# NEON USER GUIDE TO SMALL MAMMAL BOX TRAPPING (DP1.10072.001)

| PREPARED BY | ORGANIZATION |
| --- | --- |
| Tanya Chesney | SCI |
| Kate Thibault | SCI |
| Sara Paull | SCI |

# 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 |
| B | 06/22/2020 | Included general statement about usage of neonUtilities R package and statement about possible location changes. Updated taxonomy information. |
| C | 11/01/2020 | Added Sampling Design Changes section and included information about suspended sampling in Puerto Rico |
| D | 04/11/2022 | 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 |
| E | 02/20/2023 | Added information about identificationHistory tables and new fields for additionalParasites and samplePrepMethod |
| E.1 | 08/28/2023 | Updated sampling frequency and number of plots sampled following analyses of published data to determine sufficient sampling intensity |

# 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 (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 (DP0.10001.001) (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., '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 Data Products Catalog |
| AD[03] | Available with data download | Validation csv |
| AD[04] | Available with data download | Variables csv |
| 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] | 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 |
| AD[14] | NEON.DOC.005346 | OS Standard Operating Procedure: FRZ – Preparation and Use of Dry Ice Alternative Freezing Materials |

# 3 DATA PRODUCT DESCRIPTION

The small mammal box trapping data product (DP1.10072.001) 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 gradient sites using standard mark-recapture techniques. Mammals are trapped using box traps (models LFA, XLK, H.B. Sherman Traps, Inc., Tallahassee, FL, 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.

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 and Puerto Rico. Box traps are arrayed in three to six (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.

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/

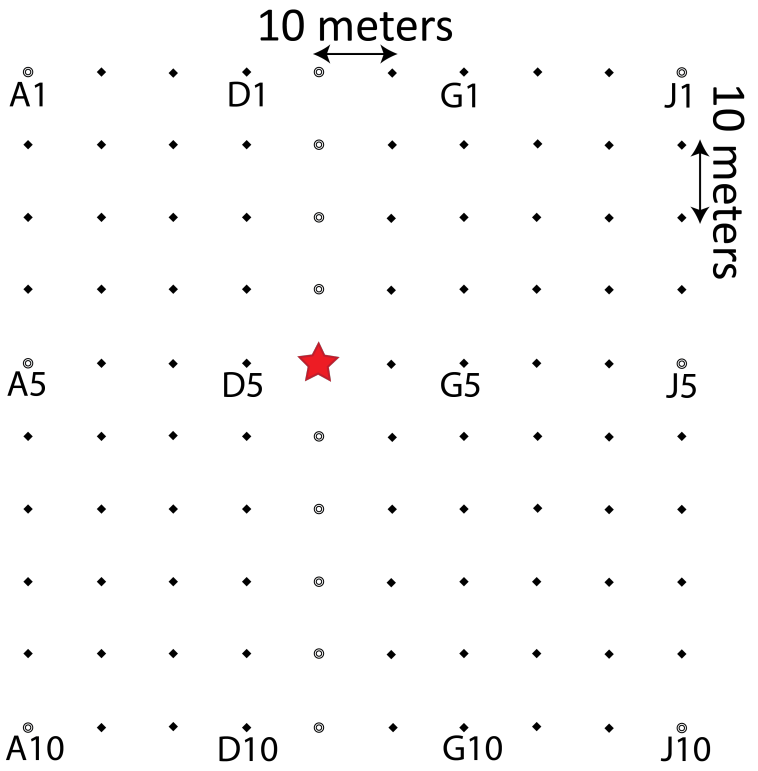


Figure 1: Design of the small mammal trapping grid, consisting of 100 box traps separated by 10 m. Trap coordinates are designated by a unique letter and number combination (e.g., A1, G10). The plot-level coordinates provided in the data product are from trap E5, designated by the red star.

### 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 capture success. The majority of sites are sampled during 4 bouts each year; however a few sites are sampled at a low intensity frequency of one bout per year. Low intensity sites are those sites that meet the following three criteria for two or

more years in a row: i) < 5% capture success per year, ii) > 80% of captures from 3 or fewer species, and iii) < 5% samples test positive for any pathogen each year. 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 Sampling Design Changes

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

In the 2015 field season, sample bout duration was decreased from three nights to one at diversity grids. By maintaining three sampling nights on the remaining three pathogen grids at each site, this design change balanced the sampling of a diversity of land cover classes with the recapture data required for density estimation. Beginning with this field season, only ten samples of a given type (e.g., hair, fecal, ear) are collected from priority species when capture rates are high to reduce the time animals spend in traps.

In the spring of 2016 small mammal trapping in Puerto Rico was suspended to prevent unintended capture of invasive mongoose (*Herpestes auropunctatus*). Data on small mammal communities in Puerto Rico are of low scientific value given that they are comprised exclusively of invasive species of Rattus and Mus. The larger Tomahawk traps required to capture the Rattus spp. risk capturing invasive mongoose (*Herpestes auropunctatus*). The site host would not permit release of this invasive species once captured, and the IACUC would not permit euthanasia by NEON staff due to risks to staff safety.

In the 2016 field season the presence/absence of ticks was recorded by tick life stage on small mammals. Beginning in the 2020 field season, technicians also began recording the binned number of total ticks attached to a small mammal, with groupings that include: 1-5, 6-20, >20.

In the 2017 field season the size of the small mammal trapping grid at SRER was reduced from a 10 x 10 grid of 100 traps to a 7 x 7 grid of 49 traps. This change was necessary to enable completion of the grids in a timely manner due to consistently very high capture rates.

In the 2019 field season, the weight of hair samples being collected was increased from < 1 mg to a target size of 5-8 mg. This was to accommodate the sample size required for stable isotope and other analyses.

Beginning in the 2020 field season, the field **samplingImpractical** was added to the data to allow for the generation of a record when a plot could not be sampled for a particular night of sampling. If field sampling was not possible **samplingImpractical** is populated with a value other than 'OK' (e.g., 'location flooded'). If sampling could not be conducted for all or part of the bout, the samplingImpractical field will communicate such missing records and the reason therefore.

Starting in 2023, the field **additionalParasites** was added to the data to allow for formalized tracking of the presence of non-tick ectoparasites such as botflies and fleas. Due to dry ice shortages, we also developed alternative initial storage methods for frozen samples and tracked their use with the new field

entitled **samplePrepMethod**. Additional details on these alternative freezing methods can be found in AD[14].

In 2023, analyses of the published NEON small mammal dataset were completed and it was determined that reductions in the sampling intensity could be realized with minimal impacts to data quality. Prior to 2023, sampling occurred 6 times per year at core sites and 4 times per year at gradient sites (except in locations where the sampling season is shorter due to weather or snow). After 2023, the number of bouts was capped at 4 bouts per year at all sites. Similarly, the number of grids sampled per site was capped at 6, which resulted in a reduction to the total number of grids sampled at 9 sites (CLBJ, CPER, ORNL, SERC, SJER, TALL, TOOL, BONA and UNDE). At these sites, 1-2 diversity grids were dropped from sampling according to a randomly generated plot priority matrix so as not to bias estimates generated for the site. Additionally, a low intensity sampling option was introduced for sites with consistently low capture rates, diversity, and pathogen presence such that only one bout per year would be completed at those sites.

## 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 Small Mammal Abundance and Diversity (DP0.10001.001) (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 (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.

## 3.5 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 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 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.

### 3.6 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

### 3.7 Associated Data Streams

**bloodSampleID** is a linking variable that ties specific samples and associated metadata to the Rodent pathogen status, hantavirus data product (DP1.10064.001) and to the Rodent pathogen status, tick-borne data product (DP1.10064.002).

**earSampleID** is a linking variable that ties specific samples and associated metadata to the Rodent pathogen status, tick-borne data product (DP1.10064.002), and the Small mammal sequences DNA barcode data product (DP1.10076.001).

### 3.8 Product Instances

There are a maximum of 6 sampling bouts per year and 8 trapping grids at each core site, while, at gradient 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 gradient sites.

### 3.9 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). The 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 or ear samples and/or tissue not sent for pathogen analyses) or one child record(s) in the Rodent pathogen status, hantavirus data product (DP1.10064.001) rpt_bloodtesting table. There may also be from 0-18 child records in the Rodent pathogen status, tick-borne (DP1.10064.002) data product rpt2_pathogentesting table for each record from mam_pertrapnight with a blood or ear sampleID. The mam_pertrapnight and rpt_bloodtesting tables may be joined on bloodSampleID, while mam_pertrapnight and rpt2_pathogentesting may be joined by either bloodSampleID or earSampleID, depending on sample type. 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 (DP1.10076.001) 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 and/or earSampleIDs used in Rodent-borne pathogen status data products (DP1.10064.001 and DP1.10064.002).

**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.

mam_identificationHistory.csv (expanded package only) - > One or more records expected per identificationHistoryID. Records are only created when data corrections to taxonomy are made. If errors in identification are detected through QAQC processes after data publication, then corrected taxonomy will be provided in the mam_pertrapnight table. The mam_identificationHistory table is populated with all prior names used for specimen(s) in the data product. When data are populated in the mam_identificationHistory table, identificationHistoryID is used as a linking variable between the mam_identificationHistory table and the mam_pertrapnight table where updates were made.

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

# 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 mammals is Wilson and Reeder (2005). Taxon ID codes used to identify taxonomic concepts in the NEON master taxonomy list are 4-8 character alpha-numeric codes, derived from the accepted scientific name. Each code for a single defined species is composed of the first two letters of the genus, followed by the first two letters of the specific epithet. A number is added to the end when necessary to distinguish between duplicate codes. For species that are part of a cryptic pair that are difficult to differentiate in the field, an 8-character ID is used that consists of the 4 character taxon ID code for each species in the pair. The list includes all small mammal species from the continental United States and Alaska. NEON plans to keep the taxonomy updated in accordance with the current literature, starting in 2020 and annually thereafter.

The master taxonomy list also indicates the expected geographic distribution for each species by NEON domain and whether it is known to be 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. Geographic ranges and native statuses used in this data product are from Nature Serve (https://www.natureserve.org/conservation-tools/data-maps-tools/digital-distribution-maps-mammals-western-hemisphere).

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.

Beginning in 2022 the mam_identificationHistory table was added to track any changes to taxonomic identifications that have been published in NEON data. Such taxonomic revisions may be necessary when errors are found in QAQC checks, or when evidence from genetic analysis of samples or re-analysis of archived samples indicate a revision is necessary. Requests for taxonomic changes are reviewed by NEON science staff. Proposed changes are evaluated based on evidence in the form of photographs, existing samples, genetic data, consultation with taxonomic experts, or range maps. Upon approval, the existing record in the mam_pertrapnight table is updated with the new taxonomic information and a unique identifier is added to the **identificationHistoryID** field. A record with the same identificationHistoryID is created in the mam_identificationHistory table where the previous taxonomic information are archived along with the date the change was made.

# 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 Small Mammal Abundance and Diversity (DP0.10001.001), 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[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.

## 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

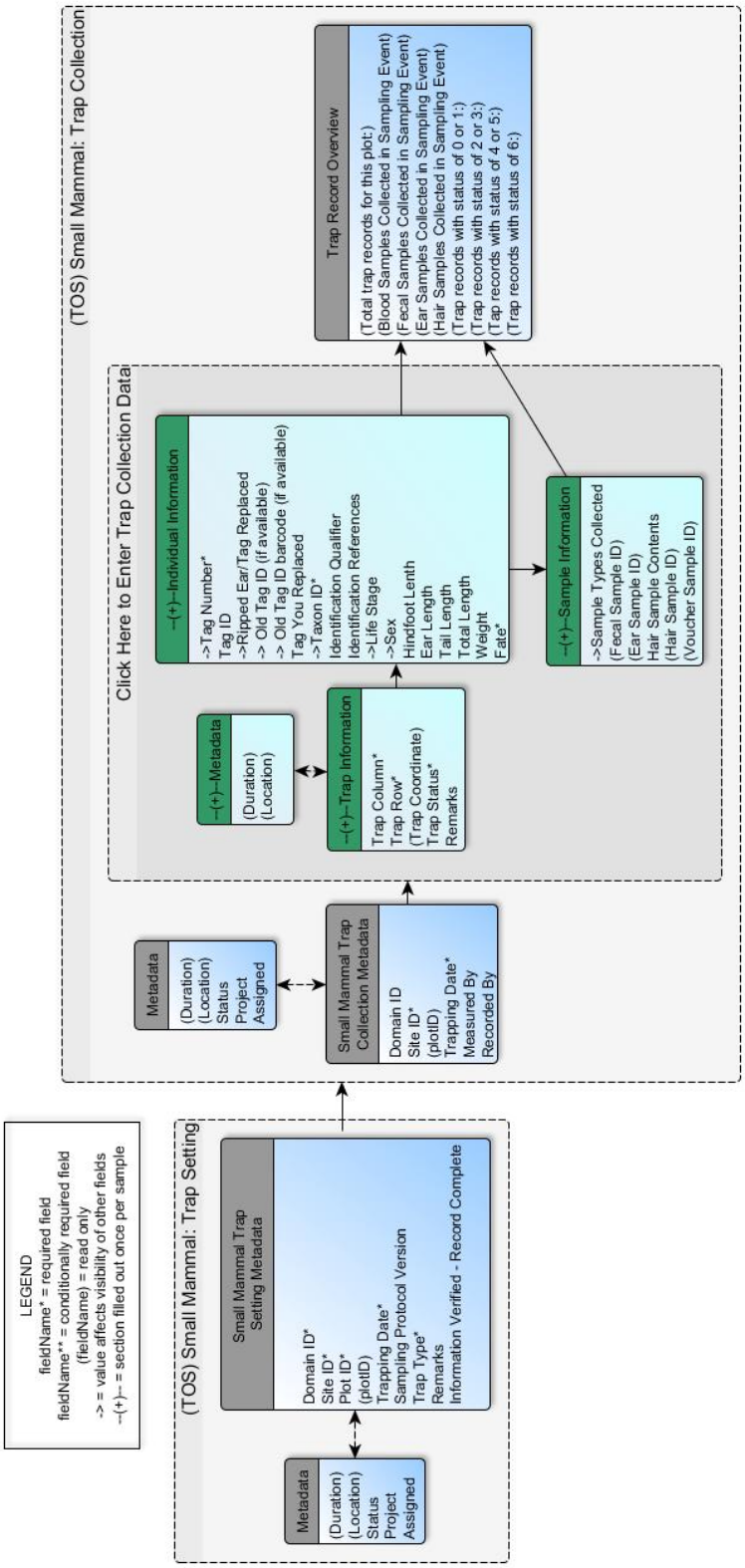


Figure 2: Schematic of the applications used by field technicians to enter small mammal field data

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 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.

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 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.

Wilson, D. E., and D. M. Reeder. 2005. Mammal Species of the World. A Taxonomic and Geographic Reference. 3rd ed. 2,142pp. Johns Hopkins University Press.