NEON USER GUIDE TO GROUND BEETLES SAMPLED FROM PITFALL TRAPS (NEON.DP1.10022)

PREPARED BY	ORGANIZATION	DATE
Katherine LeVan	SCI	07/25/2017

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	07/19/2017	Initial Release

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 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 (NEON.DP1.10022) (AD[05]) and NEON Data Variables for Ground beetle sequences DNA barcode (NEON.DP1.10020) (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 (NEON.DP0.10022) (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 Level 1, Level 2, and Level 3 Data Products Catalog
AD[04]	NEON.DP0.10003.001_dataValidation.csv	NEON Raw Data Validation for Ground beetles sampled from pitfall traps (NEON.DP0.10022)
AD[05]	NEON.DP1.10022.001_variables.csv	NEON Data Variables for Ground beetles sampled from pitfall traps (NEON.DP1.10022)
AD[06]	NEON.DP1.10020.001_variables.csv	NEON Data Variables for Ground beetle sequences DNA barcode (NEON.DP1.10020)
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]	OS_Generic_Transitions.pdf	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[12]	Nicl Language.pdf	NEON's Ingest Conversion Language (NICL) specifications

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. Four traps are deployed in each of 10 plots at each terrestrial NEON site (40 traps per site), with traps arrayed approximately 20 meters from the center of the plot in each of the four cardinal directions. 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]).

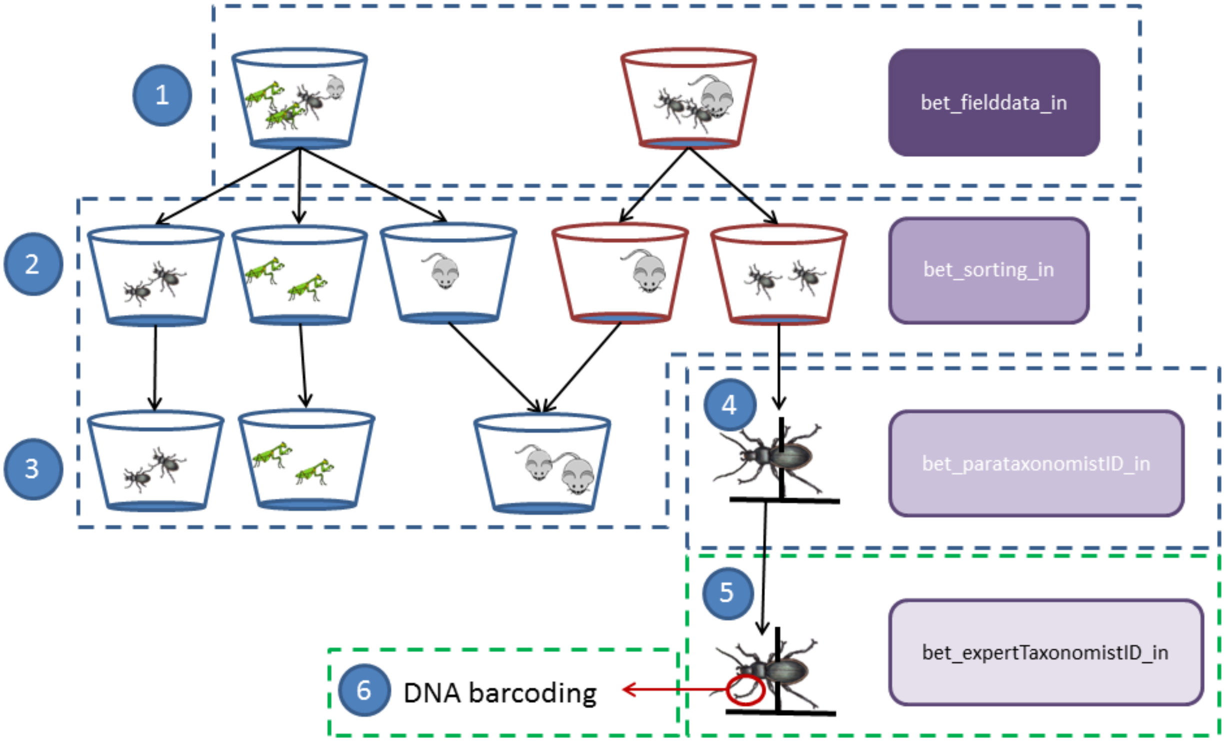

Figure 1: A workflow illustrating the process of data collection for ground beetles

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.

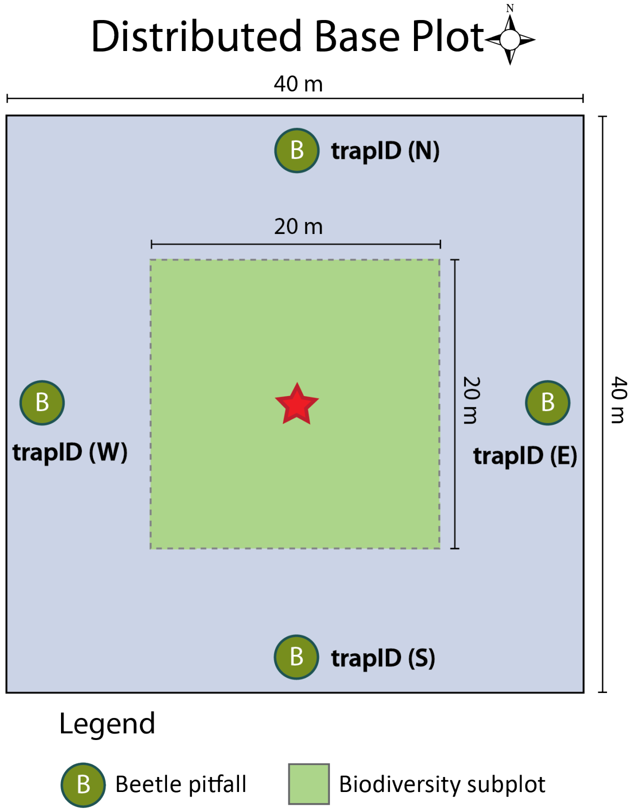

Figure 2: Layout of 4 pitfall traps at a plot. The plot-level centroid is trap E5, designated by the red star.

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 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]).
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 (NEON.DP0.10022) (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 (NEON.DP1.10022) (AD[05]) and NEON Data Variables for Ground beetle sequences DNA barcode (NEON.DP1.10020) (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 Terrestrial Observation System (TOS) spatial data employs the World Geodetic System 1984 (WGS84) for its fundamental reference datum and Earth Gravitational Model 96 (EGM96) 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.

3.3 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**

3.4 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/api; e.g. http://data.neonscience.org/api/v0/locations/HARV_026.basePlot.bet.E) to query for elevation ("locationElevation"), easting("locationUtmEasting"), nor-

thing(“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.

3.5 Associated Data Streams

individualID is the linking variables that tie specific samples and associated metadata between the Ground beetles sampled from pitfall traps (NEON.DP1.10022) and Ground beetle sequences DNA barcode (NEON.DP1.10020).

3.6 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. 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. The number of individuals identified varies with the abundance of organisms at the site.

3.7 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 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 abberations 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

3.8 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 specimens that are relatively easy to identify, only a parataxonomist identification will be required (though even these identifications will undergo quality control checks). For specimens that are more difficult to identify, the parataxonomist 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** 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 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 (NEON.DP0.10022) (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.

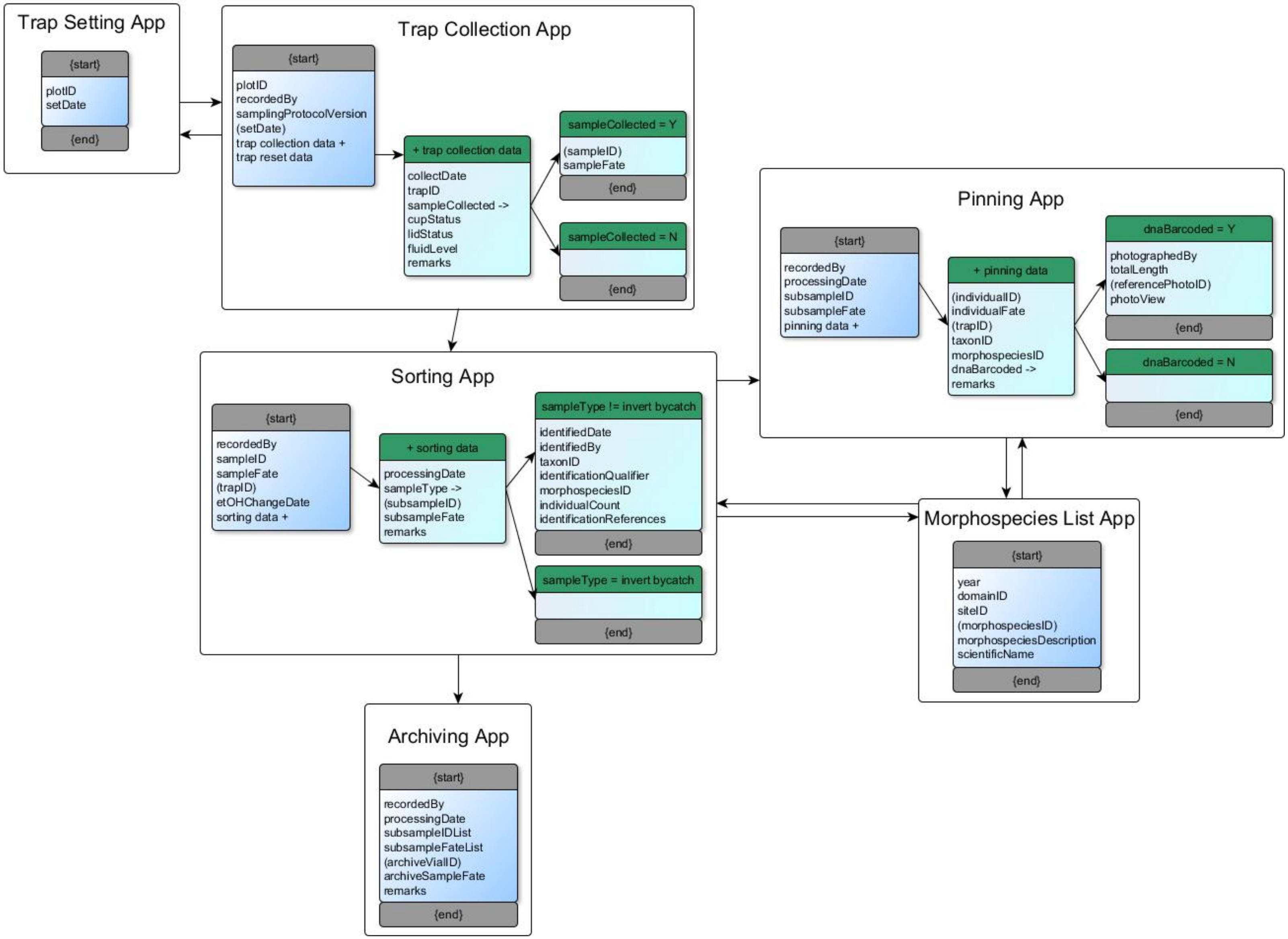

Figure 3: Schematic of the applications used by field technicians to enter ground beetle field data

5 DATA PROCESSING STEPS

Following data entry into a mobile application of 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]).

6 REFERENCES

Kotze, D. J., Brandmayr, P., Casale, A., Dauffy-Richard, E., Dekoninck, W., Koivula, M. J., … Zetto, T. (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.

Rainio, J., & Niemela, J. (2003). Ground beetles (Coleoptera: Carabidae) as bioindicators. Biodiversity and Conservation, 12, 487–506.