### Overview

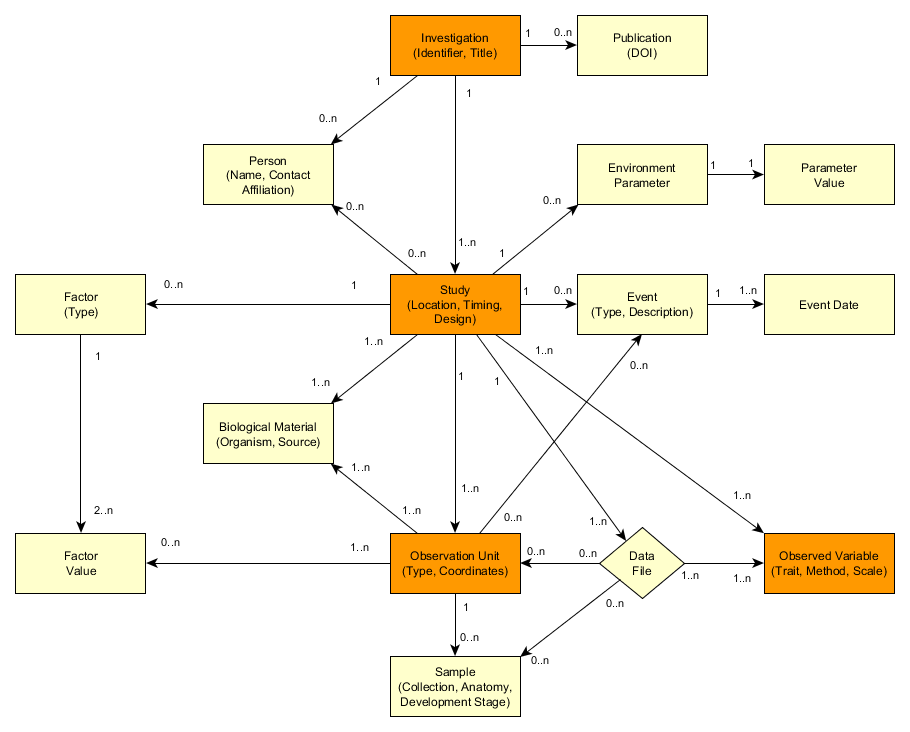
The present revision of MIAPPE, carried out by ELIXIR's plant use case, aimed at achieving four main goals:

1. The extension of MIAPPE’s scope, namely to accommodate woody plants as an additional use-case.
2. The specification of a data model for MIAPPE, to facilitate its implementation in various formats and enable its automatic validation.
3. The interoperability between MIAPPE and related external resources, namely ISA-Tools and the Breeding API.
4. The enrichment of MIAPPE with clear definitions and examples for all fields, to improve its accessibility to users.

In this document, we review MIAPPE’s data model specification and explain the changes motivated by the first three goals (those motivated by the last goal are self-explanatory).

Overall, we believe that this MIAPPE revision will enhance its accessibility and usability, leading to its adoption by a wider community.

### MIAPPE Data Model



**Figure 1** – Schematic representation of MIAPPE’s data model specification.

MIAPPE’s data model specification, illustrated in Figure 1, was heavily influenced by the ISA data model, as interoperability with the ISA-Tools was one of our main goals. Of its four central objects (represented in orange in the schema), three have a direct correspondence with ISA: Investigation and Study are homonymous in both models, and Observed Variable corresponds to ISA Assay. The fourth, Observation Unit, is a central object in the data model of the Breeding API, and together with Observed Variable, is key to enable interoperability with the latter.

The objects Investigation and Study merit further attention as they were not explicitly defined in MIAPPE 1.0. Only a section devoted to general metadata was included therein, with a reference to ISA-Tab, but with no formal separation between these two objects. If we want to support ISA-Tab as a submission format, then Investigation and Study must be explicit objects in the MIAPPE specification or we risk users not understanding the separation between them and/or leaving one of them unfilled. The proposed data model is that publication and submission metadata be listed at the Investigation level only, and that links to data files, timing and location data be listed at the Study level only, with the restriction that a Study must have a single location.

### Explanation of Proposed Changes

#### The changes made in the proposed version of MIAPPE can be roughly divided into four categories: terminological changes, organisational changes, scope extensions, and modelling simplifications.

#### Terminological Changes

#### These changes were aimed at making field labels less ambiguous and/or promoting interoperability with external resources. They were the following:

1. Renaming of *Biosource* to *Biological Material* which was deemed more intuitive
2. Renaming of *Treatment* to *Factor*  
   This matches ISA, which uses *Factor* rather than *Treatment* to refer to this concept. The Breeding API uses *Treatment* to refer to the same generic concept, but each *Treatment* is described by a *Factor* and a *Modality*, which now correspond to MIAPPE *Factor* *Type* and *Factor Value* respectively. Thus, this change promotes interoperability with both ISA and the Breeding API. Furthermore, it avoids the ambiguity of the name “treatment”, which may be used in a broader sense to refer to any cultivation practices. A *Factor* is expected to be a controlled variable, the effect of which is the object of the study.

#### Organisational Changes

#### These changes were motivated by the need to convey the data model intuitively, even at the checklist level. For this to be possible, each section in the MIAPPE checklist should correspond to data model object or pair of objects (in the case of paired objects like *Environment Parameter* and *Parameter Value*). Conversely, each data model object for which we need to list multiple attributes should have its own section in the MIAPPE checklist. The changes falling under this category were the following:

1. Creation of new sections: *Investigation*, *Study*, and *Observation Unit*
2. Abolishment of old sections with no representation in the data model, and redistribution of their metadata fields into the new sections:
   1. *General Metadata* – fields split between *Investigation* and *Study*; some duplicated in both (e.g., *Title* and *Description*)
   2. *Timing and Location* – fields moved to *Study* section
   3. *Experimental Design* – fields split between *Study* and *Observation Unit*
3. Grouping unique and mandatory *Environment* fields under the *Study*, namely: *Type of growth facility* and *Cultural practices*

#### Scope Extensions

#### These changes were motivated by the need to accommodate the woody plant use-case or capture more types of information. They were the following:

1. Addition of a *License* field under Investigation for FAIR compliance
2. Addition of *Person* section to describe parties involved in the investigation or any of its studies, including information about their roles and affiliations.
3. Addition of fields for listing *Data Files* and providing their version at the *Study* level  
   Data files were not listed in MIAPPE 1.0 and the version is necessary for FAIR compliance.
4. Extension of the *Biological Material* section to accommodate woody plants:
   1. Addition of geographical coordinates as a means of uniquely identifying woody plants in the field, instead of accession.
   2. Renaming of *Seed Source* to *Material Source* due to the fact that woody plants (and not only) may not be necessarily propagated by means of seeds.
   3. Replacement of the fields *Pretreatments* and *Conservation conditions*, which pertained to *Seed preparation*, by the more generic field *Biological material preprocessing* which englobes both aspects (preprocessing was chosen instead of pretreatment to avoid the ambiguity of the name “treatment”).
5. Addition of *Event* section to describe discrete occurrences that happened during the experiment, and that must be associated with a date. These can be natural or artificial interventions, can encompass elements of planning (e.g. planting) including parts of *Factors* (e.g. watering), or sporadic events (e.g. rain). They may apply globally, to the whole study, or locally, to specific observation units.
6. Generalisation of the *Observation Unit* description  
   Since the layout of experiments may vary greatly, users should be able to define their layout instead of having to provide specific levels. This was enabled by the paired fields *Spatial distribution: type* and *Spatial distribution: value*, where users can provide specific coordinates under the logic of key-value pairs (e.g. block: 5; plot: 1; latitude: +2.341). In combination with the *Observation unit levels* field under *Study*, this enables the description of hierarchies.

#### Modelling Simplifications

#### These changes were aimed at simplifying the checklist by avoiding exhaustive lists of categories for aspects such as *Factors* or *Environment* parameters. Given the broad scope of MIAPPE, the categories that make sense in one experimental setting may not make sense in another, so it is impossible to enforce mandatory fields of these types. Furthermore, it would be virtually impossible to be fully exhaustive, and thus it is best to give the users some flexibility. As such, we opted for a streamlined common representation under which any category from an exhaustive list can be provided by the users when adequate in their particular experiments. The aspects for which this type of simplification was made were:

1. *Factor*, which is described using three fields: *Factor type* (or name), a *Factor description* (elaborating on the specific applications and details in free text), and *Factor values* (a list of all different modalities for this specific factor). The list of possible factor types is to be supplied as an appendix to the MIAPPE checklist.
2. *Environment*, which is described using two fields: *Environment parameter name* and *Environment parameter value*. This section supports the description of only environmental conditions that were constant throughout the *study*. Environment parameters that were actively measured during the experiment should be recorded as observation variables. This is an additional reason why an exhaustive list of environment parameters would be inadequate. Nevertheless, as in the case of Factors, the exhaustive list of environment parameters is to be supplied as an appendix to the MIAPPE checklist, in order to guide users.
3. *Samples*, in which a number of fields describing sample conservation conditions (e.g. salinity, oxygenation, storage temperature) have been condensed into a single, free-text field, *Sample description*.