



Reporting data to the #GlobalSoilMacrofauna database v.2.1

contact: jerome.mathieu@upmc.fr

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

The goal of the database is to get estimates of the density and biomass of the major soil macrofauna groups in a variety of situations. Density means here the number of individuals of the taxa by a unit of surface measured during one sampling occasion. We only accept data at the monolith or layer level.

As a general rule, please fill the template with as much as possible details. 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 fit it in a controlled vocabulary.

Please feel free to add any information that is not in the template, which you judge important.

If possible, please send any published paper associated to the dataset. This is important for repeatability and data quality checks.

When possible, please send pictures (CC-BY-4 licence) of each transect, monoliths or soil profiles. Please name the file in a way that we can match them to the unit they are representing. For instance:



Reporting abundances

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

- If a taxa would have been reported if it was found, but it was not present at all, then you should fill the abundance column with 0. Do not leave empty cells. This way we will know that the taxa was absent (instead of "not reported").
- Columns with only empty cells mean "No data": the taxa was not reported, even if it was present.
- You need to do this for all taxa that are mandatory (indicated in the explanation sheet of the template). This way we have a complete view of the community.
- Preferred units are ind.m⁻² and mg.m⁻² dry weight.

Reporting depth

Please indicate the top and bottom depth of the layer, in separate columns. If litter is covered, please indicate it with "L". If this info is not filled, we cannot use your data.



Hierarchical organization of the samples and data

Samples and data are organized in a hierarchical manner, and have an ID at each level:

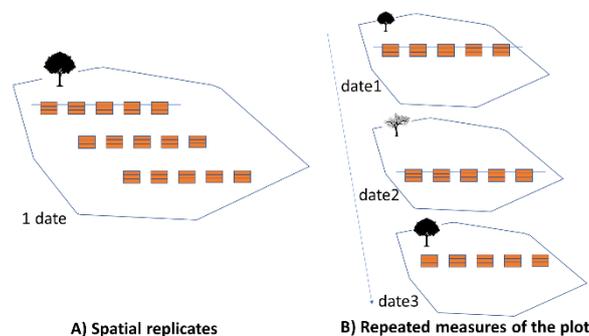
- Layers belong a monolith (they have the same monolith ID)
- Monoliths belong to a transect (they have the same transect ID)

Available data	Illustration	IDs required
Density at each monolith		transect_ID monolith_ID
Density at each layer of each monolith		transect_ID monolith_ID layer_ID

Illustration of the IDs required as a function of the the available data

- Transects are part of a plot only when
 - A) There are several transects in the same plot, at different locations (spatial replicates).
 - B) The same plot was sampled several times (repeated measures) and you have the data for each sampling date.

In such case your data need a plot ID, and all transects from the same plot have the same plot ID. The column 'repeated' in the template indicates temporal replicate.



⇒ *There is one column ID for each level. You need to use the ones that are relevant to your data.*

Clarifications

- A plot is a physical spatial area, such as a crop, a patch of forest, which is homogenous. A plot is typically 1ha large.
- A transect is a sampling event in space and time: a group of samples taken near each other at the same date. Large transect such as elevational gradient, or distance to the sea, are not considered transects here. In such case, each distance class constitutes a transect, made of one or more monoliths.



How to construct samples' IDs

- **The general structure of the IDs is:**

'(plot_ID)_transect_ID_monolith_ID_layer_ID'

Remember:

- **plot_ID** is only necessary in case of repeated measures or spatial replicates.
- **Monolith and layer IDs** are only necessary when you have data at these resolutions.
- When there is only one monolith per transect, there should be a transect and monolith ID.
- **transect_ID** is mandatory.

- **Convention**

In the IDs, you can use '_pXX' for plot, '_tXX' for transect, '_mXX' for monolith, '_sXX' or '_lXX' for layer and litter respectively

- **Examples (one row = 1 record)**

Study 1: just 1 transect in a forest plot, with only 1 sample, with three layers (litter and two soil layers)

transect_ID	monolith_ID	layer_ID
t1	t1_m1	t1_m1_L
t1	t1_m1	t1_m1_s1
t1	t1_m1	t1_m1_s2

Study 2: 1 Sampling in two transects, e.g. one in a forest, one in a crop, only one monolith per transect, no layers

transect_ID	Monoliths_ID
t1	t1_m1
t2	t2_m1

Study 3: Repeated sampling within the same plot, 2 monoliths per transect, 2 dates, no layers

Plot_ID	transect_ID	Monoliths_ID
p1	p1_t1	p1_t1_m1
p1	p1_t1	p1_t1_m2
p1	p1_t2	p1_t2_m1
p1	p1_t2	p1_t2_m2



Study 4: Repeated measures: 8 plots were sampled two times, but with different cover each time, 1 or 2 monoliths per transect (= per sampling event), no layer data

plot_ID	transect_ID	Monoliths_ID	lu_cover
p1	p1_t1	p1_t1_m1	maze 2020 glyphosate
p1	p1_t1	p1_t1_m2	maze 2020 glyphosate
p2	p2_t1	p2_t1_m1	maze 2020 glyphosate
p3	p3_t1	p3_t1_m1	maze 2020
p4	p4_t1	p4_t1_m1	maze 2020
p5	p5_t1	p5_t1_m1	wheat 2020 glyphosate
p6	p6_t1	p6_t1_m1	wheat 2020 glyphosate
p7	p7_t1	p7_t1_m1	wheat 2020
p8	p8_t1	p8_t1_m1	wheat 2020
p1	p1_t2	p1_t2_m1	corn 2021 glyphosate
p2	p2_t2	p2_t2_m1	corn 2021 glyphosate
p3	p3_t2	p3_t2_m1	corn 2021 glyphosate
p4	p4_t2	p4_t2_m1	corn 2021 glyphosate
p5	p5_t2	p5_t2_m1	fallow 2021 glyphosate
p6	p6_t2	p6_t2_m1	fallow 2021 glyphosate
p7	p7_t2	p7_t2_m1	fallow 2021
p8	p8_t2	p8_t2_m1	fallow 2021

Here plot 1 (_p1) was under maze the first year ('_t1'), but corn the second year ('_t2'), and was sampled with two monoliths first year, with one sample the second year.