Uploading metadata into Metatree

Introduction

In Metatree data can be entered manually adding each entry by hand. Figure 1 shows an example of how details of an object can be viewed and or updated.
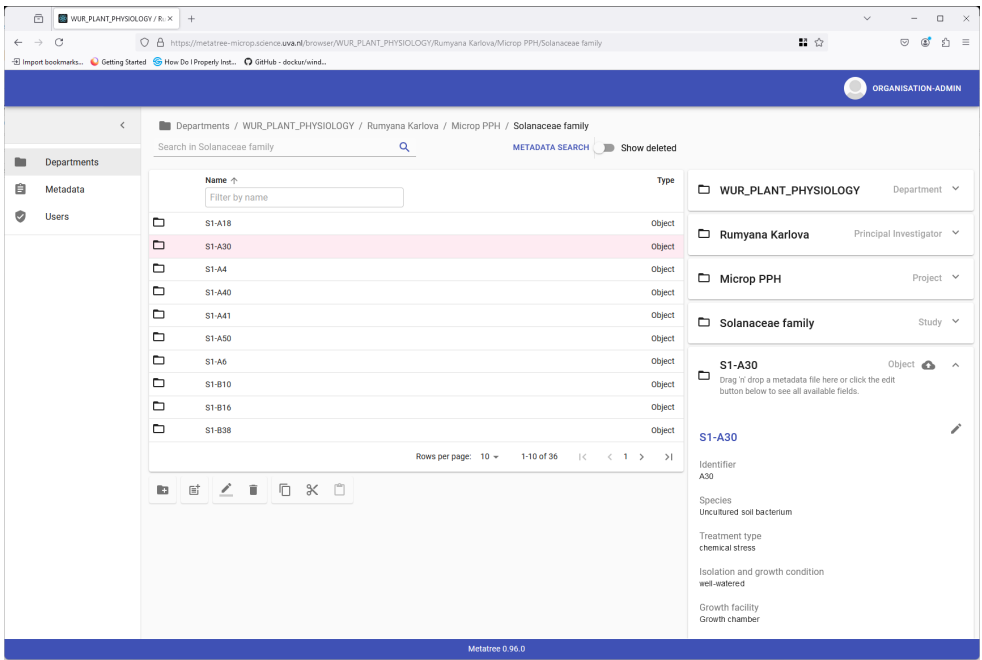

Figure 1, Metatree per-entry reviewing/updating

In figure 2 the treatment type is updated with a value defined by the ontology.
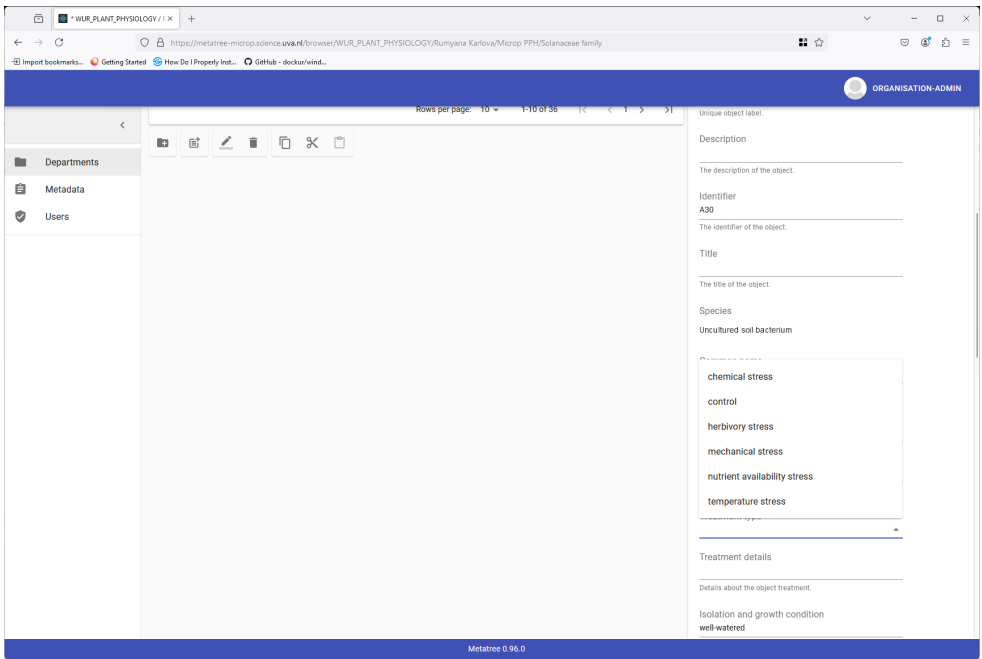

Figure 2, Metatree, select from predefined set of values.

The manual process of uploading is cumbersome and time consuming. It's more convenient to upload the data in bulk. An import routine was implemented for uploading Excel contents to do just that. The import format, however, must adhere to some standards which are discussed below.

Excel format

The hierarchy of Metatree is as follows:

1) Department
    a. Principal Investigator
        i. Study
            1. Project
                a. Object
                    i. Sample
                        1. Assay
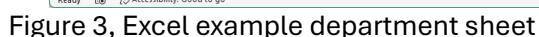   a. (External) File

At every level the entity can be annotated by fields specific to that entity. For example, at object level the species can be defined while at assay level the type of sample that was measured is specified. **Only by linking an external file to an assay the entry will be listed as a metadata result.**

Within every hierarchy level the identifiers must be unique. This way the system does not impose restrictions on naming samples for example allowing different researchers to use the same sample codes for different samples. By including the upper identifiers, the samples can still be uniquely identified throughout the system.

To upload the data in bulk, the Excel has different sheets, one for each hierarchy level with a corresponding name. If there is a typo in the sheet name, an error will be generated. In every sheet the different entities (rows) can be defined with their values for the specific properties (attributes).

In figure 3 the **department** details for plant physiology in Wageningen is shown/defined. Every can contain an entry and every column points to a property of that entry. It's not necessary that each attribute has a value, but for meaningful queries later, it is advised to enter as much as possible. **Where in all subsequent sheets (hierarchy levels) the researcher/uploader is free to choose the Label value, in this case the Label value must exist in the Metatree database and is created by the administrator of the database.**

Every user has an access level that is restricted by this Department (value). This prevents overwriting of values by different users.

Figure 3, Excel example department sheet

In figure 4 the details for **principal investigator** are shown. Note that there is a department column that points to the department label defined in the previous sheet. Whenever an Excel file is uploaded, the contents of the file will be first processed to see if all entries all properly linked to each other. This means for example that if there is a typing error in the department reference, an error will be generated, and the data will not be uploaded. This process of validation is cascaded throughout every hierarchy level ensuring that entries are uploaded correctly.

Figure 4, Excel example of Principal Investigator sheet

The process of including the previous fields for a unique mapping is again visible in figure 5, where the different **studies** are defined. Every row has a reference to the previous sheets by defining the **Department**, **Principal Investigator** and **Project**.

Figure 5, Excel example of Study sheet

By specifying the upper levels on every sheet, every sheet can contain multiple entries that belong to different values of these upper levels. This way, a researcher only needs a single Excel file that contains all metadata of his/her samples of multiple studies, projects etc.. Furthermore, by updating some values or adding some entries the Excel file can be re-used for upload and the database entries will be updated accordingly and/or new entries will be added.

Specific fields/Ontologies

Some property values are not free to choose and are defined in some ontology. Where and when possible, public ontologies are used but in cases the ontology itself is not mature enough so own ontologies are used/defined. There are ontologies defined for: AssayType, GrowthFacility, LCMSApparatus, LCMSApparatusType, PlantDevelopmentalStage, PlantGrowthMedium, PlantHealthState, PlantStructure, Ploidy, SequencePlatform, SequenceTechnology, Species and Treatment.

The values for all these ontologies are by no means complete and can be extended.

In figure 6, an example is shown of the project data sheet in which the organisms and treatment types are defined. Both follow an ontology and as such, these values cannot be freely chosen (but are case insensitive). When the sheet is uploaded these values are checked and if they do not correspond an error will be thrown indicating what was not properly defined on which sheet.

At project level the organism field can hold more than one entry which should be separated by a pipeline symbol ("**|**").
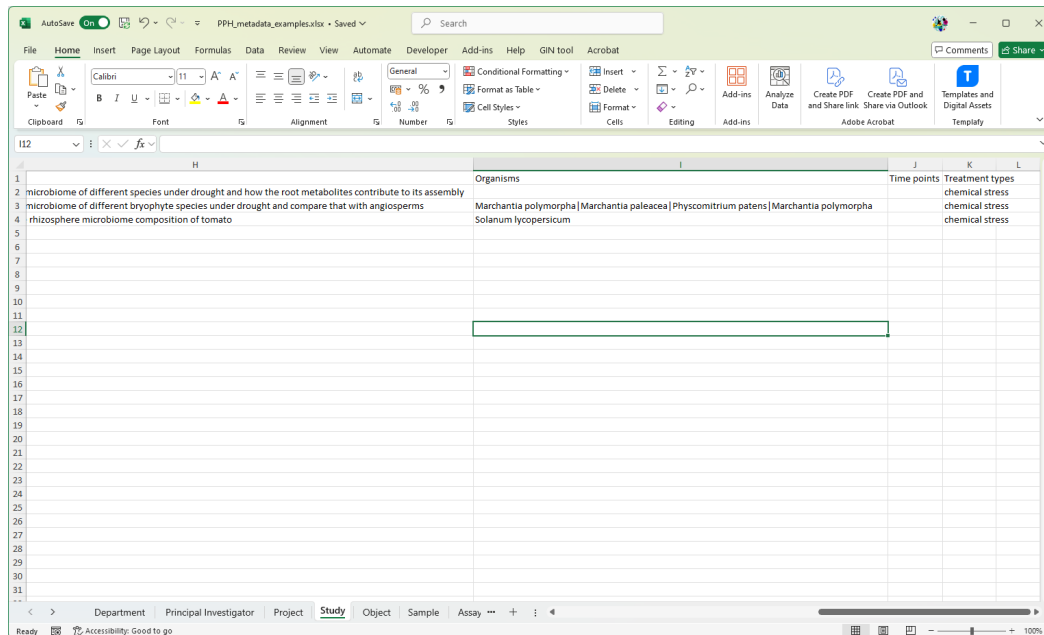

Figure 6, Study sheet with different ontology items

There's no comprehensive list for every ontology what are valid values but the Metatree application itself will show its options when selecting the specific field, see figure 2.
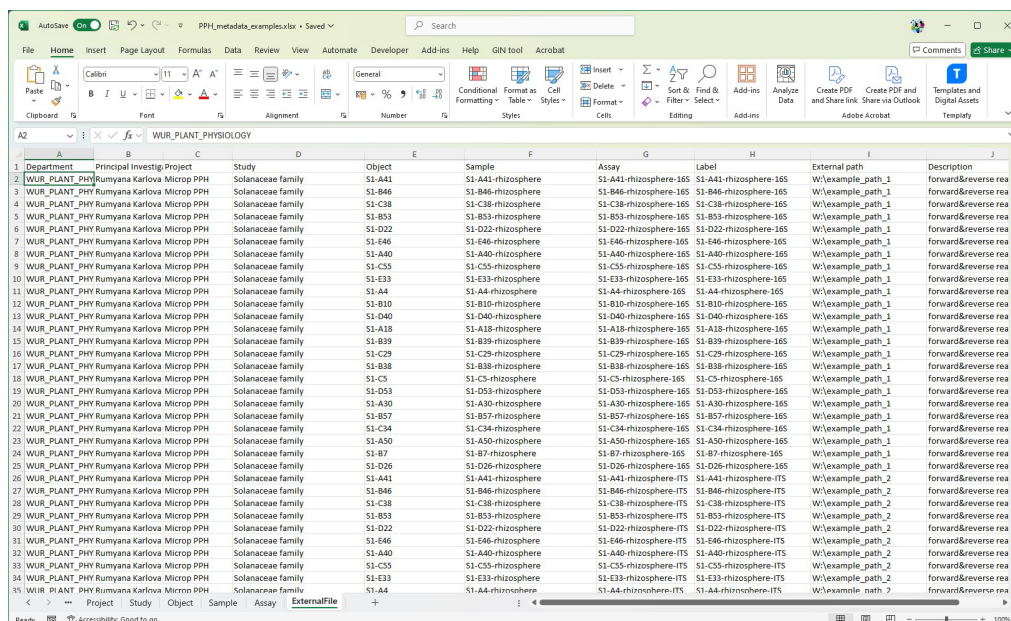

Figure 7, The ExternalFile sheet

As mentioned earlier, **only** if the assay is connected to an external file as shown in figure 7 the entries will be included in the metadata search results. The entries here are therefore of significance. The external path, however, can be any path but logically points to a location where the data actually is stored. This does not mean that it needs to be accessible to everybody. The Metatree database only contains meta-data and not the actual data itself.

Uploading the data

The actual upload of the data can be done by selecting the document+ icon (second from the left, see figure 8). A dialog box will open prompting you to select an Excel file to upload. Remember not to have this Excel file open otherwise an error will pop-up. The supported format for upload is **.xlsx**. The xlsx files generated with **LibreOffice** also work.
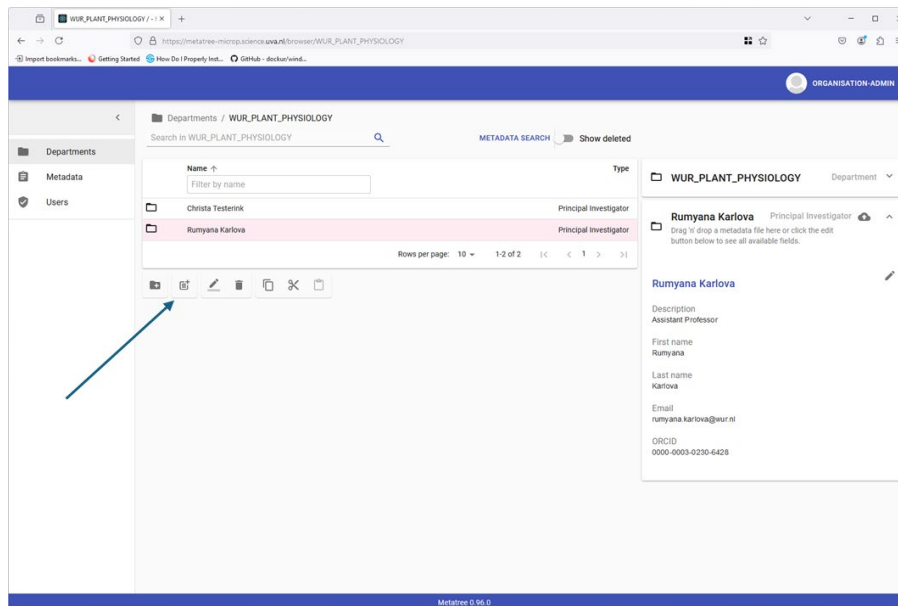

Figure 8, Choose the button second from the left to upload the Excel file.

After a while of validating and processing a prompt like in Figure 9 should pop-up or indeed an error message telling you what went wrong.
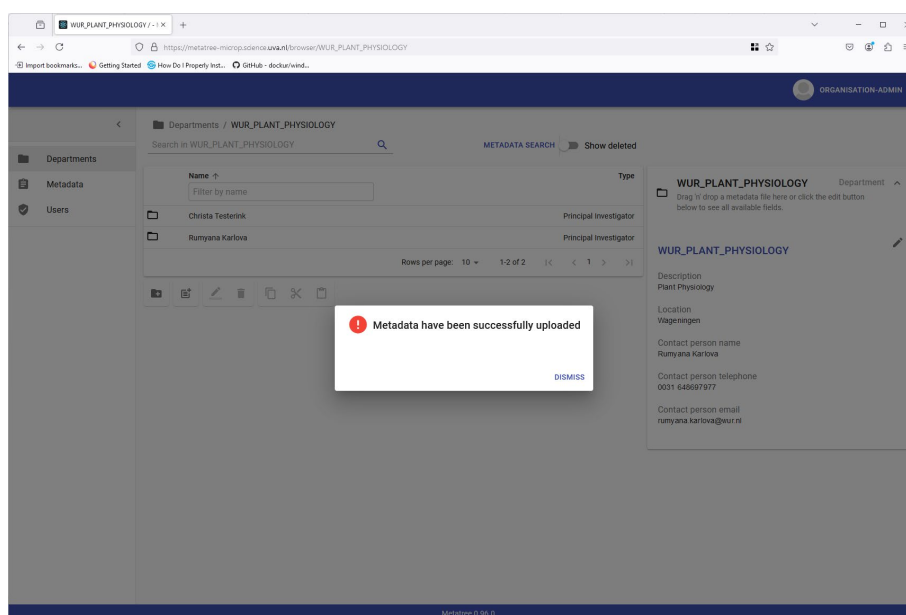

Figure 9, Successful upload of metadata