



NEON USER GUIDE TO RODENT PATHOGEN STATUS, HANTAVIRUS (DP1.10064.001)

PREPARED BY	ORGANIZATION
Kate Thibault	FSU



CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	09/12/2017	Initial Release
B	07/06/2020	Included general statement about usage of neonUtilities R package and statement about possible location changes.
C	11/01/2020	Included information about the change from Hantavirus to tick-borne disease testing for 2020 field season.
D	04/13/2022	Updated data product title from Rodent-borne pathogen status to Rodent pathogen status, hantavirus. Updated section 4.3 Data Revision with latest information regarding data release. Updated information regarding the geoNEON package.
E	02/18/2025	Added information about colocation with Tick Plots in section 3.1. Updated the url for spatial data in section 3.5. Added information about the new neonUtilities Python package.



TABLE OF CONTENTS

1	DESCRIPTION	1
1.1	Purpose	1
1.2	Scope	1
2	RELATED DOCUMENTS	2
2.1	Associated Documents	2
3	DATA PRODUCT DESCRIPTION	3
3.1	Spatial Sampling Design	3
3.2	Temporal Sampling Design	5
3.3	Sampling Design Changes	5
3.4	Variables Reported	6
3.5	Spatial Resolution and Extent	7
3.6	Temporal Resolution and Extent	7
3.7	Associated Data Streams	8
3.8	Product Instances	8
3.9	Data Relationships	8
4	DATA QUALITY	9
4.1	Data Entry Constraint and Validation	9
4.2	Automated Data Processing Steps	9
4.3	Data Revision	9
4.4	Quality Flagging	9
4.5	Analytical Facility Data Quality	10
5	REFERENCES	10

LIST OF TABLES AND FIGURES

Table 1	Descriptions of the dataQF codes for quality flagging	9
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.	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 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, Rodent Pathogen Status, hantavirus - the presence/absence of a pathogen (or antibodies to a pathogen) in a single rodent sample - 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 Rodent Pathogen Status, hantavirus (DP1.10064.001) (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 Rodent Pathogen Status, hantavirus (DP0.10064.001) (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	NEON Raw Data Ingest Workbook for Rodent Pathogen Status, hantavirus (DP0.10064.001)
AD[04]	Available with data download	NEON Data Publication Workbook for Rodent Pathogen Status, hantavirus (DP1.10064.001)
AD[05]	NEON.DOC.000911	TOS Science Design for Vectors and Pathogens
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



3 DATA PRODUCT DESCRIPTION

The Rodent Pathogen Status, hantavirus data product (DP1.10064.001) includes the results of the testing of blood samples from individual small mammals for seropositivity to pathogens. Small mammal blood samples were tested for hantaviruses from 2014-2019. Beginning in 2020 rodent pathogen testing focuses on tick-borne diseases. The blood samples are collected as part of the mark-recapture, box trapping effort for small mammals (i.e., rodents (Rodentia) < 600 grams), with the field capture results available separately via the Small mammal box trapping data product (DP1.10072.001). Blood samples are collected from individuals of target species of rodents, including all species in the families Cricetidae, Muridae, and Dipodidae, if an individual weighs more than 10 grams and is in good physical condition. Samples are collected only from individuals captured on three of the mark-recapture mammal grids at each NEON site, designated as pathogen grids and trapped for three consecutive (or nearly) nights per sampling bout. Samples are only collected once per individual per sampling bout, but an individual may be resampled in subsequent bouts. When capture rates are high (>20%), blood samples are only collected from up to 20 individuals per plot per trapping date.

Sampling frequency therefore depends on the frequency of captures and successful blood collection from target species. Blood samples are collected via the retro-orbital (Mills et al. 1995) or submandibular (Golde et al. 2005) method. Samples are placed immediately after collection on dry ice in a cooler and then stored in a -80 °C freezer at the completion of the field day. One to two shipments of blood samples are made to the external laboratory for pathogen analysis throughout the field season. A minimum of 20 microliters is required for testing; samples received by the external laboratory that are less than this volume are indicated as Quantity Not Sufficient (QNS) in the data. In some years, all blood samples have been tested. In most years, only a subset of samples collected will be tested, up to approximately 140 samples per NEON site. Testing is primarily targeted at adults. For additional details, see AD[05] and AD[06]. The testing protocol used for a particular sample is provided in the testProtocolVersion field of the rpt_bloodtesting table.

3.1 Spatial Sampling Design

Rodent Pathogen Status, hantavirus data rely on the small mammal box trapping that is executed at all NEON terrestrial sites, except for Hawaii, Puerto Rico, and Yellowstone National Park where not permitted. 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. Each grid consists of 100 traps, with 10m spacing between traps (Figure 1). Rodent Pathogen Status, hantavirus data are typically collected only on three of the mammal trapping grids at each site, with these grids designated as 'pathogen grids'. At a handful of sites where capture rates are extremely low (<5%), blood samples may be collected from all grids to achieve a higher sample size for pathogen testing. Pathogen grids are trapped for three consecutive (or nearly) nights per sampling event to facilitate the generation of robust density estimates, whereas the remaining grids, 'diversity grids' are only sampled for one night per sampling event. Pathogen grids are selected by the field crews to achieve the goals of 1) long-term sampling of target and opportunistic species, which requires moderate to high average capture rates, and 2) representation of the site's dominant vegetation type (National Land Cover Database (NLCD) class). If there are more or fewer than 3 trapping grids that occur within the dominant vegetation type(s) and trapping data from previous years are available, field crews target grids with the



highest combined abundances of target and opportunistic species. If the majority of captures in the dominant vegetation type at a site are heteromyids (not targeted for pathogen analysis), then an alternate habitat type may be targeted. Otherwise, pathogen grids are chosen at random or based on an educated assessment of habitat quality. Once a grid has been designated as a pathogen grid (after an initial assessment period of two years), that classification applies for all subsequent trapping seasons for consistent, long-term data collection.

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

To facilitate analyses connecting the tick and small mammal as well as associated pathogen data products, when possible, mammal grids are located within 500m of a co-located tick sampling plot. Co-located plots have the same plotID and occur in the same NLCD cover. For example: ABBY_002.mammalGrid.mam and ABBY_002.tickPlot.tck.

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/>

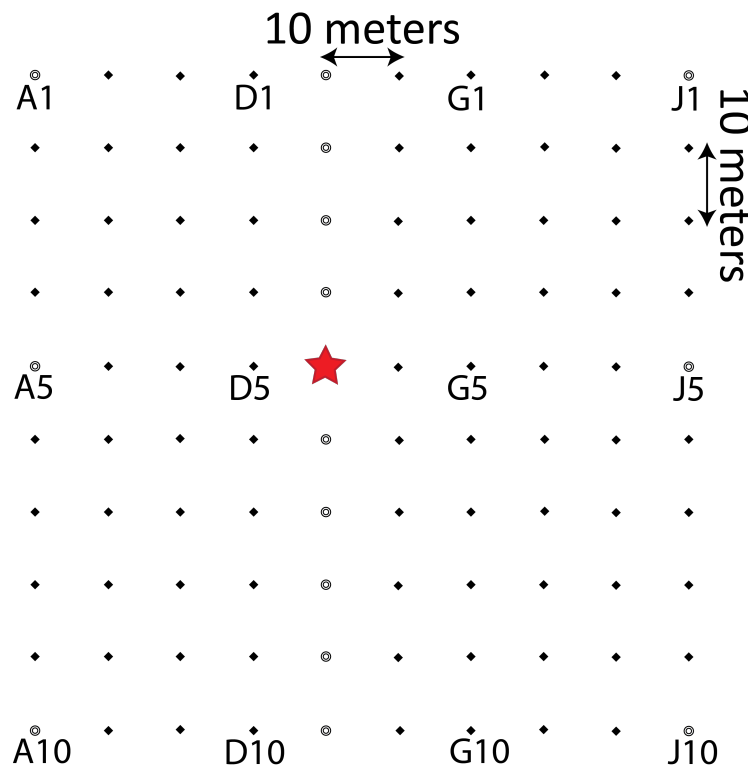


Figure 1: Design of the small mammal trapping grid, consisting of 100 box traps separated by 10 m. Trap coordinates are designated by a unique letter and number combination (e.g., A1, G10). The plot-level coordinates provided in the data product are from trap E5, designated by the red star.

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. 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. Blood sampling occurs in all small mammal trapping bouts. See TOS Science Design for Vectors and Pathogens (AD[05]) and the TOS Protocol and Procedure: Small Mammal Sampling (AD[06]).

3.3 Sampling Design Changes

There have been some design changes that have been implemented to the small mammal data products 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. A more complete list of design changes for small mammal data collection, including those less likely to influence interpretation of rodent-borne pathogen data, is reported in the NEON User Guide to Small Mammal Box Trapping.

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. Blood samples are collected from all pathogen grids. 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 auro punctatus*). 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 auro punctatus*). 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.

Beginning in the 2020 field season, small mammal blood samples will no longer be tested for hantaviruses. In 2020 a pilot study will be conducted such that blood samples from 5 sites will be tested for a range of tick-borne diseases, and ear samples will be tested for *Borrelia burgdorferi*. In the 2021 field season, the rodent-borne pathogen testing will be conducted at all NEON sites and will involve testing for tick-borne diseases. This change will improve linkages between the tick and small mammal datasets. Blood samples collected in excess of our capacity to test will continue to be archived at the Biorepository and can be requested by external users for use in hantavirus studies.

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 Rodent Pathogen Status, hantavirus (DP0.10064.001) (AD[03]). All variables reported in the published data (L1 data) are also provided separately in the file, NEON Data Publication Workbook for Rodent Pathogen Status, hantavirus (DP1.10064.001) (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/ncceas/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 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.



3.5 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 on the NEON science webpage at <https://www.neonscience.org/data-samples/data/spatial-data-maps>.

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 Small mammal box trapping data product (DP1.10072.001).

3.8 Product Instances

Only blood samples tested for pathogens are included in this data product. Typically, a maximum of 140 is expected per site per year, although in some years all samples have been tested.

3.9 Data Relationships

There should be one record per **bloodSampleID** in the `rpt_bloodtesting` table. This record can be joined on **bloodSampleID** with the corresponding record in the `mam_pertrapnight` table of the Small mammal box trapping data product (DP1.10072). 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.

`rpt_bloodtesting.csv` -> One record expected per **bloodSampleID**. The basic package includes the date and **plotID** on which the sample was collected along with the test result. The expanded package includes multiple data fields describing the technical specifications of the batch-level analytical run, such as the positive control antibody lot number and the diluent dilution. **bloodSampleIDs** are also used in Small mammal box trapping data product (DP1.10072).

`mam_pertrapnight.csv` -> From the Small mammal box trapping data product (DP1.10072). 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) records in `rpt_bloodtesting.csv`

bloodSampleIDs 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 are analyzed for pathogens (up to 140 per site per year). Where not consumed in analyses, these samples are available by request.

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 DATA QUALITY

4.1 Data Entry Constraint and Validation

Constraints and data validation 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 Rodent Pathogen Status, hantavirus (DP0.10064.001), 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]]).

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[10]).

4.3 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.

4.4 Quality Flagging

The **dataQF** field in each data record is a quality flag for known errors applying to the record. Please see below for an explanation of **dataQF** codes specific to this product.

Table 1: 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)



4.5 Analytical Facility Data Quality

Samples are tested against negative control antigen at the same time they are tested for the presence of antigen specific antibodies in the serum. As internal positive controls, serum known to contain antibodies against the selected antigens is used. As the negative controls, serum known to contain no antibodies against the selected antigens is used. Positive and negative control sera are used at minimum sensitivity to ensure the reliability of the assay. More details can be found in the SOP.

5 REFERENCES

- Botten, J., K. Mirowsky, D. Kusewitt, M. Bharadwaj, J. Yee, R. Ricci, R.M. Feddersen, and B.Hjelle. 2000. Experimental infection model for Sin Nombre hantavirus in the deer mouse (*Peromyscus maniculatus*). Proceedings of the National Academy of Sciences 97: 10578-10583.
- Golde, W., P. Gollobin, and L. Rodriguez. 2005. A rapid, simple, and humane method for submandibular bleeding of mice using a lancet. Lab Animal, 34:39–43.
- Hjelle, B., S. Jenison, N. Torrez-Martinez, B. Herring, S. Quan, A. Polito, S. Pichuanes et al. 1997. Rapid and specific detection of Sin Nombre virus antibodies in patients with hantavirus pulmonary syndrome by a strip immunoblot assay suitable for field diagnosis. Journal of Clinical Microbiology 35: 600-608.
- Mills, J.N., J.E. Childs, T.G. Ksiazek, C.J. Peters, and W.M. Velleca. 1995. Methods for trapping and sampling small mammals for virologic testing. Centers for Disease Control and Prevention.