

<i>Title:</i> NEON Algorithm Theoretical Basis Document: TOS Ground Beetle Abundance and Diversity - QA/QC of Raw Field and Lab Data	<i>Author:</i> Natalie Robinson	<i>Date:</i> 26 August, 2014
<i>NEON Doc. #:</i> NEON.DOC.001238		<i>Revision:</i> A_DRAFT

# NEON Algorithm Theoretical Basis Document: TOS Ground Beetle Abundance and Diversity - QA/QC of Raw Field and Lab Data

PREPARED BY	ORGANIZATION	DATE
Natalie Robinson	FSU	8/20/2014
David Hoekman	FSU	05/20/2014

APPROVALS (Name)	ORGANIZATION	APPROVAL DATE

RELEASED BY (Name)	ORGANIZATION	RELEASE DATE

See Configuration Management System for approval history.

© 2014 NEON Inc. All rights reserved.

The National Ecological Observatory Network is a project solely funded by the National Science Foundation and managed under cooperative agreement by NEON, Inc. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

<i>Title:</i> NEON Algorithm Theoretical Basis Document: TOS Ground Beetle Abundance and Diversity - QA/QC of Raw Field and Lab Data	<i>Author:</i> Natalie Robinson	<i>Date:</i> 26 August, 2014
<i>NEON Doc. #:</i> NEON.DOC.001238		<i>Revision:</i> A_DRAFT

## Change Record

REVISION	DATE	ECO #	DESCRIPTION OF CHANGE

## TABLE OF CONTENTS

<b>TABLE OF CONTENTS_TOC396735972 .....</b>	<b>I</b>
<b>LIST OF TABLES AND FIGURES.....</b>	<b>III</b>
<b>1 DESCRIPTION.....</b>	<b>1</b>
1.1 Purpose .....	1
1.2 Scope.....	1
<b>2 RELATED DOCUMENTS AND ACRONYMS .....</b>	<b>2</b>
2.1 Applicable Documents .....	2
2.2 Reference Documents.....	2
<b>3 DATA PRODUCT DESCRIPTION.....</b>	<b>3</b>
3.1 Variables Reported .....	3
3.2 Temporal Resolution and Extent .....	4
3.3 Spatial Resolution and Extent .....	5
3.4 Associated Data Streams .....	5
3.5 Product Instances.....	5
<b>4 SCIENTIFIC CONTEXT .....</b>	<b>5</b>
4.1 Theory of Measurement/Observation.....	5
4.2 Theory of Algorithm .....	7
4.2.1 Summary of Algorithm for the Field Data .....	7
4.2.2 Summary of Algorithm for Sorting Data.....	7
4.2.3 Summary of Algorithm for ID and Pinning Data.....	7
4.2.4 Summary of Algorithm for Taxonomist ID Data .....	8
4.2.5 Summary of Algorithm for Vertebrate Bycatch Data .....	8
4.2.6 Summary of Algorithm for Morphospecies.....	8
4.3 Special Considerations .....	8
<b>5 ALGORITHM IMPLEMENTATION .....</b>	<b>8</b>
5.1 Run the following processing steps for data in bet_fielddata_in .....	9
5.2 Run the following processing steps for data in bet_sorting_in .....	10
5.3 Run the following processing steps for data in bet_IDandpinning_in.....	11

<b>6</b>	<b>UNCERTAINTY .....</b>	<b>12</b>
6.1	Analysis of Uncertainty .....	12
6.2	Reported Uncertainty .....	12
<b>7</b>	<b>VALIDATION AND VERIFICATION .....</b>	<b>12</b>
7.1	Algorithm Validation .....	12
7.2	Data Product Validation .....	12
7.3	Data Product Verification .....	12
<b>8</b>	<b>SCIENTIFIC AND EDUCATIONAL APPLICATIONS .....</b>	<b>12</b>
<b>9</b>	<b>FUTURE MODIFICATIONS AND PLANS .....</b>	<b>12</b>
<b>10</b>	<b>BIBLIOGRAPHY .....</b>	<b>14</b>
<b>11</b>	<b>CHANGELOG .....</b>	<b>15</b>
<b>12</b>	<b>APPENDIX: PDA &amp; WEBUI DATA ENTRY AND VALIDATION REQUIREMENTS .....</b>	<b>15</b>
12.1	Common User Interface Requirements .....	15
12.1.1	General Requirements .....	15
12.1.2	fielddata User Interface Requirements .....	16
12.2	PDA User Interface Requirements .....	16
12.2.1	fielddata User Interface Requirements .....	16
12.3	WEBUI User Interface Requirements .....	16
12.3.1	fielddata User Interface Requirements .....	16
12.3.2	sorting User Interface Requirements .....	17
12.3.3	IDandpinning User Interface Requirements .....	18
12.4	Common Validation Requirements .....	19
12.5	PDA Validation Requirements .....	19
12.6	WEBUI Validation Requirements .....	19
12.6.1	fielddata Validation rules .....	19
12.6.2	sorting Validation rules .....	20
12.6.3	IDandpinning Validation rules .....	20

<i>Title:</i> NEON Algorithm Theoretical Basis Document: TOS Ground Beetle Abundance and Diversity - QA/QC of Raw Field and Lab Data	<i>Author:</i> Natalie Robinson	<i>Date:</i> 26 August, 2014
<i>NEON Doc. #:</i> NEON.DOC.001238		<i>Revision:</i> A_DRAFT

## LIST OF TABLES AND FIGURES

<b>Figure 1.</b> Distributed plot layout including 4 pitfall traps (one at each cardinal direction).....	3
<b>Table 1.</b> List of subproducts produced in the data product: ground beetle trapping.....	4

## 1 DESCRIPTION

### 1.1 Purpose

This document details the algorithms used for creating a subset of NEON Level 1 data products that are the quality controlled products generated from raw Level 0 data, and associated metadata.

In the NEON data products framework, the raw data collected in the field are considered the lowest level (Level 0). Raw data that have been quality checked via the algorithms detailed herein, as well as simple metrics that emerge from the raw data, such as total species richness, are considered Level 1 data products. This document relates only to the former group of L1 data products, the quality controlled pass-through products from the Level 0 data products.

Ground beetles will be collected from all NEON sites and identified to species, where possible, or the lowest taxonomic rank that can be defined (e.g. family). This document details the processes necessary to QA/QC Level 0 ground beetle capture data. It includes a detailed discussion of measurement theory and implementation, appropriate theoretical background, data product provenance, quality assurance and control methods used, approximations and/or assumptions made, and a detailed exposition of uncertainty resulting in a cumulative reported uncertainty for this product.

### 1.2 Scope

This document describes the theoretical background and entire algorithmic process for creating a subset of quality controlled and calibrated L1 data products from input data. These data products include the field collection and NEON laboratory component of ground beetle trapping data (NEON.DOM.SIT.DP1.10022). Later revisions, along with additional ATBDs, will add species identifications from external laboratories and/or barcoding facilities, diversity and abundance data. It does not provide computational implementation details, except in cases where these stem directly from algorithmic choices explained here. This document also provides details relevant to the publication of the data products for NEON Data Publication Workbook for TOS Ground Beetle Abundance and Diversity (AD[11]) via the NEON data portal.

This document describes the algorithms for ingesting and performing automated quality assurance and control procedures on the ground beetle data collected in the field as described in Field and Lab Protocol: Ground Beetle Sampling (AD[08]). The raw data that are processed in this document are detailed in the NEON Raw Data Ingest Workbook for TOS Ground Beetle Abundance and Diversity (AD[10]).

## 2 RELATED DOCUMENTS AND ACRONYMS

### 2.1 Applicable Documents

AD[01]	NEON.DOC.000001	NEON Observatory Design (NOD) Requirements
AD[02]	NEON.DOC.005003	NEON Scientific Data Products Catalog
AD[03]	NEON.DOC.005004	NEON Level 1-3 Data Products Catalog
AD[04]	NEON.DOC.005005	NEON Level 0 Data Product Catalog
AD[05]	NEON.DOC.005011	NEON Coordinate Systems Specification
AD[06]	NEON.DOC.002162	NEON Algorithm Theoretical Basis Document: Taxonomic Consensus Identifications and Uncertainty – in prep
AD[07]	NEON.DOC.000909	TOS Science Design for Ground Beetle Abundance and Diversity
AD[08]	NEON.DOC.000451	Field and Lab Protocol: Ground Beetle Sampling
AD[09]	NEON.DOC.001100	Lab Protocol for Beetles and Mosquitoes
AD[10]	NEON.DOC.001400	NEON Raw Data Ingest Workbook for TOS Ground Beetle Abundance and Diversity
AD[11]	NEON.DOC.001411	NEON Data Publication Workbook for TOS Ground Beetle Abundance and Diversity
AD[12]	NEON.DOC.004309	NEON Field Site Information
AD[13]	NEON.DOC.000261	TOS Spatial Data
AD[14]	NEON.DOC.002255	NEON Taxonomic Name List for Ground Beetles
AD[15]	NEON.DOC.002259	NEON Taxonomic Name List for Small Mammals
AD[16]	NEON.DOC.002258	NEON Taxonomic Name List for Reptiles and Amphibians
AD[17]	NEON.DOC.00XXXX	External facilities look-up table – in prep
AD[18]	NEON.DOC.002253	List of Valid Identification Qualifiers

### 2.2 Reference Documents

RD[01]	NEON.DOC.000008	NEON Acronym List
RD[02]	NEON.DOC.000243	NEON Glossary of Terms

### 3 DATA PRODUCT DESCRIPTION

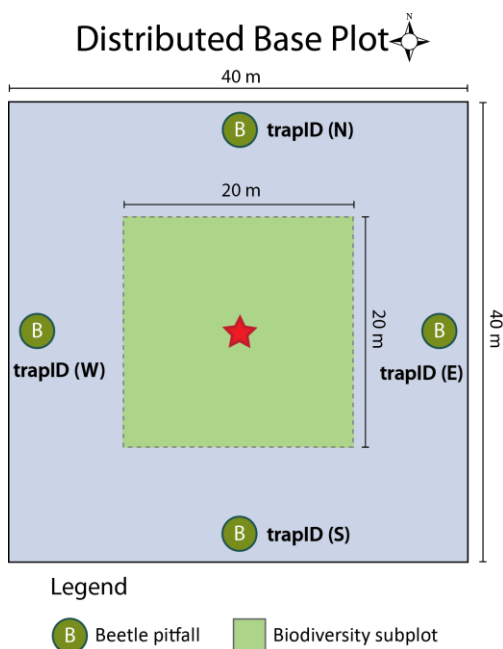


Figure 2: Distributed plot layout including 4 pitfall traps (one at each cardinal direction).

#### 3.1 Variables Reported

This ATBD describes the steps needed to generate the L1 data products: ground beetles sampled from pitfall traps (NEON.DOM.SIT.DP1.10022). Subproducts for this data product are listed below (Table 1). Detailed lists of the associated subproducts and metadata products are provided separately, along with example data in publication-ready spreadsheets in the NEON Data Publication Workbook for TOS Ground Beetle Abundance and Diversity (AD[11]). 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), and with the Bird Monitoring Data Exchange standards (<http://www.avianknowledge.net>; accessed 16 February 2014), where applicable. Geospatial data shall conform to the standards set forth in the NEON Coordinate Systems Specification (AD[05]).



**Table 1.** List of subproducts produced in this ATBD for the dataproduct, ground beetle trapping (NEON.DOM.SIT.DP1.10022). The list is not exhaustive and a variety of supporting data will also be made available.

Number	Field Name	Description
NEON.DOM.SITE.DP1.10022.001.00077.002.001.001	targetTaxaPresent	Indicator of whether the sample contained individuals of the target taxa
NEON.DOM.SITE.DP1.10022.001.00081.002.001.001	morphospeciesID	Identifier for morphospecies
NEON.DOM.SITE.DP1.10022.001.00079.002.001.001	individualCount	Number of individuals of the same type.
NEON.DOM.SITE.DP1.10022.001.00012.002.001.001	taxonID	Species code, based on one or more sources.
NEON.DOM.SITE.DP1.10022.001.00055.002.001.001	scientificName	Scientific name, associated with the taxonID. This is the name of the lowest level taxonomic rank that can be determined.
NEON.DOM.SITE.DP1.10022.001.00056.002.001.001	taxonRank	The lowest level taxonomic rank that can be determined for the individual or specimen.
NEON.DOM.SITE.DP1.10022.001.00013.002.001.001	identificationQualifier	A standard term to express the determiner's doubts about the Identification.
NEON.DOM.SITE.DP1.10022.001.00047.003.001.001	individualID	Domain-level unique identifier for an individual: NEON.MOD.D##.#####

### 3.2 Temporal Resolution and Extent

The finest temporal resolution that ground beetle data will be tracked is at the level of a ~14-day trapping bout. This temporal unit will be used for calculating abundance estimates and other higher level data products, such as species richness and diversity that may be subsequently aggregated to derive measurements over coarser temporal resolutions (e.g., month, season) by end users of the data. The total number of bouts per year will vary among sites, based on the length of the growing season. It is expected that 6-10 bouts of trapping will occur annually during the warmer part of the year. Each year, bout numbering at each site will begin at 1 and increase sequentially.

Documentation of sampling effort (in the form of daysOfTrapping) facilitates calculation of abundance measures, on a per-day basis, by end users, although this information is not calculated as an L1 dataproduct.

### 3.3 Spatial Resolution and Extent

The finest spatial resolution at which ground beetle data will be tracked is per individual pitfall trap (4 traps per **plotID**, one at each cardinal direction). Traps are nested within plots (10 per site), which are placed in order to be representative of a site (e.g., spread across dominant vegetation types). Three sites, a core and two relocatables, are together representative of a domain. Every individual ground beetle specimen will be tied to a specific time and place (**eventID**) where it was collected. Overall, this results in a spatial hierarchy of:

*spatial hierarchy = trapID (ID of trap within plot) → plotID (ID of plot within site) → siteID (ID of NEON site) → domainID (ID of a NEON domain)*

### 3.4 Associated Data Streams

### 3.5 Product Instances

Pitfall traps will be set at all core and relocatable sites. Up to 40 pitfall traps will be set at every site and sampled/reset every 2 weeks during the growing season. It is expected that trapping will result in 100-1000 individual ground beetles per domain per year. The number of specimens that are subsequently pinned/pointed will depend on the proportion of common, easily-identifiable ground beetles. It is expected that 100-400 beetles per domain per year will be selected for barcoding and/or identification by taxonomists.

## 4 SCIENTIFIC CONTEXT

### 4.1 Theory of Measurement/Observation

The Terrestrial Observation System (TOS) is charged with monitoring the responses of biodiversity and ecosystems to environmental change. While several different invertebrate groups were considered, a NEON design committee (AIBSnews, 2007) selected ground beetles (Coleoptera: Carabidae) as a focal taxon for measurement.

Title: NEON Algorithm Theoretical Basis Document: TOS Ground Beetle Abundance and Diversity - QA/QC of Raw Field and Lab Data	Author: Natalie Robinson	Date: 26 August, 2014
NEON Doc. #: NEON.DOC.001238		Revision: A_DRAFT

Ground beetles have been extensively used as indicator species of arthropod biodiversity, environmental change, and the impacts of urbanization on ecological systems (Koivula, 2011; Niemelä & Kotze, 2009; Rainio & Niemela, 2003), land use (Vanbergen, Woodcock, Watt, & Niemelä, 2005) and land management (Work et al., 2008). The ground beetle family is large, widespread, and associated with the ground layer, rather than any particular plant community, ensuring that some ground beetles are present in virtually all terrestrial habitats over the entire extent of the observatory (Lövei & Sunderland, 1996). The family is rich in species (over 40,000 species described globally, circa 3,000 species in NEON's spatial extent), abundant in individuals, and well-known taxonomically (Laroche & Larivière, 2003). Though ground beetles range in body size from less than a millimeter to several centimeters, they are straightforward to sample and identify (Kotze et al., 2011).

Ground beetles are sensitive to environmental conditions, and ground beetle communities form well-defined richness gradients (e.g. along the North American latitudinal gradient). Consequently, ground beetles are considered "sentinels" of arthropod biodiversity (Koivula, 2011; Rainio & Niemela, 2003), and are valuable model organisms for studies in the fields of population biology, landscape ecology, and conservation biology (Kotze et al., 2011). In addition to what can be learned from ground beetle abundance and diversity measures themselves, these organisms also form an important component of terrestrial food webs, and can thus influence terrestrial trophic structure. For example, many ground beetle species are common prey for small mammals, birds, reptiles, amphibians and other larger arthropods (Laroche & Larivière, 2003). Furthermore, ground beetles play important roles in agricultural landscapes, through contributing to the biological control of pests (Kromp, 1999a), and reducing weed seed abundance through predation (Gaines & Gratton, 2010).

To promote utility to the scientific community, ground beetles will be sampled using pitfall traps. 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, 1999b; Rainio & Niemela, 2003) was selected to maximize comparability across time and between domains within the observatory. In addition, NEON data collected in this way will be comparable to datasets collected by many other ecologists.

Samples collected from pitfall traps in the field are then sorted back in the NEON domain laboratory. 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. Finally, carabid beetles are sorted by NEON technicians, identified to species or morphospecies, and a subset of these are pointed or pinned. Data recorded during pinning/pointing include entries in the 'specimenSource' field, which helps end user avoid double-counting individuals for which data are generated both during sorting and pinning/pointing (indicated by a value of 'vial of pooled individuals' in the specimenSource field). Following pinning/pointing, a subset of individuals is sent to an expert taxonomist and/or a DNA barcoding facility for identification.

## 4.2 Theory of Algorithm

This document describes the algorithms for assessing the integrity of the L0 data stream generated by the field sampling and lab analysis of ground beetle abundance and diversity. The approaches described herein are simple yet necessary components of quality control and quality assurance, and include such processing steps as: verifying that all required data are recorded for each capture event, comparing data values to pre-defined (through provided validation rules or specified lookup tables) ranges of possible values, and tracking the individual-level data for consistency and accuracy through time.

### 4.2.1 Summary of Algorithm for the Field Data

1. Check for duplicate data.
2. Generate eventID.
3. Assign spatial information (plot-level NLCD class, elevation, latitude, longitude, and uncertainty in the latter three measurements).
4. Verify that date entered for trap collection is later than that entered for trap setting.
5. Calculate the number of days in each trapping bout.
6. Generate a bout identifier.

### 4.2.2 Summary of Algorithm for Sorting Data

1. Check for duplicate data.
2. Assign taxonID, taxonRank and scientificNameAuthorship values using the reported acceptedTaxonID and the taxonomic name list for ground beetles lookup table (AD[14]).
3. Verify that entries in the associatedSampleID field are present in the sampleID field of AD[10]:bet\_fielddata\_in.

### 4.2.3 Summary of Algorithm for ID and Pinning Data

1. Check for duplicate data.
2. Assign taxonID, taxonRank and scientificNameAuthorship values using the reported acceptedTaxonID and the taxonomic name list for ground beetles lookup table (AD[14]).
3. Verify that entries in the sampleID field are present in the sampleID field of AD[10]:bet\_fielddata\_in OR AD[10]:bet\_sorting\_in.

#### 4.2.4 Summary of Algorithm for Taxonomist ID Data

This section will be updated in future versions of this ATBD

#### 4.2.5 Summary of Algorithm for Vertebrate Bycatch Data

This section will be updated in future versions of this ATBD

#### 4.2.6 Summary of Algorithm for Morphospecies

This section will be updated in future versions of this ATBD

### 4.3 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 or updated 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). These steps are not part of this ATBD and will be addressed in the taxonomy ATBD (AD[06]). They are not necessary for the initial release of raw data.

## 5 ALGORITHM IMPLEMENTATION

\* Throughout the algorithm implementation section of this ATBD, 'nodata', 'null', and/or 'NA' indicates a blank cell. All variables reported from the field or laboratory technician (LO) are listed in the data ingest workbook (AD[10]) notated here as `bet_tablename_in`. Unless otherwise specified in the algorithm below, all variables that appear in tables `bet_tablename_pub` have been passed directly from the LO variables with the same name, as listed in the data publication workbook (AD[11]). Algorithm

implementation should proceed in the order of the subsections provided here, e.g., the processing of field collected data (subsection 5.1) should occur prior to that of sorting data (subsection 5.2), and so on.

## Automated Processing Steps for Validation of Field Collected Data

### 5.1 Run the following processing steps for data in `bet_fielddata_in`

1. Check for duplicate values
  - a. List of fieldNames = [**plotID**, **trapID**, **boutNumber**, year component of **collectDate** field]
  - b. Create quality flag field and populate with zeroes
    - i. QF Name = **duplicateCollectionEventQF**
  - c. For any records where **trapID** is NULL
    - i. set **duplicateCollectionEventQF** to -1
  - d. Check for duplicate records where **trapID** is not NULL, based on exact matches in the fields listed in (a).
  - e. If duplicate records are identified:
    - i. If values in all remaining fields (e.g., **samplingProtocol**, **cupStatus**) *except* **uid** AND/OR **enteredBy** are also the same:
      - A. Pass only one of the records into `bet_fielddata_pub`
      - B. Enter a '1' into the corresponding cell in the **duplicateCollectionEventQF** field
    - ii. Else:
      - A. Pass duplicate records into `bet_fielddata_pub`
      - B. Enter a '2' into the corresponding cell in the **duplicateCollectionEventQF** field
2. Generate **eventID**
  - a. For each record, concatenate: 'BET' + value in **plotID** field + year component of **collectDate** field + value in **boutNumber** field, separated by '.'
3. Check that the sampling bout was sufficiently documented
  - a. Create quality flag field and populate with zeroes
    - i. QF Name = **missingRecordsPerBoutQF**
  - b. For each **eventID**, verify that there is one of each of the following values in the **trapID** field: 'N', 'S', 'E', 'W'. If any of these trapIDs are missing:
    - i. create n new records in `bet_fielddata_pub`, where n = number of missing trapIDs from list in (b), and assign the value(s) of the missing trapID(s) from list (b) into the **trapID** field(s) of the new record(s)
    - ii. assign the values for **uid**, **eventID**, **domainID**, **siteID**, **plotID**, and **boutNumber** accordingly
    - iii. populate the **collectDate** field for the given **eventID** with the value from the **collectDate** field of any of the existing records in list (b)
    - iv. insert '1' into **missingRecordsPerBoutQF** of all new records for the given **eventID**
    - v. insert 'This record was created by the system' into the **remarks** field of all new records

4. Generate spatial location and uncertainty information
  - a. Insert corresponding values of **nlcdClass**, **decimalLatitude**, **decimalLongitude**, **geodeticDatum**, **elevation**, **coordinateUncertainty**, **elevationUncertainty** values for each **plotID** and **trapID** combination by matching each {**plotID**, **trapID**} combination with a {**plotID**, **pointID**} combination in AD[13] (and where AD[13]:**applicableModule** = 'bet').
5. Confirm that **collectDate** values are later than the **setDate** values
  - a. Create quality flag field and populate with zeroes
    - i. QF Name = **compareSetCollectDateQF**
  - b. For each record: if collectDate<setDate, set **compareSetCollectDateQF=1**
6. Generate **daysOfTrapping** values
  - a. For all records where **compareSetCollectDateQF=1**, set **daysOfTrapping** = NULL
  - b. For all records not included in (a), calculate the **daysOfTrapping**
    - i. Convert **setDate** and **collectDate** to day of year: starting with January 1<sup>st</sup> as day one of the year, calculate the day of year for each date with attention to whether the year in question is a leap year.
    - ii. Insert the difference between the **setDate** and **collectDate** (collect day of year minus set day of year) into the corresponding cell in the **daysOfTrapping** field. If the **collectDate** is in a later year than **setDate**, use one of the following equations:  
**daysOfTrapping** = (365 – set day of year) + collect day of year [normal year]  
**daysOfTrapping** = (366 – set day of year) + collect day of year [leap year]

## 5.2 Run the following processing steps for data in **bet\_sorting\_in**

1. Check for duplicate values
  - a. List of fieldNames: [**sampleID**]
  - b. Create quality flag and populate with zeroes
    - i. QF Name = **duplicateSampleIDQF**
  - c. Check for duplicate records, based on exact matches in the fields listed in (a).
  - d. If duplicate records are identified:
    - i. If values in all remaining fields (e.g., **taxonID**, **individualCount**) *except* **uid** AND/OR **enteredBy** are also the same:
      - A. Pass only one of the records into **bet\_sorting\_pub**
      - B. Enter a '1' into the corresponding cell in the **duplicateSampleIDQF** field
    - ii. Else:
      - A. Pass duplicate records into **bet\_sorting\_pub**
      - B. Enter a '2' into the corresponding cell in the **duplicateSampleIDQF** field
2. Assign L1 **scientificName**, **taxonRank** and **scientificNameAuthorship** values, based on matching **acceptedTaxonID** values to those in beetle, bycatch, and small mammal lookup tables:
  - a. For each record in **bet\_sorting\_in**, use the value in the **acceptedTaxonID** field (if not null) to locate the correct record in one of the following taxon-specific lookup

Title: NEON Algorithm Theoretical Basis Document: TOS Ground Beetle Abundance and Diversity - QA/QC of Raw Field and Lab Data	Author: Natalie Robinson	Date: 26 August, 2014
NEON Doc. #: NEON.DOC.001238		Revision: A_DRAFT

tables: NEON Taxonomic Name List for Ground Beetles (AD[14]) (for sample type = 'common carabid'), NEON Taxonomic Name List for Small Mammals (AD[15]) (for sample type = 'vert bycatch mam'), NEON Taxonomic Name List for Reptiles and Amphibians (AD[16]) (for sample type = 'vert bycatch herp').

- b. Copy data from the **scientificName**, **taxonRank** and **scientificNameAuthorship** fields of the lookup table to the same fields in bet\_sorting\_pub.
- c. If the **acceptedTaxonID** value is null in bet\_sorting\_db, insert nodata into the **scientificName**, **taxonRank** and **scientificNameAuthorship** fields of bet\_sorting\_pub.
- d. Assign **beetle\_sorting\_pub:taxonID**:
  - a. **beetle\_sorting\_pub:taxonID** = bet\_sorting\_in:acceptedTaxonID
  - b. **beetle\_sorting\_in:taxonID** is L0 data (possibly containing synonyms) and retained for traceability only and not passed on to users.
3. Do not pass records of species of conservation concern to the L1:
  - a. For each record in bet\_sorting\_pub:
    - i. Use **acceptedTaxonID** and **domainID** to lookup the dXXNativeStatus for the relevant domain (e.g., d10NativeStatus where domainID=D10), based on AD[14], AD[15], or AD[16].
    - ii. If dXXNativeStatus = 'STATE ' or 'FEDERAL':
      - A. Remove record from bet\_sorting\_pub

### 5.3 Run the following processing steps for data in bet\_IDandpinning\_in

1. Check for duplicate values
  - a. List of fieldNames: [individualID]
  - b. Create quality flag and populate with zeroes
    - i. QF Name = **duplicateIndividualIDQF**
  - c. Check for duplicate records, based on exact matches in the fields listed in (a).
  - d. If duplicate records are identified:
    - i. If values in all remaining fields (e.g., **taxonID**, **identificationQualifier**) *except* **uid** AND/OR **enteredBy** are also the same:
      - A. Pass only one of the records into bet\_IDandpinning\_pub
      - B. Enter a '1' into the corresponding cell in the **duplicateIndividualIDQF** field
    - ii. Else:
      - A. Pass duplicate records into bet\_IDandpinning\_pub
      - B. Enter a '2' into the corresponding cell in the **duplicateIndividualIDQF** field
2. Assign L1 **scientificName**, **taxonRank** and **scientificNameAuthorship** values
  - a. For each record, run algorithm 5.2.2 with the same inputs, but with bet\_IDandpinning\_in rather than bet\_sorting\_in
3. Do not pass records of species of conservation concern to the L1:
  - c. For each record in bet\_IDandpinning\_pub:



- i. Use **acceptedTaxonID** and **domainID** to lookup the dXXNativeStatus for the relevant domain (e.g., d10NativeStatus where domainID=D10), based on AD[14])
- ii. If dXXNativeStatus = 'STATE ' or 'FEDERAL':
  - A. Remove record from bet\_IDandpinning\_pub

## 6 UNCERTAINTY

### 6.1 Analysis of Uncertainty

### 6.2 Reported Uncertainty

*Although no quantitative algorithms are available to incorporate many of these sources of uncertainty into the associated data products, NEON can produce summary uncertainty reports for these observational data products. The algorithm for assessing error rates in taxonomic identifications is presented elsewhere [AD[07]].*

## 7 VALIDATION AND VERIFICATION

### 7.1 Algorithm Validation

### 7.2 Data Product Validation

### 7.3 Data Product Verification

## 8 SCIENTIFIC AND EDUCATIONAL APPLICATIONS

## 9 FUTURE MODIFICATIONS AND PLANS

Future additions to this ATBD will include activities related to constructing new webUIs and processing data from outside facilities (e.g., taxonomist laboratories) and also to DNA barcoding. Some ground beetles will be photographed and tissue extracted (leg removed) for DNA barcoding, after pinning/pointing. These specimens will also be identified by taxonomists and some may be added to the domain voucher collection to aid identification in future seasons. The photographs will be submitted along with tissue samples and associated collection metadata to add to the barcode record.

One potential webUI development step would be to use sampleIDs generated in the field to populate a dropdown menu of options for the sorting activity, such that users could pick from a list of available sampleIDs (now representing the associatedSampleID) instead of these IDs being generated anew. Similarly, sampleIDs generated during field collection and sorting could be used to populate a dropdown

menu in the IDandpinning activity, again providing users with a list of sampleIDs from which the individual being pinned/pointed was taken.

Some of the webUI development steps that will be required for data from outside facilities include:

*For data in a vertebratebycatch table:*

1. Provide filtered drop-down menus for location fields in the [bet\_dataingest\_2014] workbook:
  - a. domainID, siteID, plotID
2. Provide drop-down menus for:
  - trapID
  - taxonRank
  - identificationQualifier
  - scientificName
  - iucnThreatStatus

*For data in a taxonomistID table:*

1. Provide drop-down menus for:
  - taxonRank
  - sex
  - identificationQualifier
  - scientificName
  - shipmentCondition
  - sampleProcess
2. individualIDs:
  - a. There will be one individualID with each beetle pinned or stored in ethanol, this coming from a pre-printed label
    - i. PDA: Scan the QR code to populate field
    - ii. UI: Tech records information from pre-printed label
      - A. Example: BET.D10.000025
3. Populate taxonID:
  - a. lookup table: NEON.DOC.##### NEON Taxonomic Name List for Ground Beetles
  - b. filtering rules: select (unique(taxonID, scientificName)) WHERE
    - i. lookupable.domainID == current domainID AND
    - ii. D##NativeStatusCode != 'A'
    - iii. Other options:
      - A. OTGB: Other – ground beetle
      - B. OTHE: Other – other
      - C. If either of these are selected, prompt user to enter additional information in the remarks
  - c. taxonID and acceptedTaxonID are sent to Maximo/PDR
4. Populate family, genus, subgenus, and specificEpithet:

- a. lookup table: NEON.DOC.##### NEON Taxonomic Name List for Ground Beetles
- b. filtering rules: select (unique(taxonID, scientificName)) WHERE
  - i. lookuptable.domainID == current domainID AND
  - ii. D##NativeStatusCode != 'A'
  - iii. Other options:
    - A. OTGB: Other – ground beetle
    - B. OTHE: Other – other
    - C. If either of these are selected, prompt user to enter additional information in the remarks
5. Add family, genus, subgenus, and specificEpithet to table 1.

Some of the algorithms that will be developed to handle data from outside facilities include:

*For data in a vertebratebycatch table:*

1. Flag duplicate records
2. Assign **scientificName**, **taxonRank** and **scientificNameAuthorship** in a manner similar to that used in section 5.2.2.

*For data in a taxonomistID table:*

1. Check for duplicate records, based on the **individualID** field
2. Assign **scientificName**, **taxonRank**, **scientificNameAuthorship**, **family**, **genus**, **subgenus** and **specificEpithet** values in a manner similar to that used in section 5.2.2
3. Verify that entries in the **individualID** field are present in the **individualID** field of bet\_IDandpinning\_in in a manner similar to that used in section 5.2.3

In addition, rules may be created for: 1) governing the creation of individualIDs in a constrained manner, and one that differs from the current procedure in which domain labs generate individualIDs, and 2) updating records to reflect correct taxonomic identifications when specimens that were pinned/pointed and then sent to a taxonomist for identification are returned.

## 10 BIBLIOGRAPHY

AIBSnews. (2007). NEON design 2007. *BioScience*, 57(2), 198–200.

Gaines, H. R., & Gratton, C. (2010). Seed predation increases with ground beetle diversity in a Wisconsin (USA) potato agroecosystem. *Agriculture, Ecosystems & Environment*, 137(3-4), 329–336. doi:10.1016/j.agee.2010.03.003

Koivula, M. J. (2011). Useful model organisms, indicators, or both? Ground beetles (Coleoptera, Carabidae) reflecting environmental conditions. *ZooKeys*, (100), 287–317.

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. (1999a). Carabid beetles in sustainable agriculture : a review on pest control efficacy , cultivation impacts and enhancement. *Agriculture Ecosystems & Environment*, 74, 187–228.

Kromp, B. (1999b). Carabid beetles in sustainable agriculture : a review on pest control efficacy , cultivation impacts and enhancement. *Agriculture Ecosystems & Environment*, 74, 187–228.

Larochelle, A., & Larivière, M. C. (2003). *A Natural History of the Ground-Beetles (Coleoptera : Carabidae) of America North of Mexico*. *Faunistica No 27* (p. 583). Pensoft.

Lövei, G. L., & Sunderland, K. D. (1996). Ecology and behavior of ground beetles (Coleoptera: Carabidae). *Annual review of entomology*, 41(1 12), 231–56. doi:10.1146/annurev.en.41.010196.001311

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

Vanbergen, A. J., Woodcock, B. A., Watt, A. D., & Niemelä, J. (2005). Effect of land-use heterogeneity on carabid communities at the landscape scale. *Ecography*, 28, 3–16.

Work, T. T., Koivula, M. J., Klimaszewski, J., Langor, D., Spence, J., Sweeney, J., & Hebert, C. (2008). Evaluation of carabid beetles as indicators of forest change in Canada. *Canadian Entomologist*, 140(4), 393.

## 11 CHANGELOG

## 12 APPENDIX: PDA & WEBUI DATA ENTRY AND VALIDATION REQUIREMENTS

### 12.1 Common User Interface Requirements

#### 12.1.1 General Requirements

1. Follow guidelines as specified in the ground beetle data ingest workbook: betFieldSummary\_PDA (AD[10]).
  - a. When a new record is created follow guidelines for default values, as specified in **defaultValuePDA** and **defaultValueUI**.
  - b. All fields must follow guidelines for **case**.

2. The user shall have the ability to delete any record which is not finalized.

### 12.1.2 fielddata User Interface Requirements

1. Provide drop-down menu for domainID
  - a. For subsequent record, auto-populate domainID with value entered previously
2. Provide drop-down menu of siteID, filtered by the selected domainID
  - a. For subsequent record, auto-populate siteID with value entered previously
3. Provide drop-down menu of plotID, filtered by the selected siteID
  - a. To generate menu of plotIDs, use the TOS plot-level spatial data lookup table (AD[13]) and where 'bet' is an element of **applicableModules**
  - b. For subsequent record, auto-populate plotID with value entered previously
4. Provide drop-down menu of available versions of samplingProtocol
5. Provide drop-down menu of trapID
6. Provide drop-down menu of cupStatus
7. Provide drop-down menu of lidStatus
8. Provide drop-down menu of fluidLevel
9. Provide drop-down menu of trapReset
10. recordedBy – Defaults to the user logged in to the Mobile Data Recorder (MDR) app or web user interface when the record is created.

## 12.2 PDA User Interface Requirements

### 12.2.1 fielddata User Interface Requirements

1. sampleIDs: User will scan the QR code to populate field
2. PDA will send values from recordedBy as entered on the PDA to both 'recordedBy' and 'enteredBy' in PDR

## 12.3 WEBUI User Interface Requirements

### 12.3.1 fielddata User Interface Requirements

1. sampleIDs: generate by concatenating: plotID+'.'+trapID+'.'+collectDate
  - a. Allow user to indicate that no sample was collected, in which case no sampleID should be generated
2. Auto-populate boutNumber with values entered previously, if available

### 12.3.2 sorting User Interface Requirements

1. Provide drop-down menu for domainID
  - a. For subsequent record, auto-populate domainID with value entered previously
2. Provide drop-down menu of siteID, filtered by the selected domainID
  - a. For subsequent record, auto-populate siteID with value entered previously
3. Provide drop-down menus of plotID, filtered by the selected siteID
  - a. To generate menu of plotIDs, use the TOS plot-level spatial data lookup table (AD[13]) and where 'bet' is an element of **applicableModules**
  - b. For subsequent record, auto-populate plotID with value entered previously
4. Provide drop-down menu of trapID
5. Provide drop-down menu of available versions of samplingProtocol
6. Provide drop-down menu of targetTaxaPresent
7. processingDate - Auto-populate values in subsequent records with value entered previously
8. taxonID:
  - a. lookup table for sampleType = 'common carabid' or 'other carabid':  
NEON.DOC.002255 NEON Taxonomic Name List for Ground Beetles (AD[14])
  - b. lookup table for sampleType = 'vert bycatch mam': NEON.DOC.002259 NEON Taxonomic Name List for Small Mammals (AD[15])
  - c. lookup table for sampleType = 'vert bycatch herp': NEON.DOC.002258 NEON Taxonomic Name List for Reptiles and Amphibians (AD[16])
  - d. field is a typeahead in which the user can type in taxonID or scientificName
  - e. populate list of values with taxa that are known to occur within the relevant domain (D##), using the D##NativeStatusCode field in the lookup table. Taxa to be included are indicated by all values in this field that are not 'A' for absent.
9. identificationReferences should be a free-form text field (until a list can be compiled by Science)
10. recordedBy – Defaults to the user logged in to the Mobile Data Recorder (MDR) app or web user interface when the record is created
11. associatedSampleID: generate by concatenating: plotID+'.'+trapID+'.'+collectDate
12. sampleType:
  - a. Do not allow value in sampleType field to be 'common carabid' if user selected 'N' in the targetTaxaPresent field
  - b. If the user has selected 'other carabid' as the sampleType, do not require values in the taxonID, morphospeciesID, or individualCount fields
13. sampleIDs:
  - a. If taxonID != NULL, generate by concatenating:  
plotID+'.'+trapID+'.'+collectDate+'.'+taxonID+'.'+##
    - i. ## indicates the tube number
  - b. If taxonID == NULL:

- i. If the user has checked the box indicating that 'This is not a pooled sample', generate by concatenating:
    - I. plotID+'.'+trapID+'.'+collectDate+'.'+sampleType+'.'+##
    - II. sampleType = 'ib', 'vb', 'vb', 'oc', or 'cc' for recorded sampleTypes of 'invert bycatch', 'vert bycatch mam', 'vert bycatch herp', 'other carabid', and 'common carabid', respectively
    - III. ## indicates the tube number
  - ii. Else, generate by concatenating:
    - I. plotID+'.'+'.'+collectDate+'.'+sampleType+'.'+##
    - II. sampleType = 'ib', 'vb', 'vb', 'oc', or 'cc' for recorded sampleTypes of 'invert bycatch', 'vert bycatch mam', 'vert bycatch herp', 'other carabid', and 'common carabid', respectively
    - III. ## indicates the tube number
14. identifiedBy – Maximo user list for FOPs or 'Other'

### 12.3.3 IDandpinning User Interface Requirements

1. Provide drop-down menu for domainID
  - a. For subsequent record, auto-populate domainID with value entered previously
2. Provide drop-down menu of siteID, filtered by the selected domainID
  - a. For subsequent record, auto-populate siteID with value entered previously
3. Provide drop-down menus of plotID, filtered by the selected siteID
  - a. To generate menu of plotIDs, use the TOS plot-level spatial data lookup table (AD[13]) and where 'bet' is an element of **applicableModules**
  - b. For subsequent record, auto-populate plotID with value entered previously
4. Provide drop-down menu of trapID
5. Provide drop-down menu of available versions of samplingProtocol
6. sampleIDs:
  - a. If specimenSource == 'field sample', generate by concatenating:  
plotID+'.'+trapID+'.'+collectDate
  - b. If specimenSource == 'vial of pooled individuals', generate by:
    - i. concatenating: plotID+'.'+trapID+'.'+collectDate+'.'
    - ii. Prompting users to add the remainder of the sampleID on the tube from which the specimen is extracted
7. identifiedBy – Maximo user list for FOPs or 'Other'

## 12.4 Common Validation Requirements

All data must pass the field specific validation rules found in the ground beetle data ingest workbook: betFieldSummary\_PDA (AD[10]).

The rules are to be evaluated in the following order:

1. Follow guidelines for fields in which no data have been entered, as specified in **noDataOutcomePDA (for mobile devices)** and **noDataOutcomeUI (for WEBUI)**:
  - a. If **noDataOutcome** = fail:
    - i. do not let user finalize record until a value is provided
    - ii. warning message text, 'Please enter a value for [fieldname] to continue', unless an alternative is provided in **warningText**
  - b. Else if **noDataOutcome** = warn:
    - i. warn user that a value is missing prior to finalizing record, but allow selection of 'OK' to continue without a value
    - ii. warning message text, 'Please confirm that there is no value for [fieldname] to continue', unless an alternative is provided in **warningText**
  - c. Else if **noDataOutcome** = pass:
    - i. allow user to finalize record with no values in this field
2. If a field has data, constrain the entered value to the correct **dataType**. (If the field is null or empty only **noDataOutcome** is performed.) If the field has data and the data is not of the correct type display an error message stating 'The value for [fieldname] must be a [datatype]' and do not allow the record to be finalized.
3. If a field has data of the correct data type, constrain entered values to valid entries, as described in the **entryValidationRules** and do not allow the record to be finalized if it does not pass validation. Display an error message based on the type of error: For example, for range errors, display an error message stating 'The value for [fieldname] must be between [minval] and [maxval]'.

## 12.5 PDA Validation Requirements

## 12.6 WEBUI Validation Requirements

### 12.6.1 fielddata Validation rules

1. cupStatus: if user has indicated that no sample was collected, do not let user finalize record if cupStatus is set to 'Ok'



<i>Title:</i> NEON Algorithm Theoretical Basis Document: TOS Ground Beetle Abundance and Diversity - QA/QC of Raw Field and Lab Data	<i>Author:</i> Natalie Robinson	<i>Date:</i> 26 August, 2014
<i>NEON Doc. #:</i> NEON.DOC.001238		<i>Revision:</i> A_DRAFT

### 12.6.2 sorting Validation rules

1. remarks: if morphospeciesID != NULL then the remarks cannot be empty

### 12.6.3 IDandpinning Validation rules

1. remarks: if morphospeciesID != NULL then the remarks cannot be empty