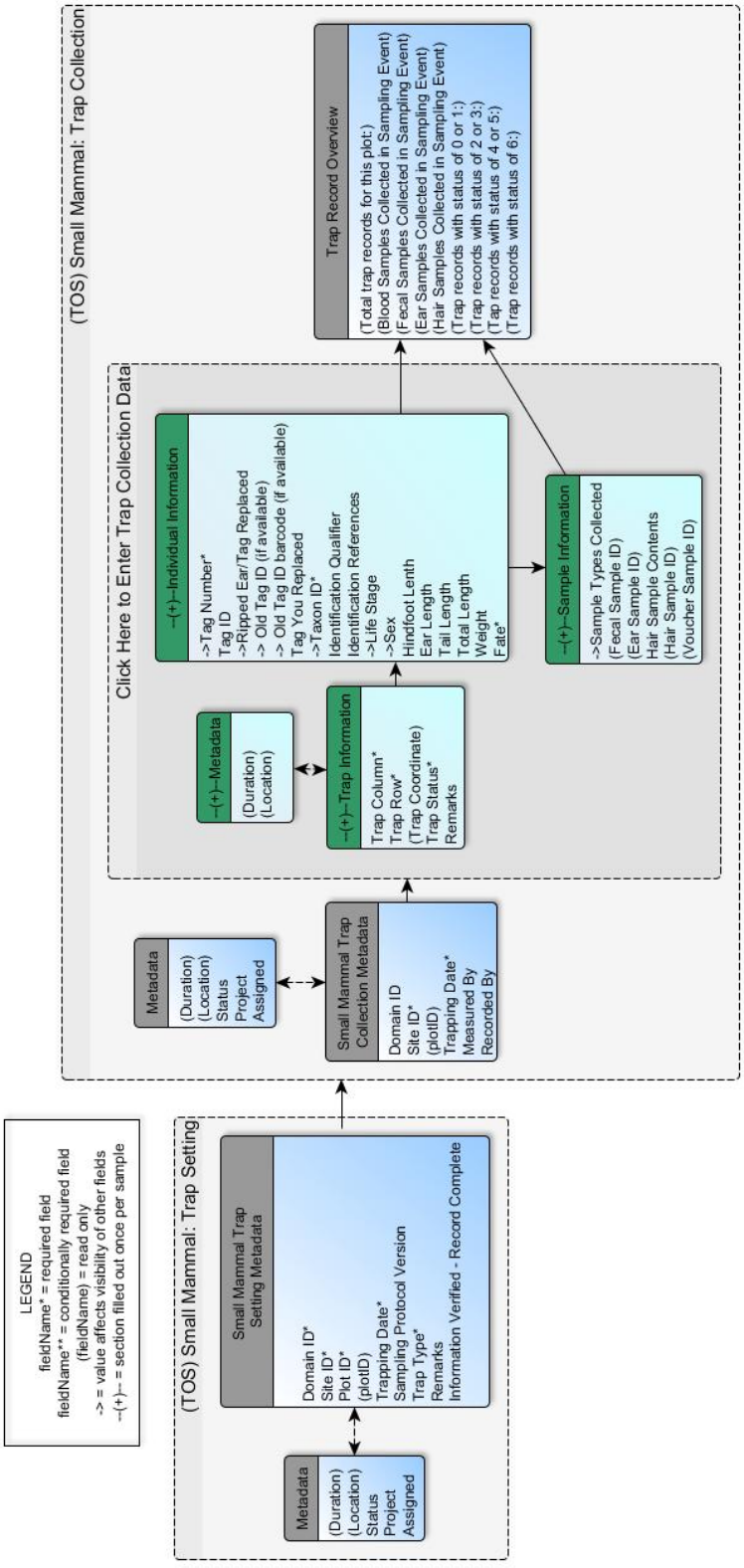

Figure 2: Schematic of the applications used by field technicians to enter small mammal field data

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 data record is a quality flag for known errors applying to the record. There are currently no dataQF codes in use in this data product.

5 REFERENCES

Kao, R. H., C. M. Gibson, R. E. Gallery, C. L. Meier, D. T. Barnett, K. M. Docherty, K. K. Blevins, P. D. Travers, E. Azuaje, Y. P. Springer, K. M. Thibault, V. J. McKenzie, M. Keller, L. F. Alves, E. S. Hinckley, J. Parnell, and D. S. Schimel. 2012. NEON terrestrial field observations: designing continental-scale, standardized sampling. Ecosphere 3:115.