

How to report soil macrofauna data

contact : jerome.mathieu@upmc.fr

Frist of all, thanks a lot for your time and effort. This is very valuable for the project!

The global goal of the database is to be able to compute the density of the major soil macrofauna groups in a variety of locations – situations. Density means here the number of individuals of the group by a unit of surface measured on a short time. Taxa should be either in the soil or in the litter.

As a general rule, please fill template with data with as much detail as possible. We will recode afterward your inputs. So for example you can describe land cover and land management in your own words, in a verbatim manner, with sentences for instance. Don't worry about space. We will recode afterwards this information to match it with data information services. Another example: if you have data at sampling unit level, please use them instead of data aggregated, at for instance, the plot level.

Please feel free to send pictures of your samples, transects plots, specimens and so on, we are interested by them!

If you have data at finer taxonomical resolution or with traits measured on individuals, please let us know, we are compiling a separate database for this type of info.

If you have both data on density (nb of individuals) and biomass (individual biomass or total by group, or total all groups together) please add them in the 1_macrofauna sheet by adding necessary columns (see template for details).

As a general rule, please feel free to add any information that is not in the template, but that you judge important.

If possible, please send the data along with the published paper associated to it. This is important for repeatability and sanitary checks of data quality.

Thanks again and good luck!

Missing data vs no individuals

A very important point is that we need to be able to distinguish "0 abundances" from "no data":

- 0 individuals, or "absence", means that the group was in the scope of the sampling (ie was sampled and reported when present), but no individuals were found in a particular sample (or transect). 0 data can be indicated by "nothing" (=empty cell), or 0.
- "No data" means that the faunal group was not surveyed in the datasets. Punctual No data can be identified by "NA", for example if you lost the data of a specific group in a specific sample, but overall, it is better to delete the columns for which you have no data for the taxonomic groups. **Missing data of abundance should not be coded by empty cells.** In other

columns, you can code missing data with empty cells, but again, not in the abundance (or biomass) data.

- Particular case: if you are 100% that a group was not present, but was not reported in the dataset (because it was indeed absent), you can leave the corresponding column, with blank cells, instead of deleting it. This can be the case for termites in particular, which are extremely rare in temperate areas.

Comments specific to the sheets of the template

• Sheet "0_data_set_info"

This sheet summarizes the context of the dataset and the terms of use. Please fill in carefully this sheet until the end.

Please note that by default we plan to release the global data base on open access model. If you do not want your data to be released, please advise us. This can be dealt with.

Please also check with your eventual co-workers that they agree on this term and also on authorship.

Co authors : if the data you provide were acquired with the help of other people who asked to be listed as a co author, please indicate it. We will use an inclusive authorship system for data with more than three sites (three transects). Though, co authors are asked to participate to the elaboration of the manuscripts.

• 1_macrofauna

The column "other_groups" in the data sheet refers to the groups that were observed but not reported individually in previous columns.

Please indicate the list of these other taxa in the sheet 0 (at the bottom).

Taxonomy

Please provide the data at the finest taxonomical resolution as possible and add a line above with the groups defined in the template. We will harmonize taxonomy afterwards. We use GBIF backbone classification system by default.

We'll take the data at the best grain you have (best resolution) and will aggregate them ourselves in a second step. This should reduce your workload.

Please provide the density of individuals in original format: if it is your data, please provide the raw observed number of individuals per sample. For legacy data from literature, please use the numbers from the publication. In all case, don't forget to explain the unit used, on the "0_data_set_info sheet"

Very important: please indicate the geometric dimension of the samples size (size of the soil monolith and depth). This information is necessary to compute the density of the group per unit of surface.

- **2_lu_sampling**

Please be as much complete and accurate in the lu_cover column. We will recode this info afterwards. Geographic coordinates should be in decimal degrees (not hexadecimal : ° ' "), with WGS84 geodetic system (in doubt, use the google maps coordinates).

- **3_soil**

Please explain how you defined the depth: is "0" starting on top of humus or at the transition from humus to mineral matrix.

For soil properties, don't forget to describe the depth of the measurement, this is an important point!

Thanks again and looking forward hearing from you!

The Soil MACROFAUNA team