

Appendix S2: Graphical User Interface

On the left side of the screen there is a panel with a menu containing nine options organized to represent a typical workflow and that allow the user to: (1) introduce the general information of a given study; (2) create new systems, factors, levels or scales; (3) specify the spatial design of the study (see “Database description”); (4) define new variable and data types; (5) add new dependent (i.e. response) variables; (6) insert data; (7) delete a given insertion event (i.e. rollback a successful insertion of data -6-); (8) delete all data associated with a given study -1-; (9) visualize the available data. The main panel varies depending on the option selected, and most of the fields they contain are usually autocompleted based on available data. Note that the “submit” button is not enabled until all mandatory fields are filled and there are no inconsistencies (e.g. the same column selected in two different fields; text inserted where numbers are expected).

The first introduction of data necessarily requires going through options 1-6 of the lateral panel. Since then, only options 1 and 6 are mandatory when introducing another study, although it is quite likely that 2-5 will also be needed. Thus, it is highly recommended to carefully and sequentially check all options (and fields) when inserting a new study.

We hope that all the fields throughout the different options of the main menu (lateral panel) are self-explanatory. Thus, below we briefly present the features of each option of the lateral menu, focusing only on the main fields that might require special attention. In “New study” (option 1), one should define the user responsible for the study. Given that the database is not designed to grant specific privileges to users, this field was included only to simplify management (i.e. the first filter applied to stored data, defining which insertions and deletions are associated with that user). The other essential field in this option is Study details; here the user is expected to include a detailed explanation of the study design, sampling and/or experimental procedure, etc.

In “New system, factor, level or scale” (option 2) the user can scan previously inserted systems (e.g. tundra, savannah, tropical forest, coral reef), factors (e.g. N addition, habitat fragmentation), levels (e.g. control, 10 g m⁻²; natural, fragmented) or scales (e.g. Block, Plot, SubPlot, SubSubPlot, Quadrat). If the required value is not registered, then the user should add it exactly as it will be named in the spreadsheet. The user should be very careful here, adding different names for qualitatively similar values will reduce the output length in future queries regarding that value. Analogously, merging different values by using the same name could lead to inconsistent results. Providing different names from those used in the spreadsheet will inevitably lead to a failure during the process of data insertion.

In “New scale for a given study” (option 3) the user is expected to upload a spreadsheet containing at least the information of the spatial design of the study being inserted. Here and elsewhere, uploaded files can be in Excel, Libre/OpenOffice native formats or csv (comma separated values), and the user can select which sheet will be used. Although not mandatory, the user can specify the spatial coordinates for each largest scale (remember that it is named “Block” but can refer to a plot, individual, region, etc.). If supplied, once data is inserted (option 6), the location of the blocks will be shown in the map (option 9) along with their data.

The “New variable & data type” option is very similar to the “New system, factor, level or scale” option. There is a summary of inserted values, and it is possible to add new if the ones inserted are not useful. Data type refers to a term that groups data according to the user's needs. For example, a typical ecological database will likely store species x site and environmental factors x site matrices (species and environmental factors in columns and sites/replicates in rows). Then, a useful way to group data could be into chemical, physical (for environmental variables) and kingdoms (for species). Similarly, dependent variables can also be grouped according to what their values represent (i.e. variable type; e.g. biomass, cover, density, concentration, count). For example, we can be intending to insert a species abundance x site matrix where some columns belong to plants (g; i.e. biomass) and some others to arthropods (counts). Moreover, the same species (data) can be added twice, one as variable type = biomass and the other as variable type = cover.

The “New dependent variable” option intends to simplify the insertion of all dependent variable names from a given study. The spreadsheet can contain data stored in the wide format (one row per replicate/sampling unit) or in the long format (only one column with numeric values per variable type, and dependent variable

names in rows). Given that ecological studies dealing with species usually have a large number of dependent variables (species), the process of inserting these dependent variables can be simplified by uploading a supporting taxa file (see previous section).

“Add data” (option 6) is likely the key feature of this GUI. As a first step, the user should upload a spreadsheet and select whether it is in the wide or long format, as described above. This will enable some fields that will be prefilled but that the user should carefully check before submission to prevent erroneous insertions. Obviously, options 2-5 can be repeated as many times as needed per study to be able to include all the necessary values before inserting data.

The following two options (“Delete insertion event” and “Delete study”) are quite simple. The first step is to select a certain user responsible for a given study, and then select the desired study’s name. In the case of deleting insertions, the last 10 insertion events will be visible and possible to rollback (previous insertions can be accessed using MySQL Workbench). An insertion event is any of the times when data was successfully submitted in the “Add data” option. The deletion of an insertion event will roll back all data submitted in that process. It is very important to be aware that a successful deletion of an insertion event or a study can not be undone. In other words, once a study is deleted, all its associated data will be permanently lost.

“Visualize data” (the second most important option), is where any user can see all the available data. The user should first select which variable types he/she would like to see, and then, one or more studies among those that include the selected variable type. At last, the user can specify the range of years to include in the visualization. Upon these selections, a map will show the location of all largest scale units that were registered with GPS data. In addition, the main panel will show three different boxes: Information, Summary and Data. The first will show the details of the selected studies, the second will show a summary of available years (range), as well as the number of cases per factor/level. The last box shows the available data, and at the bottom of the table there are two buttons to download these data in xls or csv formats.

Finally, the first and least important screen, the welcome screen. This screen only appears immediately after starting the GUI, and it disappears as soon after the user selects any of the options in the lateral panel. The purpose of this welcome screen is to briefly summarize the amount of data stored on the database.