

# STANDARD OPERATING PROCEDURE

## CMNH VERIFICATION OF DATA ENTRY FOR IDENTIFICATIONS

### 1. PURPOSE

- 1.1 To provide a general procedure for verifying the proper recording of specimen identifications in spreadsheets before ingesting into the NEON data repository.

### 2. SCOPE

- 2.1 This procedure applies to all CMNH personnel involved in the handling of Carabid beetle specimens for the National Ecological Observatory Network (NEON) program at CMNH.

### 3. RESPONSIBILITIES

- 3.1 The NEON Carabid beetle taxonomic identifier is responsible for: 1) identifying carabid beetle specimens to lowest taxonomic level possible (usually species or subspecies), 2) annotation of the appropriate domain identification worksheet columns with the correct species information and associated data based on the Individual ID, 3) once identification of an entire domain is complete, specimen counts and identifications for each species are verified and 4) updates are made to the spreadsheet, if necessary, before uploading to the NEON data portal.
- 3.2 CMNH data manager is responsible for 1) taking the verified CSV formatted file from the taxonomic identifier and uploading it to the NEON data portal, 2) if upload fails for some reason, the data manager consults with the taxonomic identifier or NEON staff to resolve the issue, and then attempts to upload the file again until it is successfully ingested.

### 4. BACKGROUND, DEFINITIONS & ABBREVIATIONS

- 4.1 **ID:** identification.
- 4.2 **CSV:** spreadsheet file format that is saved as a text file where the individual spreadsheet cell values are listed and separated by commas (i.e., comma separated values).
- 4.3 Specific database fields are indicated in bold below.

## 5. PROCEDURE

- 5.1 During the identification (ID) phase, the specimens of each species are sorted in the NEON supplied spreadsheet template by **genus**, **specificEpithet**, **sex**, and **individualID** number, all arranged in ascending alphabetical or numerical order.
- 5.2 After data entry is completed (i.e., species identifications are assigned to each specimen), the domain spreadsheet is checked to make sure the number of entries matches the number of specimens received (e.g., 642 specimens received has 642 lines of data entered). The spreadsheet is then sorted by **scientificName** (in alphabetical order), **sex** (in alphabetical order), and **individualID** (in ascending numerical order). Columns are expanded for easy visibility. A careful check is made for typos, inconsistencies, entry errors, especially for multiple entries of the same species. Since the entries are now all sorted together, differences or typos stand out.
- 5.3 Species by species, the numbers of entries for males and females (or unknown) are checked against the numbers of actual male and female specimens in the unit trays to make sure they match (e.g., making sure 8 female and 3 male specimens of Species A have 8 female and 3 male entries for Species A on the spreadsheet).
- 5.4 Selected specimens (but this is usually most, and sometimes all) are scanned under the microscope to confirm the identification. For each confirmed identification, the spreadsheet is then checked for the **individualID** number of the selected specimen to make sure the **scientificName** and **sex** match the specimen.
- 5.5 Once confirmed, a date is entered into the **batchID** column, thus indicating that that specimen was examined again.
- 5.6 Once verified and released by the taxonomic identifier, the spreadsheet is saved (in CSV format) and a copy is sent to the CMNH data manager, who uploads it to the NEON data portal. If the data portal flags the file as containing inconsistent data, the data manager resolves the issue(s) with either the taxonomic identifier or with NEON personnel until the file uploads successfully.