



<i>Title:</i> NEON User Guide to Ground beetles sampled from pitfall traps (DP1.10022.001)	<i>Date:</i> 10/05/2020
<i>Author:</i> Katherine LeVan	<i>Revision:</i> C

## NEON USER GUIDE TO GROUND BEETLES SAMPLED FROM PITFALL TRAPS (DP1.10022.001)

<b>PREPARED BY</b>	<b>ORGANIZATION</b>
Katherine LeVan	SCI

## CHANGE RECORD

<b>REVISION</b>	<b>DATE</b>	<b>DESCRIPTION OF CHANGE</b>
A	07/19/2017	Initial Release
B	07/15/2020	Included general statement about usage of neonUtilities R package and statement about possible location changes. Updated taxonomy information.
C	09/01/2020	Clarified that from 2018-present, three traps were deployed at each plot (30 traps/site/bout); Adds missed bout reporting, adds new field called sampling impractical



## TABLE OF CONTENTS

<b>1</b>	<b>DESCRIPTION</b>	<b>1</b>
1.1	Purpose . . . . .	1
1.2	Scope . . . . .	1
<b>2</b>	<b>RELATED DOCUMENTS AND ACRONYMS</b>	<b>2</b>
2.1	Associated Documents . . . . .	2
<b>3</b>	<b>DATA PRODUCT DESCRIPTION</b>	<b>3</b>
3.1	Spatial Sampling Design . . . . .	4
3.2	Temporal Sampling Design . . . . .	5
3.3	Sampling Design Changes . . . . .	6
3.4	Variables Reported . . . . .	6
3.5	Temporal Resolution and Extent . . . . .	7
3.6	Spatial Resolution and Extent . . . . .	7
3.7	Associated Data Streams . . . . .	7
3.8	Product Instances . . . . .	8
3.9	Data Relationships . . . . .	8
3.10	Special Considerations . . . . .	9
<b>4</b>	<b>TAXONOMY</b>	<b>9</b>
4.1	Beetle Taxonomy . . . . .	9
4.2	Mammal Taxonomy . . . . .	10
4.3	Herptiles Taxonomy . . . . .	10
<b>5</b>	<b>DATA QUALITY</b>	<b>11</b>
5.1	Data Entry Constraint and Validation . . . . .	11
5.2	Automated Data Processing Steps . . . . .	11
5.3	Data Revision . . . . .	11
5.4	Quality Flagging . . . . .	11
<b>6</b>	<b>REFERENCES</b>	<b>14</b>

## LIST OF TABLES AND FIGURES

Table 1	Descriptions of the sampling impractical codes for quality flagging . . . . .	13
Figure 1	A workflow illustrating the process of data collection for ground beetles. The red cup and blue cup demonstrate sample processing of 2 pitfall traps from the same plot. . . . .	4
Figure 2	Layout of 4 pitfall traps at a plot. The plot-level centroid is designated by the red star. In 2018, NEON discontinued sampling at the North trap location. . . . .	5
Figure 3	Schematic of the applications used by field technicians to enter ground beetle field data . . . . .	12



## 1 DESCRIPTION

### 1.1 Purpose

This document provides an overview of the data included in this NEON Level 1 data product, the quality controlled product generated from raw Level 0 data, and associated metadata. In the NEON data products framework, the raw data collected in the field, for example, the dry weights of litter functional groups from a single collection event are considered the lowest level (Level 0). Raw data that have been quality checked via the steps detailed herein, as well as simple metrics that emerge from the raw data are considered Level 1 data products.

The text herein provides a discussion of measurement theory and implementation, data product provenance, quality assurance and control methods used, and approximations and/or assumptions made during L1 data creation.

### 1.2 Scope

This document describes the steps needed to generate the L1 data product: Ground beetles sampled from pitfall traps 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 files NEON Data Variables for Ground beetles sampled from pitfall traps (DP1.10022.001) (AD[05]) and NEON Data Variables for Ground beetle sequences DNA barcode (DP1.10020.001) (AD[06]), 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: Ground Beetle Sampling (AD[08]). The raw data that are processed in this document are detailed in the file, NEON Raw Data Validation for Ground beetles sampled from pitfall traps (DP0.10022.001) (AD[04]), provided in the download package for this data product. Please note that raw data products (denoted by 'DP0') may not always have the same numbers (e.g., '10003') as the corresponding L1 data product.



## 2 RELATED DOCUMENTS AND ACRONYMS

### 2.1 Associated Documents

AD[01]	NEON.DOC.000001	NEON Observatory Design (NOD) Requirements
AD[02]	NEON.DOC.000913	TOS Science Design for Spatial Sampling
AD[03]	NEON.DOC.002652	NEON Data Products Catalog
AD[04]	Available with data download	Validation csv
AD[05]	Available with data download	Variables csv - DP1.10022.001
AD[06]	Available with data download	Variables csv - DP1.10020.001
AD[07]	NEON.DOC.000909	TOS Science Design for Ground Beetle Abundance and Diversity
AD[08]	NEON.DOC.014050	TOS Protocol and Procedure: Ground Beetle Sampling
AD[09]	NEON.DOC.000008	NEON Acronym List
AD[10]	NEON.DOC.000243	NEON Glossary of Terms
AD[11]	NEON.DOC.004285	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[12]	Available on NEON data portal	NEON Ingest Conversion Language Function Library
AD[13]	Available on NEON data portal	NEON Ingest Conversion Language
AD[14]	Available with data download	Categorical Codes csv



### 3 DATA PRODUCT DESCRIPTION

The Ground beetles sampled by pitfall traps data product provides counts of ground beetles (Coleoptera: Carabidae) and vertebrate bycatch specimens from individual sampling bouts. Pitfall traps consist of 16 oz deli containers filled with 150 or 250 mL of propylene glycol. Pitfall trapping has been used for more than a century and is still the most commonly used method for beetle collection. This standard, well-established, widely-used, and effective sampling method (Kotze et al., 2011; Kromp 1999; Rainio & Niemela 2003) was selected to maximize comparability across time and between domains within the observatory. Multiple traps are deployed in each of 10 plots at each terrestrial NEON site (from 2013-2017, 40 traps per site), with traps arrayed approximately 20 meters from the center of the plot in each of the four cardinal directions. In 2018, the number of traps deployed at each plot was reduced from 4 traps per plot to 3 traps per plot. Sampling occurs biweekly throughout the growing season (when temperatures are above 4°C).

Following trap collection, samples are then sorted back in the NEON domain laboratory (see step 2 of Figure 1). During this process, vertebrate bycatch specimens (namely, small mammals, reptiles, and amphibians) are identified, counted, and stored separately for archiving. Invertebrate bycatch, i.e., all adult invertebrates that are not identified as belonging to the family Carabidae and all larvae, are stored together for archiving purposes, but no taxonomic identifications or counts are made. Archived specimens (see step 3 of Figure 1) may be maintained at the trap-level or pooled at the plot-level, depending on specimen quantities and archival costs. Finally, carabid beetles are sorted by NEON technicians, identified to species or morphospecies (see step 2 of Figure 1), and a subset of these are pointed or pinned (see step 4 of Figure 1). Following pinning/pointing, a subset of individuals (up to 467 per site) is sent to an expert taxonomist(see step 5 of Figure 1) and/or a DNA barcoding facility for identification (see step 6 of Figure 1). Pinned specimens that technicians identify as a morphospecies (that is, no species-level identification can be determined), are prioritized for secondary identification. Only pinned specimens that have already received an identification by a taxonomist expert will be eligible for DNA barcoding. These secondary and tertiary identifications will allow for the accurate identification of taxa that are difficult to identify, cryptic or poorly-known taxonomically (i.e., new species, genera that have limited or incomplete representation in taxonomic keys, etc.). Regardless of storage method, all collections data are reported at a per trap resolution. Identifications performed on these individuals may be used to estimate uncertainty in parataxonomist identification by NEON technicians.

All beetles must be sorted prior to pinning, so the total number of beetles collected can be calculated as the sum of individualCount in bet\_sorting, though further identifications may be updated based on the downstream workflow (see Special Considerations, below).

For additional details on the sampling design and associated protocol, see the TOS Science Design for Ground Beetle Abundance and Diversity (AD[07]) and TOS Protocol and Procedure: Ground Beetle Sampling (AD[08]).

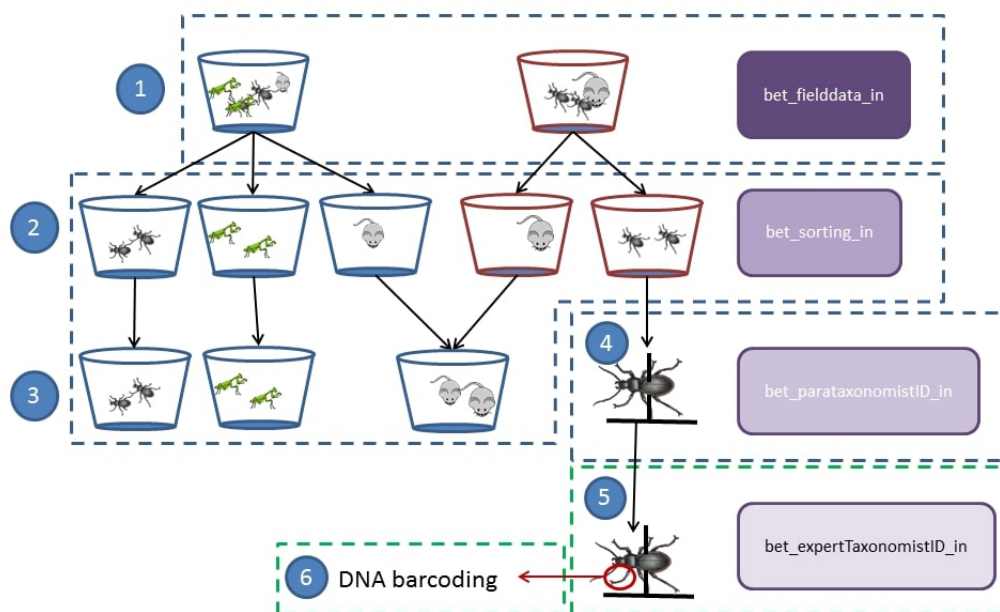


Figure 1: A workflow illustrating the process of data collection for ground beetles. The red cup and blue cup demonstrate sample processing of 2 pitfall traps from the same plot.

### 3.1 Spatial Sampling Design

Beetle sampling is executed at all terrestrial NEON sites and follows a spatially-balanced stratified random design (AD[02]). Beetles are sampled at replicate traps (Figure 2) at 10 distributed plots per site. Plots are randomly positioned within each National Land Cover Database (NLCD) class with representation within each NLCD class set as proportional to its representation at the site; NLCD classes with less than 5% representation are excluded from sampling. Additionally, plots must be separated by a minimum of 55m, plot centers must be > 50m from large paved roads and buildings, plot edges must be >10m from dirt roads and plots may not be intersected by streams >1m wide. Beetle sampling is directly co-located with sampling of soils and plant diversity.

As much as possible, sampling occurs in the same locations over the lifetime of the Observatory. However, over time some sampling plots may become impossible to sample, due to disturbance or other local changes. When this occurs, the location and its location ID are retired. A new location is established, and given a new location ID. Refer to the data product change log for details about plots that have been retired.

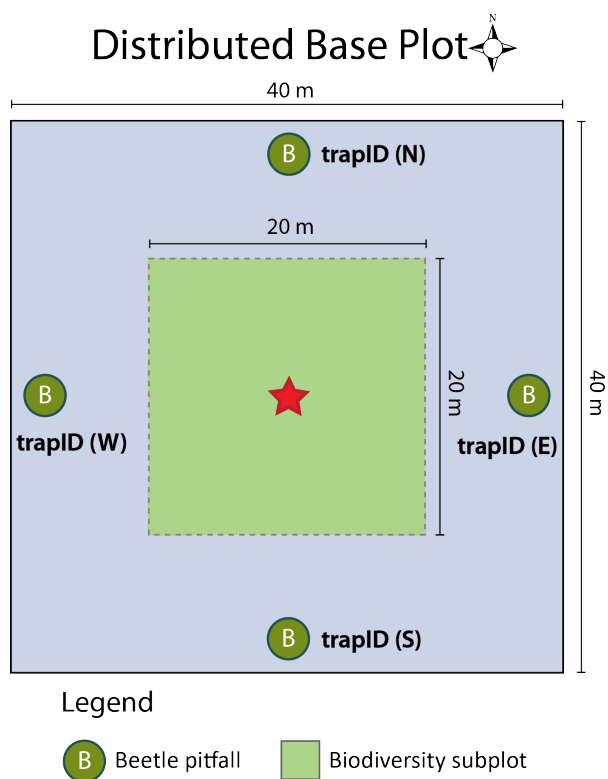


Figure 2: Layout of 4 pitfall traps at a plot. The plot-level centroid is designated by the red star. In 2018, NEON discontinued sampling at the North trap location.

### 3.2 Temporal Sampling Design

The finest temporal resolution at which beetle data (for the purposes of species richness and abundance) will be tracked is trapping bout, a ~14-day interval during which pitfall traps are deployed. The `setDate` (indicating when the trap was set) and `collectDate` (indicating when the trap was collected) are recorded for each sample collected during a bout. Bouts are grouped using the `eventID` designation (a descriptor that includes the year of sampling, the site ID, and the calendar week in which a sampling bout occurred). Infrequently, a bout may be scheduled over 2 dates that span different ISOweeks.

The total number of bouts per year varies among sites based on seasonality of each site (e.g., stopping during winter at temperate sites) up to a maximum of 13 bouts per site. During the time of year when carabids are active, sampling bouts occur every 2 weeks. After the beetle sampling season has ended (e.g., upon the onset of winter), the next season will not resume until minimum ambient temperatures (average minimum temperature over the preceding 10 days exceeds 4°C) and green-up are observed. A given scheduled sampling bout will be cancelled if minimum ambient temperature thresholds are not met. Additional details about sampling bout frequency can be found in the TOS Protocol and Procedure: Ground Beetle Sampling (AD[08]).





Title: NEON User Guide to Ground beetles sampled from pitfall traps (DP1.10022.001)	Date: 10/05/2020
Author: Katherine LeVan	Revision: C

### 3.3 Sampling Design Changes

- 2013-2016: Early protocols had a workflow where carabids were treated one of two ways. NEON pinned no or few individuals from certain, easily identified and common carabids (tagged with **sampleType** = 'common carabid') and the identification for these taxa is included at a fine grain level in the `bet_sorting` table. All carabid specimens from hard to identify taxa (tagged with **sampleType** = 'other carabid') were pinned; for these records, the `bet_sorting` table had coarse taxonomic information (`scientificName` might be 'Carabidae spp.') and all finer scale taxonomy was communicated in the `bet_parataxonomistID` table (as a result, the taxonomic information from `bet_sorting` and `bet_parataxonomistID` may differ). For all records where **sampleType** is 'other carabid' users should preferentially use taxonomic information from the `bet_parataxonomistID` over the `bet_sorting` table, where differences occur. In more recent protocols (**sampleType** is 'carabid') and in older records where **sampleType** is 'common carabid', users should preferentially use information in the `bet_sorting` table for taxonomic identifications from NEON staff.
- 2013-2017: In the original design, each plot had 4 pitfall traps. In 2018, this was modified to have 3 traps deployed per plot location.
- 2020: Before 2020, each bout of collection would have *up to* 40 records (collection years 2013-2017) or 30 records (collection years 2018-2019) in the `bet_fielddata` table. In cases where fewer than the maximum number of records were generated per bout, this reflects lower levels of sampling effort and additional records should not be expected. In 2020, NEON added the quality flag field **samplingImpractical** to this data product to assist users in understanding when data for this product are temporarily missing versus permanently unavailable. In collection year 2020 and onward, there will *always* be one record generated per scheduled pitfall trap in the `bet_fielddata` table (e.g., in 2020, there are 30 records for each of the planned 30 pitfall traps to be present in a bout).

### 3.4 Variables Reported

All variables reported from the field or laboratory technician (L0 data) are listed in the file, NEON Raw Data Validation for Ground beetles sampled from pitfall traps (DP0.10022.001) (AD[04]). All variables reported in the published data (L1 data) are also provided separately in the files, NEON Data Variables for Ground beetles sampled from pitfall traps (DP1.10022.001) (AD[05]) and NEON Data Variables for Ground beetle sequences DNA barcode (DP1.10020.001) (AD[06]).

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 Temporal Resolution and Extent

The finest resolution at which temporal data are reported is the **daysOfTrapping**, the range between **setDate** and **collectDate**.

**collectDate** (date an individual trap was collected) → **daysOfTrapping**

### 3.6 Spatial Resolution and Extent

The finest resolution at which spatial data are reported is a single trap (Figure 2).

**trapID** (unique ID given to the individual trap) → **plotID** (unique ID given to the plot) → **siteID** (ID of NEON site) → **domainID** (ID of a NEON domain).

The basic spatial data included in the data downloaded include the latitude, longitude, and elevation of the centroid of the plot where sampling occurred + associated uncertainty due to GPS error and plot width. Shapefiles of all NEON Terrestrial Observation System sampling locations can be found here: <http://www.neonscience.org/science-design/field-sites/maps-spatial-data>.

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 trap. Construct the named location of the trap of each record in `bet_fielddata` by concatenating the fields for `namedLocation` and `trapID` as: `namedLocation + ' ' + trapID`, e.g. `trapID 'E'` of `namedLocation 'HARV_026.basePlot.bet'` has a complete named location of `'HARV_026.basePlot.bet.E'`.
  2. Use the API (<http://data.neonscience.org/data-api>; e.g. [http://data.neonscience.org/api/v0/locations/HARV\\_026.basePlot.bet.E](http://data.neonscience.org/api/v0/locations/HARV_026.basePlot.bet.E)) to query for `elevation` (“`locationElevation`”), `easting` (“`locationUtmEasting`”), `northing` (“`locationUtmNorthing`”), `coordinateUncertainty` (“`Value for Coordinate uncertainty`”), `elevationUncertainty` (“`Value for Elevation uncertainty`”), and `utmZone` (“`locationUtmZone`”).
  3. Increase `coordinateUncertainty` by an appropriate amount to account for error introduced by navigating within plots. Technicians may shift pitfall traps up to 2 meters within a plot to avoid obstacles. Additional error may be introduced due to tape stretching to navigate to locations within plots.

### 3.7 Associated Data Streams

**individualID** is the linking variables that tie specific samples and associated metadata between the Ground beetles sampled from pitfall traps (DP1.10022.001) and Ground beetle sequences DNA barcode (DP1.10020.001).



### 3.8 Product Instances

There are a maximum of 13 field season collection bouts per year, with carabids collected from no more than 10 plotIDs per bout. For data collected prior to 2018, each plot will yield no more than 4 samples per bout of collection, resulting in a maximum of 520 plot-bouts per site per year. For collections 2018 and later, each plot yields a maximum of 3 samples within each bout. The number of individuals identified varies with the abundance of organisms at the site.

### 3.9 Data Relationships

The protocol dictates that each trap is collected once per bout (one expected record per trapID per plotID per collectDate in bet\_fielddata; collectDates for a given trap should be separated by a minimum of ~12 days). A record from bet\_fielddata may have zero (if no sample collected) or multiple child records in bet\_sorting depending on number of taxa contained in the sampleID. A record from bet\_sorting may have zero (if no contents of the subsampleID pinned) or multiple child records in bet\_parataxonomistID depending on the number of individuals selected for pinning from each subsampleID. A record in bet\_archivepooling may correspond to one or more records in bet\_subsampling, where multiple subsampleIDs are pooled into a single archiveVial. Each record in bet\_IDandpinning should have zero or one corresponding records in bet\_expertTaxonomistIDProcessed, depending on whether that individualID was selected for professional identification. Each record in bet\_IDandpinning should also have zero or one corresponding records in bet\_expertTaxonomistIDRaw. All beetles must be sorted prior to pinning, so the total number of beetles collected can be calculated as the sum of individualCount in bet\_sorting, though further identifications may be updated based on the downstream workflow. 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.

bet\_fielddata.csv -> One record expected per sampleID for all time; max of one record per trapID per plotID per collectDate.

bet\_sorting.csv -> One record expected per subsampleID for all time, carabid subsamples may generate zero or more children in the bet\_parataxonomistID table

bet\_parataxonomistID.csv -> One record expected per individualID for all time. The number of individualIDs pulled from a given subsampleID should not exceed the individualCount given in the bet\_sorting table.

bet\_archivepooling.csv -> One record expected per archiveID, which is a mixture of subsampleIDs (listed in the subsampleIDList). Not all subsampleID's from bet\_sorting contribute to mixtures; some are pinned or maintained at the trap-level.

bet\_expertTaxonomistIDProcessed.csv -> One expected per individualID for all time

bet\_expertTaxonomistIDRaw.csv -> One expected per individualID for all time

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>

### 3.10 Special Considerations

The ground beetle data are unusual among the TOS data products in that the identification of each ground beetle specimen will involve synthesizing information from various sources, often including a combination of parataxonomist identification, taxonomist identification and DNA barcoding data. For some specimens, only NEON staff identifications will be available. For a subset of specimens, the NEON staff identification will be confirmed by professional taxonomists and/or DNA barcoding. DNA barcoding involves the removal of a leg, or part of a leg, from a beetle, extracting and sequencing DNA from the leg, and matching that sequence data to sequences from previously identified voucher specimens. In some cases, the taxonomist identification and DNA barcoding data will conflict and a determination will have to be made based on a number of factors (e.g., the physical condition of the specimen, the length and quality of the DNA sequence, the phylogenetic clarity of the taxonomic designation provided by the taxonomist and DNA barcoding).

Note: Only specimens that have been identified by an expert taxonomist will be eligible for DNA barcoding. A subset of specimens that receive DNA barcoding will also be photographed. Data from all barcoded beetles (sequence data, location metadata, and photos) will be available on the Barcode of Life Database for public use. The **individualID** (found in the bet\_parataxonomistID table) of all barcoded specimens will allow the end-user to connect NEON carabid data with the barcoding data supplied via the Barcode of Life Database.

## 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 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.

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

### 4.1 Beetle Taxonomy

The master taxonomy for carabid beetles is derived primarily from Lorenz (2005), with some modifications and additions based on Bousquet (2012) and the Integrated Taxonomic Information System (ITIS) on-line database (<http://www.itis.gov>). Hawaiian species are sourced largely from Liebherr and Zimmerman (2000), but include references from 11 additional publications from Liebherr. Taxon ID codes used to identify taxonomic concepts in the NEON master taxonomy list were generated for each taxon by concatenating the first three letters of the genus

name together with the first three letters of the specific epithet to make a unique taxon ID for each scientific name. Where such concatenation would produce duplicate taxon ID codes, numbers were appended to the taxon ID until it was unique within the NEON database (e.g., *Pasimachus sublaevis* as PASSUB1 and *Pasimachus subsulcatus* as PASSUB2). The master taxonomy list includes all carabid species from the continental United States, supplemented with species that are expected to occur at NEON sites in Alaska, Puerto Rico, and Hawaii. NEON plans to keep the taxonomy updated in accordance with the current literature, starting in 2020 and annually thereafter. Geographic ranges and native statuses used in this data product are primarily from Bousquet (2012), future ranges and nativity statuses will be derived from the ITIS on-line database and the current literature.

## 4.2 Mammal Taxonomy

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. 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>).

## 4.3 Herptiles Taxonomy

The master taxonomy for herptiles is derived primarily from Crother et al. (2012). Some modifications and additions based on the Integrated Taxonomic Information System (ITIS) on-line database (<http://www.itis.gov>). Puerto Rican species have been sourced from Rivero (1998) and Schwartz and Henderson (1991). Taxon ID codes used to identify taxonomic concepts in the NEON master taxonomy list were generated for each taxon by concatenating the first three letters of the genus name together with the first three letters of the specific epithet to make a unique taxon ID for each scientific name. Where such concatenation would produce duplicate taxon ID codes, numbers were appended to the taxon ID until it was unique within the NEON database (e.g., *Crotalus pricei* as CROPRI1 and *Crotalus pricei pricei* as CROPRI2). The master taxonomy list includes all herptiles species from the continental United States, supplemented with species that are expected to occur at NEON sites in Alaska, Puerto Rico, and Hawaii. NEON plans to keep the taxonomy updated in accordance with the current literature, starting in 2020 and annually thereafter. Geographic ranges and native statuses used in this data product are derived from the ITIS on-line database and the current literature.



## 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 Validation for Ground beetles sampled from pitfall traps (DPO.10022.001) (AD[04]), 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. Also included in this file is a field named 'entryValidationRulesParser', which describes syntactically the validation rules for each field that is performed upon ingest of the data into the NEON Cyberinfrastructure, based on a standardized data validation language (NiCl) internal to NEON. Please see AD[12] for more information about the NiCl language.

A schematic of the data entry application design is depicted in Figure 3.

### 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; 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.

### 5.4 Quality Flagging

Quality flags are used in multiple tables to signal possible issues with records and/or samples.

#### *bet\_fielddata*

In the *bet\_fielddata* table, potential problems with trap deployment are indicated in the **cupStatus**, **lidStatus**, and **fluidLevel** fields. Starting in 2020, the **samplingImpractical** field in each data record is a quality flag that communicates the reason for missed sampling events. From sampling season 2020 and onwards, there will always be 30

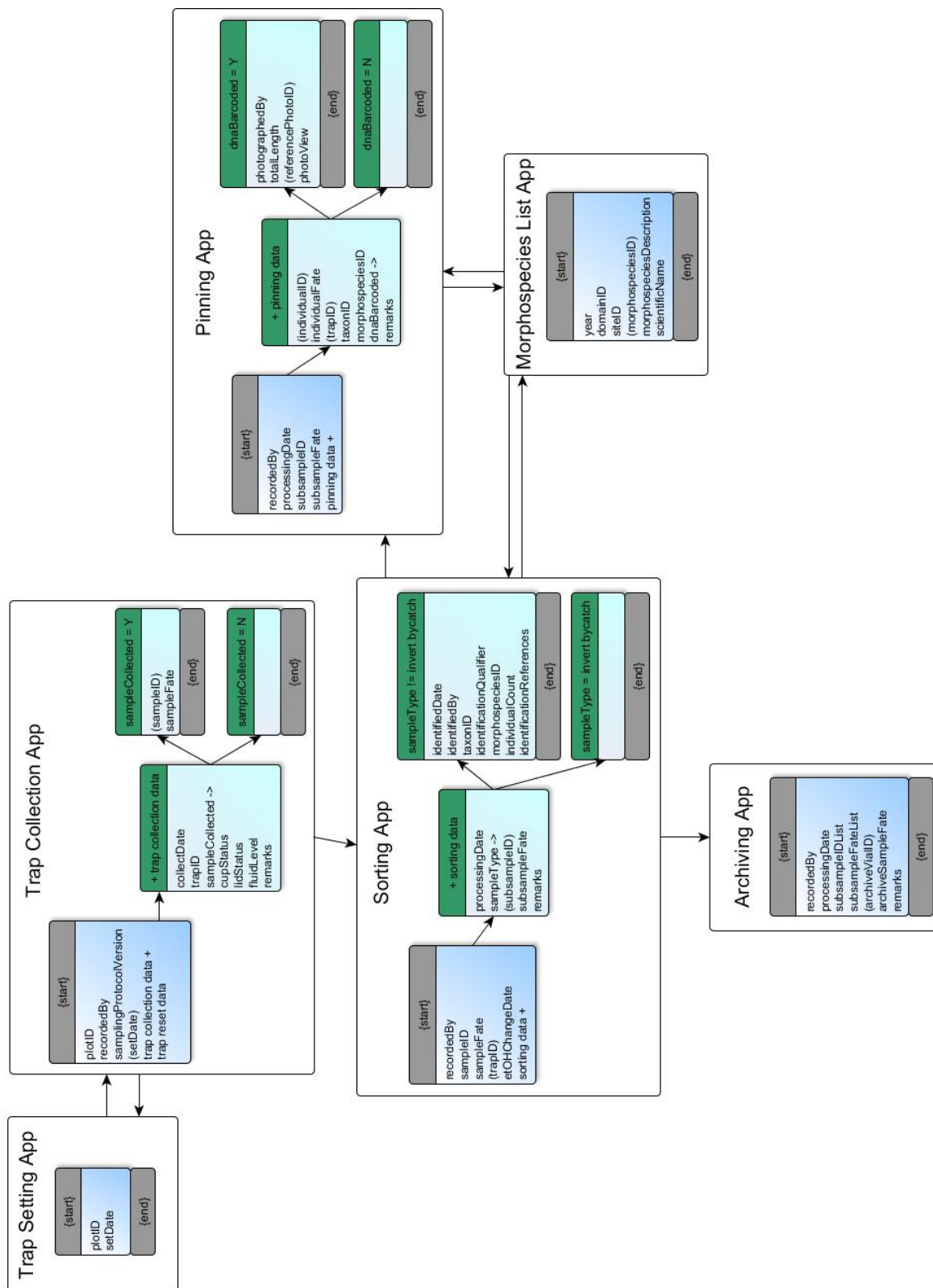


Figure 3: Schematic of the applications used by field technicians to enter ground beetle field data





sampling records per bout of pitfall trapping. 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.

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)

Table 1: Descriptions of the samplingImpractical codes for quality flagging

Value	Description
OK	Sampling occurred at the intended plot location and time
Location snow covered	Sampling did not occur at the intended plot location and time due to the presence of heavy snow cover in excess of 1.8 m at the plot or impeded access to the plot due to snow cover
Location flooded	Sampling did not occur at the intended plot location and time due to flooding at the plot or impeded access to the plot due to flooding
Temperature low	Temperatures at the site were below the threshold for off season sampling, thus off season sampling was not conducted; note that low temperatures are never a valid reason to cancel a field season collection bout
Logistical	Logistical reasons such as site access or staffing difficulties prevented sampling at the scheduled time for that plot
Management	Management activities such as controlled burn, grazing, managed hunts, etc prevented personnel from accessing the site location; see the Site Management and Event Reporting product (DP1.10111.001) for additional details
Extreme weather	Extreme weather such as tornado, hurricane, wildfire, etc present at the site prevented staff from accessing the plot on the scheduled interval
Other	Other activities prevented sampling on schedule at this plot location; these reasons are described in the remarks

*bet\_sorting*

In the *bet\_sorting* table, delays in initial processing (**etOHChangeDate** more than 24 hours after collection) can indicate sample quality issues.

*bet\_archivepooling*

Pooling is intended to save money on sample archiving costs, but is only performed on samples originating from the same bout and plot location. Occasionally, there are errors in pooling and staff inadvertently combine samples from traps at *two different plots* resulting in a very different spatial resolution compared to all other sam-



ple types. These problematic samples will be flagged with the field **pooledFromMultiplePlots** (problem samples flagged with a value of 'Y') where this issue occurs.

*bet\_expertTaxonomistIDProcessed* and *bet\_expertTaxonomistIDRaw*

In both of these tables, **identificationRemarks** is a free text field for comments or notes about the identification that may provide additional context for the identification.

*multiple tables*

Potential sample issues (across multiple tables) are communicated in the **sampleCondition** field. The **identificationQualifier** and **morphospeciesID** fields flag records to indicate level of taxonomic uncertainty with the identification. The **nativeStatusCode** in each data record indicates the nativity status for that species at that site. Nativity is based on checklists and papers from the literature, evaluated on an annual basis. Records where **nativeStatusCode** is 'A', indicate records where species are expected to be absent, signifying that these taxon identifications (if the taxonomic analysis is correct) may represent range expansions.

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.

## 6 REFERENCES

- Bousquet, Y. 2012. Catalogue of Geadephaga (Coleoptera: Adephaga) of America, north of Mexico. *ZooKeys* 245:1-1722.
- Crother, B. I., J. Boundy, F. T. Burbrink, J. A. Campbell, K. De Queiroz, D. R. Frost, D. M. Green, R. Highton, J. B. Iverson, F. Kraus, R. W. Mcdiarmid, J. R. Mendelson III, P. A. Meylan, R. A. Pyron, T. W. Reeder, M. E. Seidel, S. G. Tilley, D. B. Wake, and V. Street. 2012. Scientific and standard english names of amphibians and reptiles of North America north of Mexico, with comments regarding confidence in our understanding. 7th ed. Society for the Study of Amphibians and Reptiles.
- Kotze, D. J., P. Brandmayr, A. Casale, E. Dauffy-Richard, W. Dekoninck, M.J. Koivula, et al. 2011. Forty years of carabid beetle research in Europe - from taxonomy, biology, ecology and population studies to bioindication, habitat assessment and conservation. *ZooKeys* 100: 55–148.
- Kromp, B. 1999. Carabid beetles in sustainable agriculture: a review on pest control efficacy , cultivation impacts and enhancement. *Agriculture Ecosystems & Environment* 74: 187–228.
- Liebherr, J. K., and E. C. Zimmerman. 2000. *Insects of Hawaii, Volume 16: Coleoptera: Carabidae: Part 1: Introduction and Tribe Platynini*. Page 488. University of Hawaii Press.
- Lorenz, W. 2005. Systematic list of extant ground beetles of the world (Insecta, Coleoptera Geadephaga: Trachypachidae and Carabidae incl. Paussinae, Cicindelinae, Rhysodinae). 2nd edition. Lorenz Publishing, Tutzing, Germany
- Rainio, J., and J. Niemela. 2003 . Ground beetles (Coleoptera: Carabidae) as bioindicators. *Biodiversity and Conservation* 12: 487–506.



<i>Title:</i> NEON User Guide to Ground beetles sampled from pitfall traps (DP1.10022.001)	<i>Date:</i> 10/05/2020
<i>Author:</i> Katherine LeVan	<i>Revision:</i> C

Rivero, J. A. 1998. The Amphibians and Reptiles of Puerto Rico. 510 pp. 2nd edition. Editorial de la Universidad de Puerto Rico, San Juan.

Schwartz, A., and R. W. Henderson. 1991. Amphibians and Reptiles of the West Indies. 714 pp. University Press of Florida.

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.