

Specimen Data Documentation

Summary

Specimen data can be submitted to AntWeb in two ways: (1) emailed as a spreadsheet or (2) uploaded directly as a tab delimited .txt file. If you manage a growing (active) dataset, it will be easier to learn to upload your data directly to Antweb. If you have completed a dataset for a specific project (e.g. revision) that you want to include in Antweb, you can email the data to Antweb for uploading.

In order to upload your data directly to AntWeb, you must be sure to name the headers on your spreadsheet exactly as indicated in the template. If you email your data to CAS, we will ensure your data is complete and suitable before uploading it to AntWeb. Below we provide instructions on how to organize data and submit it to AntWeb via direct uploading and email submission.

Specimen codes (unique object identifier) are required in all types of data submission. Different approaches and case scenarios for using specimen identifiers are discussed below.

The **Specimen data template** Excel file includes the following tabs with data entry instructions and formats:

- (1) data field descriptions
- (2) specimen data template
- (3) date conversion
- (4) Latitude Longitude conversion
- (5) Country Name List
- (6) Bioregion Name List
- (7) indetermined and taxonomic uncertainty
- (8) Museum acronyms
- (9) Method and Microhabitats

Specimen data submission

Ideally, specimens bear the specimen code that is recorded in the spread sheet. For example, at CAS, we add a CASENT number to all pinned and vial specimens and record this number in AntWeb. However, there will be cases where this number is not present on the specimens. In these cases, AntWeb can provide a number (such as a CASENT# or AntWeb#) for you to apply to the specimen. If you have already recorded the data, and can't feasibly add a specimen code label to each specimen, you can add an AntWeb# in the spreadsheet. These records are referred to as collection "observations."

Antweb currently does not accept the following:

Specimen records recorded from literature or field observations that are not backed up with a specimen.

email spreadsheet to CAS

If you have a completed dataset that you no longer plan to update on a regular basis, you can submit the data via email to "Michele Esposito" <MEsposito@calacademy.org> for uploading to AntWeb. We will check over the data to ensure it fits the AntWeb format. If in the future you do update the data associated with your specimens, you can resubmit the data. Resubmission will add new records or update old records but will not delete existing records. If you delete specimen records, you need to let us know which specimen codes to delete.

Before submitting the data to Michele, your data should be formatted with the appropriate AntWeb column headers. If your data lacks locality codes or collection codes, these will be assigned at the time of submission. If we assign specimens locality codes or collection codes, we will send a new version of the spreadsheet back to you with these codes. You will need to use this new version if you plan to update these records in the future.

Direct text file upload

If you maintain an active specimen database in Excel, Biota, or any other database application that can export a tab-delimited file, you could potentially upload directly to AntWeb. However, if you upload directly to AntWeb, you must adhere strictly to column header names.

The advantage of direct upload is that you maintain control of all of your data. If you reupload your data, specimens that have been deleted are automatically deleted (not so with emailing specimen data). We strongly encourage direct file upload if you plan to maintain a growing database of specimen records.

To upload:

Save file as .txt file. This would be easy if Excel did not modify .txt files by adding double quotes to cell values. To get around this, you must select all the values and copy to a text editor such as Notepad (PC) or TexWrangler (Mac) and then save as a .txt file.

Upload .txt file to Antweb.

First Log on and go to Curator Admin Tools:

Specimen Data

Upload Specimen File:
(tab-delimited .txt file)

No file chosen

After uploading, you will get a "Success" screen if the file was successfully uploaded. You will also get a report indicating if any of the names in the data file are not valid

species names with respect to the current version of AntCat. Names in this report should be investigated to make sure you are using the current valid name and are spelling names as indicated in AntCat. Note that when verifying your data, also check your subfamily names.

Curating specimen data

Once you submit specimen data to AntWeb, you may need to go back and update specimen records, identifications, or adjust locality information based on feedback from users of AntWeb. You can receive these comments directly if you provide your email address as a data provider. Otherwise, most people will contact AntWeb with the update.

Creating and using Specimen codes

Specimens identifies (specimen codes) used by AntWeb are registered unique identifiers. If you do not already have at your institution unique identifiers, we can provide you CASNET or ANTWEB identifiers for you use. We can allocate you a set of numbers or even print the numbers on archival paper number for you to use.

Specimen Data FAQ's

- **Is "100w" ok to put in the life stage/sex column? Is that the format you use for multiple individuals/collection event? How do you denote collections that include multiple semaphoronts such as 39w, 2dQ, 4aQ, 5m? What are the acceptable abbreviations?**

You can use abbreviations or full text works for caste such as aQ, dQ, w, m or alate queen, dealate queen, workers, male. You can also include the number of each caste such as for a record for a vial: 100w, 1dq, 3m.

- **Does every specimen have to be identified to species/morpho-species? Although I am coming closer to understanding what is going on with Palauan Pheidole, many of my specimen records are still only identified to "Pheidole".**

The Species name field cannot be left blank.

For taxa that have not been named, even to species code, use "(indet)". Thus instead of just "Camponotus" or "Camponotus sp." use Camponotus (indet).

- **Can I update the taxonomy later (e.g., what if I discover that Pheidole jcm01 is really Pheidole umbonata? What if I discover that there are several cryptic species within what I was calling "Pheidole sexspinosa"?) Will I be able to update the specimen records by uploading a new spreadsheet?**

One advantage of a specimen-level database is that you can update specimen identifications. You can re-upload or email your data as many times as you want

to change the identity of specimens. Specimens may go from (indet), to morphospecies code, to named species, and each upload or submission will reflect these changes to your data. If you do email your specimens to AntWeb, note that if you delete specimen records, you must indicate in the email which specimen codes should be deleted.

- **What strategy should be used when multiple users are uploading data linked to the same collections/specimens? For example, if I upload new *Stenamma* determinations and then my colleague, as result of how he imports data to AntWeb, uploads his older determinations, will mine be overwritten?**

We want to avoid having two separate data providers manage the same specimen data. However, this situation is bound to occur eventually. A taxonomist may visit collections from multiple institutions when working on a revision of a genus, and then submit the data to Antweb. Meanwhile, institutions may be submitting data from their own collections. Currently, only the most recent submission of a specimen record is shown on AntWeb.

One way to avoid this is to submit corrections and updates to institutions that house the specimens.

- **Is there a way to be clear about the source and precision of GPS coordinates? By source, I mean whether the data come from a gazetteer, a GPS unit, or Google Earth. For precision, is it possible to indicate the number of sig figs for every entry? A problem of the current system is that if you convert latitude/longitude information from degrees and minutes to decimal degrees, you often end up with coordinates that are too precise. AntWeb should make clear to users how precise the data are.**

LocalityNotes is the best field available to put information about the accuracy and precision of coordinate information.

- **Is there a convention for stage/sex data? Many of you use a one specimen/pin system, but when there are multiple specimens/pin, as is common on older material, how should one indicate this? 1dq, 2w?**

In the **LifeStageSex** field, you can include both caste and number of individuals. However, this practice causes confusion when images of a single specimen are recorded yet the LifeStageSex field indicates more than one specimen. In the future we hope to include a field to indicate the caste of the imaged specimen.