MIAPPE v1.1

##### [IsaTab levels]: the IsaTab levels are indicated for each section in brackets.

Ontly the initals of the commenters have been kept ot be in conformance with RGPD before publication on github.

HP: What you and I aimed for in Poorter et al. (2012) was a succinct report in the M&M or results section of a paper. What I guess we are only talking about now – correct me if I am wrong - is a digital file with meta-data that could be added separately to a paper.

Answer: MIAPPE is indeed targeting something both more detailed than a Material and Method Report and more extendable. Indeed, MIAPPE provides the structure, and the list of recommended variables (Rooting and Air environment for instance) can be integrated in this structure. The existing recommended variables have been stored in the appendix and can be extended or transformed into ontologies; therefore the appendix serves the Poorter et al 2012 and MIAPPE v1.0 original purpose of listing recommended variables in each situation.

HP: hourly data would be very useful, in doing meta-analyses of 100s of experiments, I appreciate if people would have recorded higher level data, so that I do not have to redo all these calculations.

Answer: hourly data are recommended. Average and computed data can of course be published and integrated in MIAPPE, the choice of the right higher level data is up to each data manager/publisher.

PK-HCW: MIAPPE should be a high level reference list or checklist with implementation as ISA-Tab, BrAPI or free text.

Answer: Checklist is important to underline the point that must be included in a MIAPPE compliant dataset description. But the checklist without a general model independent of specific implementations is becoming difficult to describe. All description made to biologists or computer scientists quickly involved the use of an implicit model. We need therefore to be able to keep both the positive aspect of the checklist and the clarity of the model. The current proposal adress this with both a general checklist with section mappable to a descriptive model and appendix wich propose open checklist for environment, treatment and factors.

RO: low throughput long-term (100y+) agricultural experiments.

Answer: New range of needs, very interesting and challenging. It needs to be carefully evaluated for crop rotation, treatment change, history, …. It will require too much effort for this version but will be handled in future version. See github <https://github.com/MIAPPE/MIAPPE-checklist/issues/4>

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **MIAPPE Check list** | **Definition** | **Example** | **Ontology/data type** |  |
| **General metadata [Study; Investigation]** | | | |  |
| **Unique identifier¹** | ML: The study meta data elements are perfectly overlapping to the DataCite meta data schema <https://schema.datacite.org/meta/kernel-4.1/doc/DataCite-MetadataKernel_v4.1.pdf>. The DCMS is required if a DOI for a data set is assigned and is therefore established as standard in the field and widely disseminated. Because it is advisable to use DOI as persistent unique identifier for data sets, I strongly encourage to tightly link this specification. Doing so, we would be able to seamless convert MIAPPE study meta data into DCMS that support the “findable” aspect of FAIR. | Answer: This was indeed done on purpose with the data set DOI idea in mind. |  |  |
| **Submission date\*** | RR: xsd:date is supposed to be: []CCYY-MM-DD[Z|(+|)hh:mm] So the example should read 2006-02-26 <http://books.xmlschemata.org/relaxng/ch19-77041.html> | All date types have been replaced with Date/Time type, ie timestamps |  |  |
| **Title\*** |  |  | GK: start of study – this refers to a ‘study’ yet my understanding was that the unit was an ‘investigation’. The ‘investigation’ may include several ‘studies’ or not ? | Answer: The start date is adressed below. One investigation can have multiple studies. This have been clarified. |
| **Description¹** | PK-HCW: One thing that could be done is dividing this section into two subsections: a) containing the "information about the **experiment**" that must be provided by the person preparing the data (title, description, publications, contact, ORCID), and b) containing "information about the data file / dataset" that are meaningful only if the data is submitted to a repository (ID, dates). This may save many questions from biologists. | Answer: The current reorganisation seems to go in the proposed direction with the isolation of the data files section. The dates are now not mandatory but are important for a lot of datasets. |  |  |
| **Public Release date\*** |  |  |  |  |
| **Associated publication/s** |  |  |  |  |
| **Name and address of the laboratory\*** |  |  |  |  |
| **Data submitter contact (email)** |  |  |  |  |
| **Data submitter identifier** |  |  |  |  |
| **Study data file link** | CU: This looks great but I have a simple question. In which box do I put the actual value of my variable. Say my plant height is 10 cm... where do I put the value 10 within this framework? | Answer: Values goes in the data files and are therefore not part of the standard. | PK-HCW: "Study data file link" - not clear what it is. A proper unique ID of the dataset should allow to find the data. Giving links to individual files in the data set is not practical as in this way the metadata may be missed. | Answer: a section for data file has been created that clarify this. |
| **License** |  |  |  |  |
| **Timing and location of study [Study]** | | | |  |
| **Timing: start of study\*** | JR: important to support also season as valid date format (type string) | Answer: The season can be either inffered from the date or preciesed in the description of the study. | HP: This can be a bit vague. Does an experiment start when the seeds are sown, the plants germinate or when the experimental treatment is applied. These can be 3 different dates! | Answer: Relevant start date is up to the data manager. This field is to give a generale idea of the timing of the study. The precise timing of eents are hadled through the Event section.  The whole list of sowing, germination, etc… can be handled through events. |
| **timing: end date of study** |  |  |  |  |
| **Geographic location of study (country)\*** |  |  |  |  |
| **Experimental site name** |  |  |  |  |
| **Geographic location of study (latitude)\*** |  |  |  |  |
| **Geographic location of study (longitude)\*** |  |  |  |  |
| **Geographic location of study (altitude)** |  |  |  |  |
| **Biological Material [Study]** | | | |  |
| **Organism\*** | JR: Support also none-NCBI taxonomy. In particular IPK used taxonomy classification is not 100% overlap to all NCBI taxonomy branches. Maybe its better to use MCPD 2.1 attributes 5-10 as strings (<https://www.bioversityinternational.org/fileadmin/user_upload/online_library/publications/pdfs/FAO-Bioversity_multi_crop_passport_descriptors_V_2_Final_rev_1526.pdf>) | Answer: We keep this taxonomy id source because it helps doing some interoperability/integration with polymorphism and genomic data. If later on we identify other sources of interoperability, we will add it in new version of MIAPPE. You can use : <https://bioportal.bioontology.org/ontologies/NCBITAXON/?p=classes&conceptid=http%3A%2F%2Fpurl.bioontology.org%2Fontology%2FNCBITAXON%2F4577&jump_to_nav=true> |  |  |
| **Genus** | SW-PK-HCW: follow MCPD is fine, but what does it mean practically? To replace "Infraspecific name" by "Subtaxa" from MCPD or to replace whole section by MCPD list of attributes - which indeed are accepted and capable of describing what we need (also source etc.). | Answer: having the minimal identificaiton from MCPD is important (accessionNumber, DOI, ….) but everything that is MCPD specific, including subTaxa/infraspecific name must stay in MCPD rather than being duplicated in MIAPPE. Those information can be accessed in the database in charge of the accessionNumber. For the case of a lab cross, the cross is specified in the species. Any other specific case can be handled in the description. |  |  |
| **Species** |  |  |  |  |
| **Life stage** | AJ : A phenomics experiment usually covers many developmental stages (BBCH, or PO). More important is developmental stage at time of sampling (already included) or imaging (is the process of imaging covered by sampling? > nondestructive sampling) | Answer: We will move it at the Assay level |  |  |
| **Material Source: Holding Institute/Stock Centre, accession\*** | ML: Key value pairs of holding Institute,  SW: if applicable use FAO institute code <http://www.fao.org/wiews-archive/wiewspage.jsp?i_l=@@&show=DownloadinstEN.jsp>, FAO-WIEWS codes should be used for holding institutes (http://www.fao.org/wiews/en/) | Answer : This list is not very helpful (obsolete and difficult to maintain) to my knowledge. Besides, it does not address lab collections. We are a bit reluctant on this list which is not very well updated. | HP: Why only for forestry and not for experiments with wild herbaceous species? | Answer: forestry has been removed to apllicability to other species. |
| **Material source: DOI** |  |  |  |  |
| **Derived Material** | ML: Would recommend to remove this attribute, becaue difficult to track and would need be updated after the experiment processed | Answer: We agree but sometimes it it necessary to keep experrimental ids that allow linked/intagrated data. It is not mandatory anyways. | RR: Is this exclusive to trees? It could be useful for landraces or wild relatives of crops. | Mention to tree removed |
| **Derived material Treatment** | Any treatment applied to the biological material before the study. Treatment/s made for all the biological material of the study. | <http://purl.obolibrary.org/obo/EO_0007210:PVY(NTN)>; | PK-HCW: Derived material: Not quite clear. Is this the actual sample or extract that is being used in phenotyping? If so, it should be part of another section. Or an ID of the biosource/accession? | Answer: clarified as MAterial source and biological Material. |
| **Derived material for trees, landraces, *in situ* material Geographic location (latitude)** | SW: Quite often, exact coordines are not available. Thus, an origin country as well as a description of the collecting site would be helpful. | Answer : in this case, uncertainty of the coordinates should be used (see below). This level of information is kept for MCPD full traceability which is not fully copied here. | MO : in general: only one format for spatial data should be defined (there are a similar fields in 'Experimental design') | Answer: aligned to the INSPIRE recomendations have been chosen : Degrees in the decimal format (ISO 6709) |
| **Derived material for trees, landraces, *in situ* material: Geographic location (longitude)** | MO: Geographical coordinates are often subjected to a level of uncertainty. For instance, this coordinates can refer only roughly to a latitude or longitude degree. For in situ preservations a uncertainty is often wanted to ensure the privacy of the location | Uncerntainty field added |  |  |
| **Derived material for trees, landraces, *in situ* material: Geographic location (altitude)** |  |  |  |  |
| **Derived material for trees, landraces, *in situ* material: Geographic location (uncertainty)** |  |  |  |  |
| **Environment: Growth facility [study]** | | | |  |
| **Type of growth facility\*** |  |  |  |  |
| **Cultural practice** | JR: Add section Environment: Field  Elements proposed: soil type, pesticide treatment, fertilization, plot size, ph-value, sawing density, slope | Those propositions can be handled as Environment Parameter. Some are already in the Appendix I environment. The other are likely in the Agro Ontology. See github... | À trouver. Include plant shuffling in automated facilities |  |
| **Average day temperature** | HP: Is it clear that the average is over the whole experimental period, not just the average over 1 day?  FT: Air Temperature | Renamed. Moved to appendix | PK-HCW: MIAPPE was based on the publications that presented the point of view of leading physiologists around 2010. Owing to the technological advancements this has changed considerably and now they set the minimum to hourly observations. It is fine, but it seems that we have to leave space for people who are not that advanced. It is possible that only the four environmental parameters (air temperature, light, VPD, wind) should be named as required, but a description like "Air temperature | Information about the air temperature throughout the experiment; hourly observations are recommended" would be less restrictive. | Answer, this is indeed the major concern for environment, what is minimal and how it must be measured and recorded can vary from one group to another. But on the other hand, leading specialist have strong recommendation. Therefore, moving this to an appendix/recommendation and moving to a more flexible systems allows both recommendation and adaptability. |
| **Average night temperature** | FT: **Organ temperature** | Renamed. Moved to appendix |  |  |
| **Change over the course of experiment** | HP: Is this the difference between the min and max T during 1 day, or the difference over the full experiment? I think the first is probably more useful, otherwise would it not be better to take the difference between the lowest and highest maximum T?  FT: appendix | Moved to appendix. Interresting question that will be handled in MIAPPE v2. | PK-HCW There should be a place for "Other environmental parameters" with a list of examples from MIAPPE 1.0 and a recommended frequency to track them. Why not leave them all in this section with general simplified names (without ‘average’, ‘day/night period’ etc.) and mark only the four as required? | Agrred, this should go in the right direction. |
| **Average daily integrated photosynthetic photon flux density (PPFD) measured at plant or canopy level.** | HP: Also here make sure it is over the full experimental period  FT rename | Renamed. Moved to appendix | PK-HCW: "Plot size" is lost but is important for field trials so a section for "Field environment" should be constructed. However, this may not be sufficient for all cases of "environments" | Back in the environment parameter seciton |
| **Average length of the light period** | FT: appendix | Moved to appendix. |  |  |
| **Light intensity** | FT: appendix  HP: At what moment? Or could this be the average? | Moved to appendix.Will be handled in MIAPPE v2. |  |  |
| **Range in peak light intensity** | FT: appendix  HP: Make clear that this is the range over the experimental period | Moved to appendix |  |  |
| **Fraction of outside light intercepted by growth facility components and surrounding structures** | FT: appendix  HP: No, this is a number between 0 and 1 | Moved to appendix |  |  |
| **Type of lamps used** |  | Moved to appendix |  |  |
| **R/FR ratio** |  | Moved to appendix |  |  |
| **Daily UV-A radiation** | FT: appendix | Moved to appendix |  |  |
| **Daily UV-B radiation** | FT: appendix | Moved to appendix |  |  |
| **Total daily irradiance** | FT: appendix | Moved to appendix |  |  |
| **Atmospheric CO2 concentration** |  | Moved to appendix |  |  |
| **Average CO2 during the light and dark periods** |  | Moved to appendix |  |  |
| **Average VPDair during the light period** | FT: Rename | Renamed and redifined, Moved to appendix |  |  |
| **Average relative humidity during the light period** | FT: appendix | Moved to appendix |  |  |
| **Average VPDair during the dark period.** | FT: appendix | Moved to appendix |  |  |
| **Average relative humidity during the dark period** | FT: appendix | Moved to appendix |  |  |
|  |  |  |  |  |
| **Environment: Rooting conditions [study]** | | | |  |
| **Rooting medium** | AJ: Gardeners usually precultivate sedlings in smaller pots before transplanting them into bigger (final) pots. This is also quite common for phenotyping experiments (we are often starting phenotyping in the small automated system and then transplant seedlings into the large system > this even covers two growth facilities, glasshouse and phytochamber). How to represent this? Repeat the Rooting conditions part? | With this senario we would have two experiments, one in the samll and one in the big systems. The tracability of the plant can be noted in the biological material preprocessing, example added accordingly. |  |  |
| **Container type** |  | Answer : all this section has been moved to appendix and can be handled through the new environment section. |  |  |
| **Container volume** |  |  |  |  |
| **Number of plants per containers** |  |  |  |  |
| **Plot size** | HP: Medium temperature: Daily average?. Also changes with depth and distance to the pot wall | MOved to appendix. This is the temperature of the the medium upon replenishment, not over the whole day. Left as is. |  |  |
| **Sowing density** | FT: rename to **Map of experimental design** | Map Line added to the study section. Moved to appendix. |  |  |
| **pH** | HP: For hydroponics this is also a relevant issue | Sentence removed (For hydroponics, leave the depth empty.) |  |  |
| **Environment: Nutrients [study]** | | | |  |
| **Medium composition** |  | Answer : all this section has been moved to appendix and can be handled through the new environment section. |  |  |
| **Extractable N content per unit ground area before fertiliser added** | Ft: replace area by volume | done |  |  |
| **Type and amount of fertiliser added per container/m2** |  |  |  |  |
| **Volume and frequency of water added per container/m2** | FT: Volume and dates of water added per container  HP & RR: L/M2 to be replaced by L/pot or L/m3 | Done, date replaced by timing. Unit: L pot size is noted elsewhere. |  |  |
| **Composition of nutrient solutions used for irrigation** | HP : including micronutrients | done |  |  |
| **Composition of the salts** | HP: What is different from the previous one? | Removed from appendix |  |  |
| **Electrical conductivity** | HP: Pretty vague | To be improved in MIAPPE v2 |  |  |
| **Treatments [study]** | | | |  |
| **Treatment name** | HP: But not all the plants of a study may have the same treatment | The whole treatment/factor section has been clarified and refactored | HP: all comments by HP have been added in the appendix. |  |
| **Treatment description** | FT: remove the whole section and allow free text. | Replaced by a Event and factor section for dated key value pairs. The free text section is mainly the Study.description. There is also the factor.description for detailed description of the treatment/cultural practice protocol applied. | PK-HCW:  - We should not try to formalize the description of the treatments (see general comments above). As a checklist, MIAPPE should require "A comprehensive description of the treatments that were applied to the plants in order to stimulate changes that were the subject of the study" (or something like that). Probably giving the list of (all? suggested?) possible treatments was not a good idea. Recommended formats may be "free text" or "literature reference" or "ontology annotation" (the latter is difficult to find). | The refactor goes in that direction. |
| **Time Factor [study, sample/observation unit]** | | | |  |
| **Time Factor ID** | AJ: Don't know what this should mean. In a phenotyping experiment we cover periods between days to months. What is a relevant date? Sowing date is listed already above. | time factor, as it is described in the crop ontology link, relates to the experimental design - date for sample/data collection or evaluation. To be deleted here??  → refactored into Event | PK-HCW: Time factor: Unclear. | Answer: agreed, refactored in events |
| **Time Factor description** |  |  |  |  |
| **Time Factor date** |  |  |  |  |
| **observationUnitID** |  |  |  |  |
| **studyId** |  |  |  |  |
| **sampleId ?** |  |  |  |  |
| **Experimental design [study]** | | | |  |
| **ObservationUnitID\*** | PK-HCW: The central concept of "Experiment description" is in fact "experimental unit", which is a term widely accepted in practical experimentation and also in the theory of experiment. Experimental units types are plot, pot, plant, etc. MIAPPE should not avoid using this term. “Experimental unit type/description” should be given as the first attribute in this section. | Answer: observationUnit, experimentalUnit, scientific object, study subject are all synonyms accepted in different system and communities. It has been added as a synonym. We might try to push unification of the naming in futur revision of the brapi. |  |  |
| **ObservationUnitType\*** |  |  |  |  |
| **BiosampleID** |  |  |  |  |
| **Spatial coordinates: ID** | ML: should be a key-value pair to identify the type of geo location. e.g. relative plot location in field or position in greenhouse. But how to denote dynamic positions in a automated greenhouse? | Refactored as ObservationUnit.Spatial distribution |  |  |
| **Spatial coordinates: type** |  |  |  |  |
| **Unit Spatial coordinates: unit** |  |  |  |  |
| **Experiment description\*** |  |  |  |  |
| **Observation unit biological material** |  |  |  |  |
| **Observation unit treatments** |  |  |  |  |
| **Replication - Technical replication ID** | PK-HCW: Whether some information or block is repeated (duplicated) for "each accession in the study" is the matter of format or schema, not of recommendations. In the MIAPPE mapping to ISA-Tab the information is repeated. In a RDF model this may not be necessary. The point is that the required information must be clear for all biosources, samples etc. | Answer : it is important to know if those are mean values or plot values. Likewise, the degree of repetition must be explained to first be able the level of elaboration of the dataset and second be able to reuse it. It seems therefore to be an important point. |  |  |
| **Replication - technical replication level hierarchy** | PK-HCW: MIAPPE 1 did not solve the problem of description for observational studies. MIAPPE 1.1 does something, but is rather limited to trees. What about other species? | Answer: it was more a formulation problem. Reference to tree specificites have been removed and it should beclearer now. |  |  |
| **Type of statistical design\*** | JR: should be set as mandatory. Furthermore, a suggestion to add another element: "Proposed statistical model for data analysis" | It depends of the type of study. Not relevant in some case like greenhouses. The proposed statistical model goes beyond the scope of MIAPPE, but can be added in the general description of the study. |  |  |
| **Sample collection, processing, management [study]** | | | |  |
| **Plant structure development stage** | AJ: Automated systems usually shuffle plants over time. Would be nice to have a check option for this. In case shuffling is applied, exact definition of positions doens't make sense. | Leaving Observation unit empty seems sufficient. Plus possibly adding automatic shuffling in study.description. |  |  |
| **Plant anatomical entity** | Does this include non-invasive image acquisition?  How to represent daily imaging events? | Imaging is handled as data file. Daily imaging is obtained through the assay and the time series fields. |  |  |
| **Sample description** | HP: O**xygenation status of sample** field was uncleare. | Answer: The sample processing has been left solely in MIAME. This field has been removed. |  |  |
| **Sampling time factors** | GK: biosource.timeFactor/lifeStage/treatment: this could certainly be multiple dates/times, ambiguous | Answer: one date by sample, but you can have multiple samples. The sample handling has been simplified and clarified. |  |  |
| **BiosampleID** |  |  |  |  |
| **observationUnitID** |  |  |  |  |
| **Observed variables [Assay]** | | | |  |
| **Trait** | JR: harmonization with  Official Variety Testing like German Bundessortenamt <https://www.bundessortenamt.de/internet30/index.php?id=9&L=1> | * check if they have machine readable ID but interesting * I have been though the web site and I do not find obvious machine readable lists of phenotyping variables with identifiers. It would be interesting though to convert these lists of measured traits and protocoles at the Crop Ontology format in the future for research use. * To be handled by converting the nationals Variety testing trait lists to ontologies. | PK-HCW: The central concept of "Experiment description" is in fact "experimental unit", which is a term widely accepted in practical experimentation and also in the theory of experiment. Experimental units types are plot, pot, plant, etc. MIAPPE should not avoid using this term. “Experimental unit type/description” should be given as the first attribute in this section. | Answer: there is a trade off between the complexity and pushing the documentaiton of trait/method/scale. Proposing this naming convention has proven helpful when preparing data and datafiles. It is better to keep it as recommended not mandatory. |
| **Variable name\*** | JR: not mandatory | agreed |  |  |
| **Variable id** |  |  |  |  |
| **Source of the variable id** |  |  |  |  |
| **Variable Method\*** |  |  |  |  |
| **Method description** |  | 2/ Extension of Method CO\_321:0000456 (Plant Height measured with ruler) : Plant height measure at 5 years, one year after Botritis inoculation. |  |  |
| **Reference associated to the method** |  |  |  |  |
| **Scale\*** |  |  |  |  |
| **Observation Time Unit** |  |  |  |  |

¹Mandatory information when preparing a submission to BioSamples database ([http://www.ebi.ac.uk/biosamples](http://www.ebi.ac.uk/biosamples)))

\*Mandatory information