



Standardizing data on phenotyping experiments using MIAPPE



A-F Adam-Blondon and C Pommier (INRA, Fr), D Faria (IGC, Pt), C Miguel (IBET, Pt), E Papoutsoglou (WUR, NL), P. Kersey (EBI, UK)

Why should we standardize phenotyping experimental data?

- To enable anyone (including yourself) to reuse it: **metadata about the experiment (who did it, for what purpose, where and how)**
- To enable the (automatic) integration of the experimental data with other types of data: **unique identification of the concepts used to link different data sets**

Phenotype 1 = **measurement** on a **cultivar** in an environment-**GPS1-time1**
Phenotype 2 = **measurement** on a **cultivar** in an environment-**GPS2-time2**
Genotype = observed marker's alleles on a **cultivar**
Climate 1 = climatic data at **GPS1-time1**

```
graph TD; P1[Phenotype 1] --- M1[measurement]; P1 --- C1[cultivar]; P1 --- E1[environment-GPS1-time1]; P2[Phenotype 2] --- M2[measurement]; P2 --- C2[cultivar]; P2 --- E2[environment-GPS2-time2]; G[Genotype] --- O[observed marker's alleles]; O --- C3[cultivar]; C1 -.- C3; E1 --- CL1[Climate 1]; CL1 --- CD[climatic data at GPS1-time1];
```

- To enable knowledge discovery: **metadata about the experiment, controlled vocabularies, ontologies**

Different steps to standardize data

- Check list of fields required to describe the data set : [metadata recommendations](#)

-> e.g. [MIAPPE](#)

- Use [controlled vocabularies/scales/ontologies](#): e.g. Crop Ontology, Gene Ontology, BBCH scales...
- Use [Persistent Uniques Identifiers](#): e.g. gene ID, accessions ID, Trait ID, pubmed ID, DOI,...
- Use [standard file formats](#): e.g. VCF, IsaTab, ...

Metadata : Data about the data

In the context of phenotyping experiment:

- The objective of the experiment
- Who contributed to the experiment
- What were the experimental procedures
- What was the biological material experimented
- ...

MIAPPE: Minimum Information About Phenotyping Experiment

- Developed and maintained by an international community interested in plant phenotyping: large community of breeders and biologists, European infrastructure for Plant Phenotyping (EPPN/EMPHASIS), European infrastructure of Bioinformatics (ELIXIR), Planteome, Excellence in Breeding Platform...
- www.miappe.org
- Implemented in a growing number of database and repositories storing data from phenotyping experiments (should be visible at <https://fairsharing.org/FAIRsharing.nd9ce9> but not yet really)

MIAPPE - Improvements

- **EU FP7. TransPlant** : MIAPPE v1.0
- **EU H2020 ELIXIR Excelerate**: MIAPPE v1.1 revised by the international community -> MIAPPE v1.1 (in progress) proposed here:
 - Detailed specifications for each field and examples
 - Checks with crops and forest trees communities
 - Work on the data model: alignment with ISA-Tab model, BrAPI

Capturing important information about phenotyping experiments using the MIAPPE standard

Two examples to illustrate how to capture important information about the phenotyping experiments:

- [1] Bouchet *et al.* 2013 (<http://doi.org/10.1371/journal.pone.0071377>): Adaptation of Maize to Temperate Climates: Mid-Density Genome-Wide Association Genetics and Diversity Patterns Reveal Key Genomic Regions, with a Major Contribution of the Vgt2 (ZCN8) Locus.
- [2] Monclus *et al.* 2012 (<http://dx.doi.org/10.1186/1471-2229-12-173>): Integrating genome annotation and QTL position to identify candidate genes for productivity, architecture and water-use efficiency in *Populus* spp

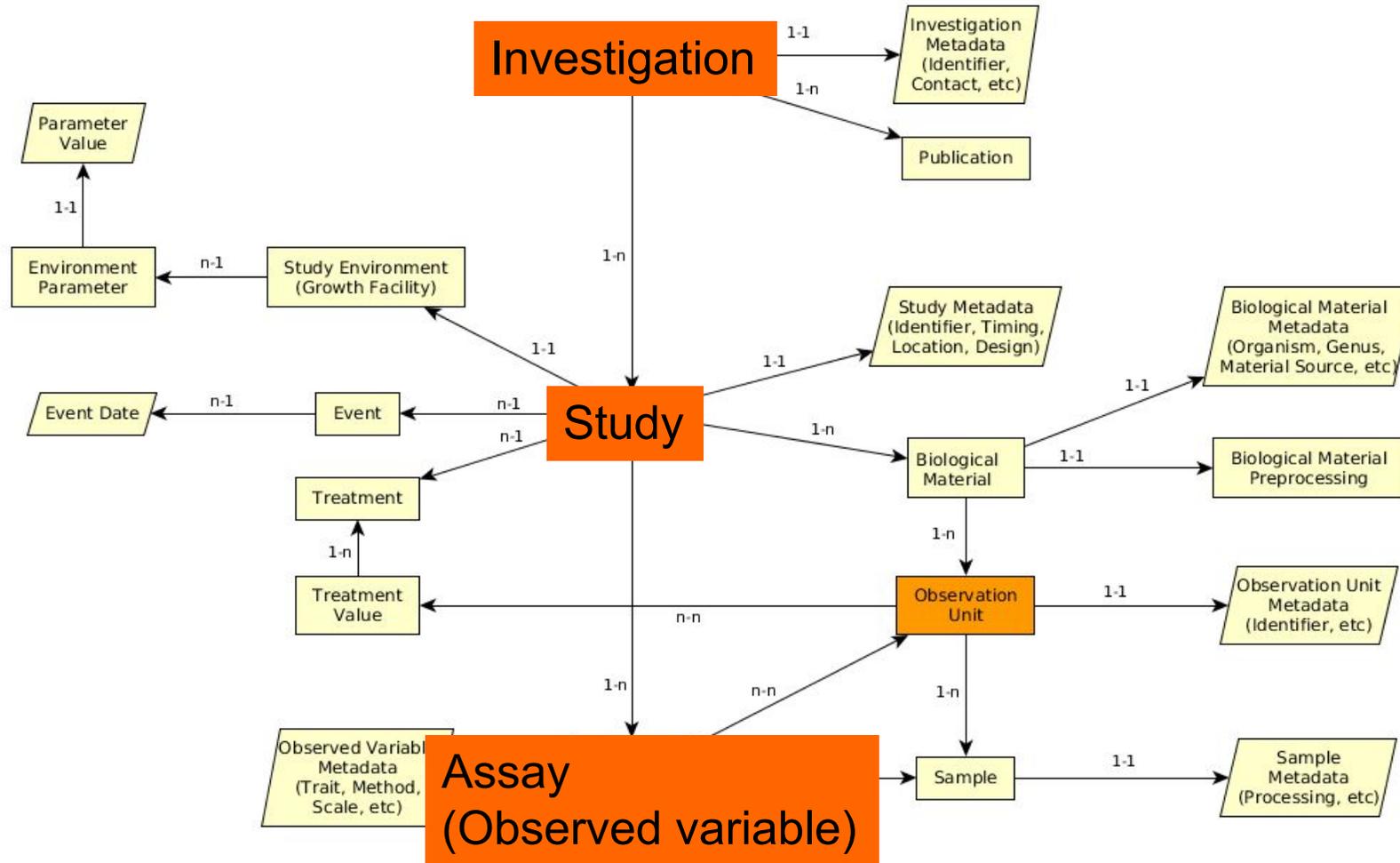
Bouchet et al 2013 [1] - Material&methods section on phenotyping experiments

- 375 inbred line collection defined to represent the European and American diversity and to cover a wide range of flowering times
- The whole panel was tested at three different locations:
 - Einbeck, Germany
 - Gif-sur-Yvette, France
 - Saint-Martin-de-Hinx, France
- French locations were evaluated over 3 years (2002–2004) and the German location over 1 year (2005).
- Evaluations were repeated twice at each location using a complete block design. In order to limit competition effects, each block was organized into four sub-blocks corresponding to earliness groups based on *a priori* information. Each individual plot consisted of a row of 15 plants at a density of approximately six plants per square meter.
- Days to anthesis for male flowering (MFLW) and days to silking for female flowering (FFLW) and anthesis to silking interval (ASI) were measured in thermal time.

Monclus et al 2012 [2] - Material&methods section on phenotyping experiments

- Field trial established in April 2003 located in Central France (Ardon, 47°49'41"N, 1°54'39"E, 110 m)
- The biological material consisted of a cloned 336 F1 progeny from an interspecific cross between the female *Populus deltoides* (Bartr. Ex Marsh.) '73028-62' from Illinois and the male *P. trichocarpa* (Torr. and Gray) '101-74' from Washington State.
- The trial was established from 25 cm- homogenous hardwood cuttings planted at a plant density of 6670 trees per ha. The trial was and consisted in 6 randomized complete blocks where each F1 genotype and each parent was represented by one replicate.
- Circumference and stem height were measured at the end of the first (winter 2003–2004) and second (winter 2004–2005) as described in Dillen et al. *Forest Ecol Manag.* 2007, 252 (1–3): 12-23). Growth increment in height and circumference during the second growing season were calculated.
- Leaf traits were measured in 2003: one fully illuminated mature leaf was collected on each tree according to Monclus et al. (<http://doi.org/10.1111/j.1469-8137.2005.01407.x>). Six calibrated discs of lamina were cut from this leaf, dried at 50 °C during 48 h and weighed, and specific leaf area (SLA, cm² g⁻¹) was computed. Leaf discs were ground to fine powder for analysis of leaf carbon isotope composition ($\delta^{13}\text{C}$), carbon (C_M) and nitrogen (N_M) contents. One-milligram subsamples of ground material were used for measuring the CO₂ produced by combustion and its ¹³CO₂/¹²CO₂ ratio by a continuous flux isotope ratio mass spectrometer. The discrimination between atmospheric CO₂ and plant material was calculated.

MIAPPE V1.1 (in progress) data model – the (ISA) backbone



MIAPPE main sections – Investigation

Investigations are research programmes with defined aims. They can exist at various scales: e.g. grant-funded programme of work with various published components; a single experiment.

One investigation holds one to many studies.

Metadata are similar to the information required for the deposition of a data set in an archive attached to a DOI: title, description, associated publications/people, ...

Examples:

- Bouchet *et al.* [1]: the whole set of multilocal and pluriannual phenotyping experiments constitute the investigation.
- Monclus *et al.* [2]: the whole set of measurements over three years constitute the investigation.

MIAPPE main sections – Study

A "**study**" (or **experiment**) comprises a series of assays (or measurements) undertaken to answer a particular biological question.

One study = one location

Metadata: all the informations describing the experiment as a whole -> timing, location, statistical design, cultural practices, etc...

Examples:

- Bouchet *et al.* [1] : sets of measurements conducted for each combination of year x location is a study (Einbeck_2005, Manguio_2002, Manguio_2003, Manguio 2004, ...)
- Monclus *et al.* [2] : One study can group all the measurements made on detached leaves in 2003 (Ardon_2003) and the growth measurements from two seasons data in another one (Ardon_2003-2005).

MIAPPE main sections – Observed Variable (Assay)

An Observed Variable (assay) corresponds to a specific measurement. The assay targets a trait of interest and the measurement conducted is associated with a method and a scale.

Metadata are the mandatory metadata of the Crop Ontology Trait Dictionary: trait, method and scale

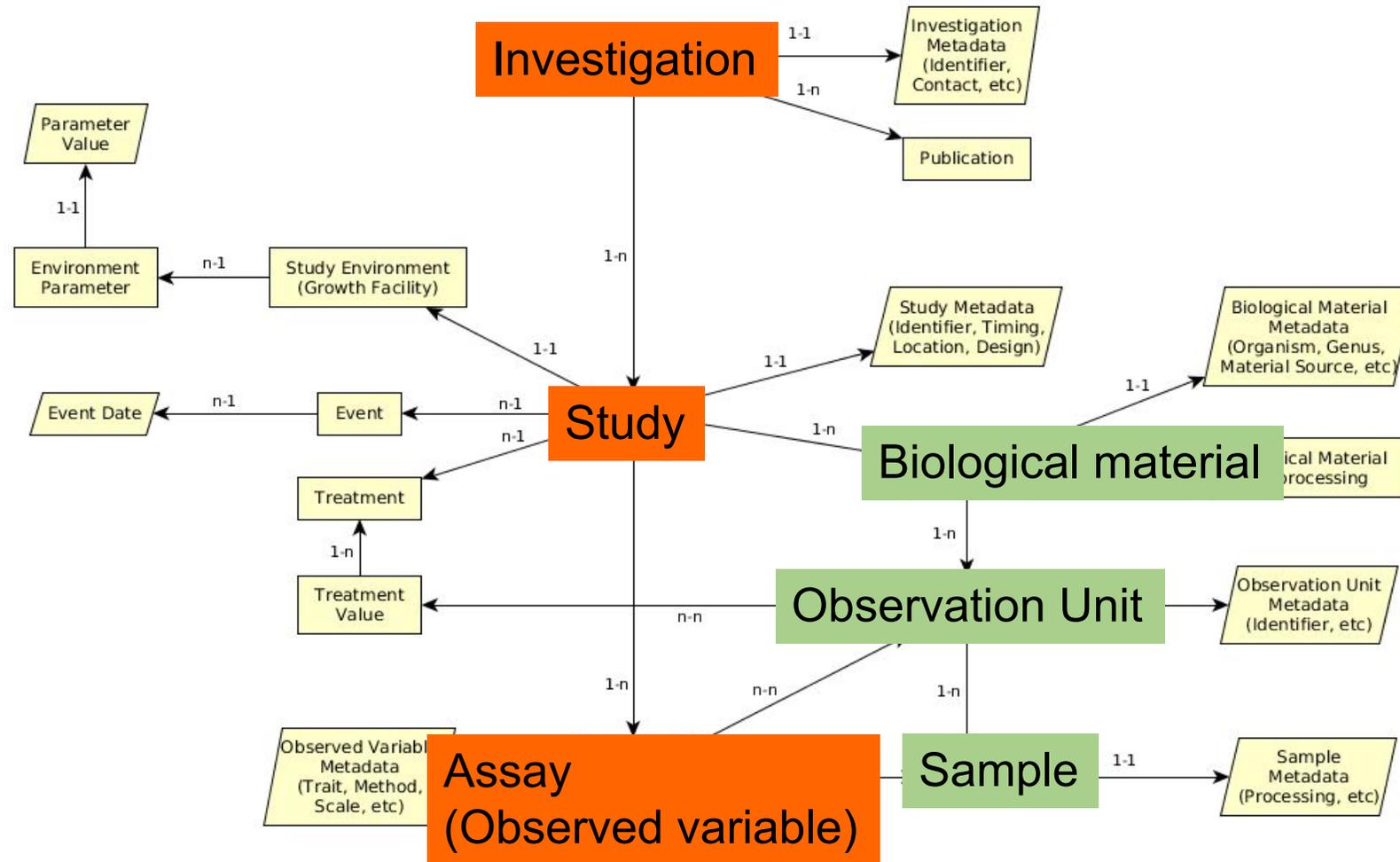
Example:

In each study of Bouchet *et al.* [1] three measurements are made:

- Daily temperature (environment)
- Female flowering developmental stage date
- Male flowering developmental stage date

This allows the authors to **generate three observed variables expressed in Growing Degree days (unit/scale) according to a defined method** associated to the data published: the male and female flowering dates and the interval between the two.

MIAPPE V1.1 (in progress) data model – The biological material assayed



MIAPPE main sections – Biological material

Biological material section: identification of the biological material being studied and of its source (if provided originally by a stock center for instance). This section is crucial for integrating phenotyping data with genomic or genetic data.

Metadata include mandatory information and identifiers from the Multicrop Passport Descriptor (MCPD) standard + the possibility to use GPS location for forest tree / in situ material provenance

Source of the material used:

accession, cultivar/variety, region of provenance, ...



(Biological) material used in the study: seed lot, cuttings...



Plant Samples used in the study: detached leaves, ...

MCPD identification system:

- Genebank/Lab + Species + accession number (mandatory)
- DOI

- Lab + internal accession number (mandatory)
- URI

- Lab + internal accession number (mandatory)
- BioSample ID

MIAPPE main sections – Observation Unit, Samples

Observation units are the objects in the study about which observations are made, and which may take different treatment values. They typically correspond to one or more individual plants. They can also be used for environmental variables.

Metadata are specific to MIAPPE: identifiers, location, replication, treatments, ...

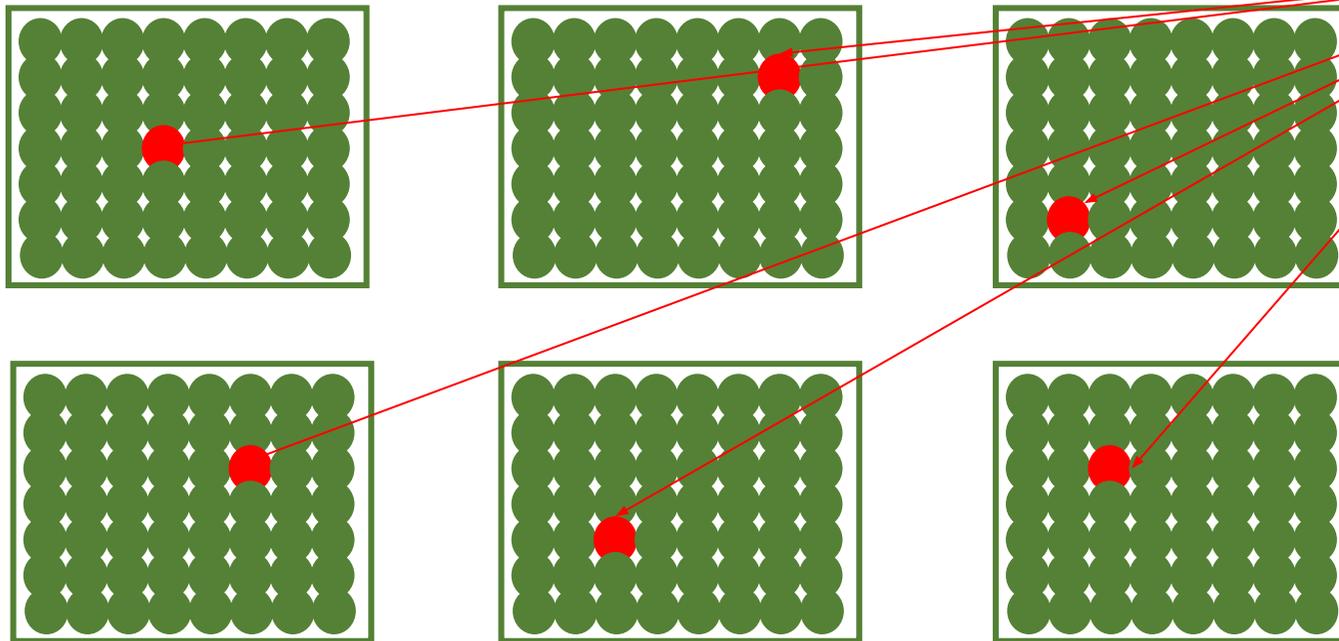
A sample is a portion of plant tissue extracted from an observation unit for the purpose of sub-plant observations and/or molecular studies.

Metadata are aligned with the BioSample DB generic model: identifiers, information about processing, ...

MIAPPE main sections – Observation Unit, Samples

Example: Monclus *et al.*, 2012, [2]

Populus deltoïdes
INRA:73028-62
doi:10.15454/1.4921786081024297E12



6 randomized blocks

1 observation unit = one tree

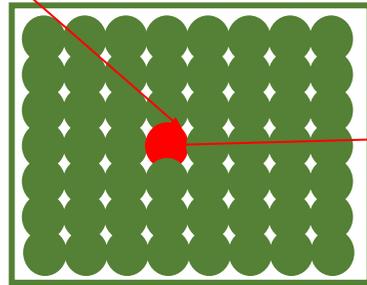
No treatment

6 replicates defined by their position in each block: row and column

MIAPPE main sections – Observation Unit, Samples

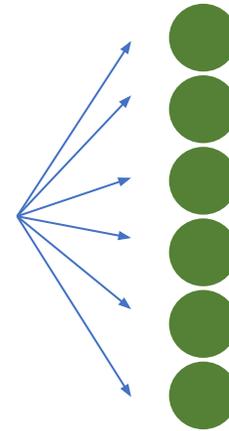
Example: Monclus *et al.*, 2012, [2]

Populus deltoïdes
INRA:73028-62
doi:10.15454/1.4921786081024297E12
[observation unit] : Block1-Row4-Col4



One leaf

Set of samples

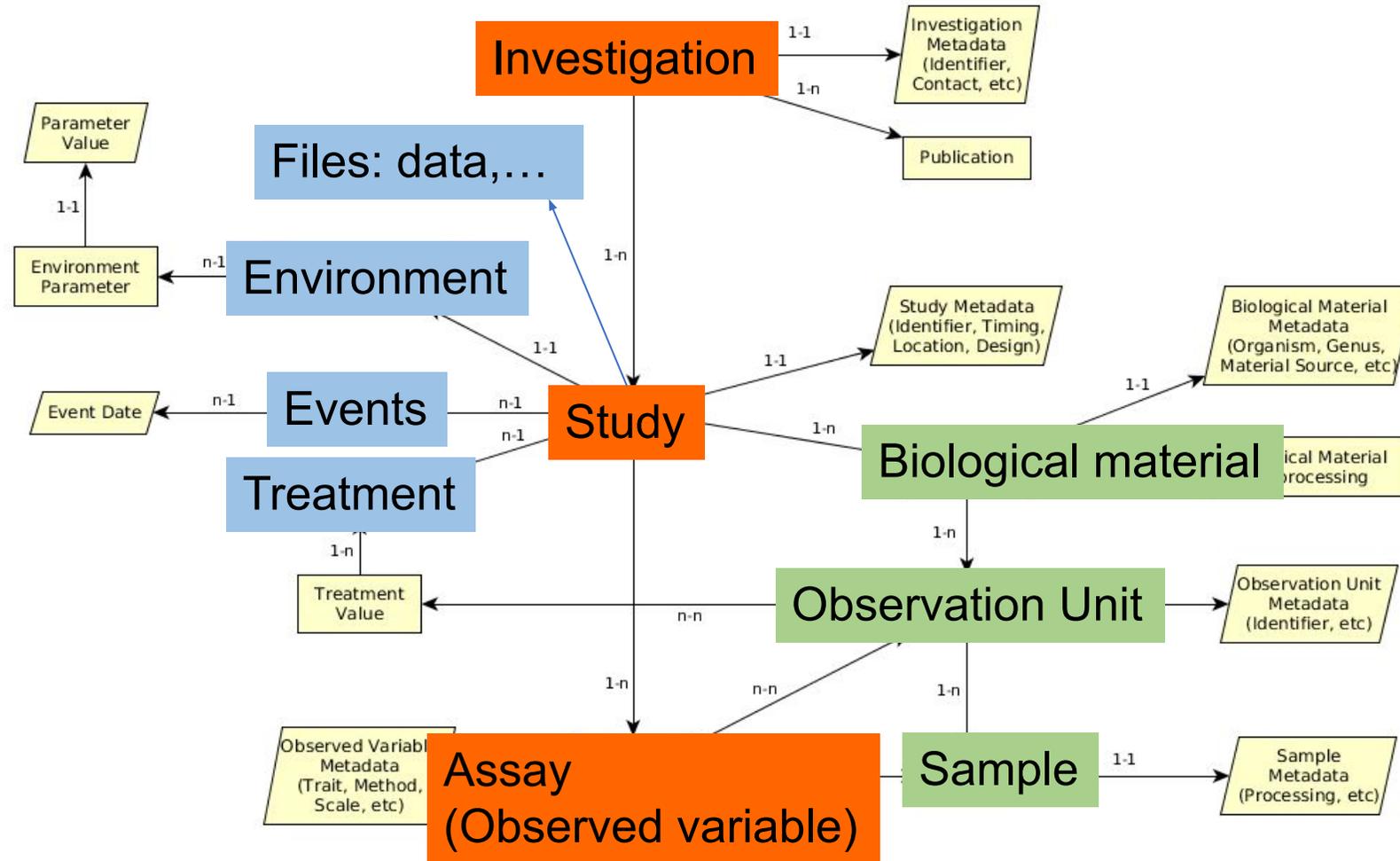


6 leaf disks

For each tree of each block
= for each observation unit

Different types of processing of the leaf disks depending on the measurement: can be captured by different sets of samples (e.g. if subsampling generates repetitions) or in the method of the observed variable.

MIAPPE V1.1 (in progress) data model – Other Important sections of information



MIAPPE main sections – Treatment

A treatment is a biotic or abiotic experimental factor the effects of which are being ascertained in the study. **It should take different values among the various observation units of the study.**

Metadata include a name, a description and a value of the treatment

Example (fiction): In Monclus et al 2012 [2] we could imagine 3 blocks watered and 3 blocks unwatered. The purpose of the experiment could then be to study the genetic adaptation to water deficit through its impact on plant growth.



Observation unit	Value[Watering]
Blocks 1-3	Watered
Blocks 4-6	Unwatered

MIAPPE main sections – Event

An event is discrete occurrence at a particular time in the experiment (which can be natural, such as rain, or unnatural, such as planting, watering, etc). Can be applied at the **whole study level** or at **the observation unit level** but in the latter case, it is not the factor being studied but an additional information on the course of the experiment (e.g. unwanted pathogen attack).

Metadata include a name, a description and a time/date

Examples: In Monclus *et al.*, 2012, [2]

- the field establishment date, 2003.
- the orchard was subjected to 15mm of rain on March 15, 2012 (fiction).



Study	Event		
	Name	Description	Date
Monclus <i>et al.</i> , 2012	Rain	15mm of rain on the orchard	2012-03-15

MIAPPE main sections – Environment

Environment: Environmental parameters and other experimental conditions that are controlled throughout the study and that are not a direct subject/factor of the experimentation. If the environmental variables are measured (recorded), they are described in the observed variable section.

Metadata include the description and the value of the environmental variable

Example:

Environment parameter	Environment parameter Value
Air temperature	22°C
Rooting medium composition	Ca (XEO:00058): 5 mg/L; ...



Let's work on your examples

- Documents for the session (both in progress):
 - MIAPPE V1.1 current version
 - MIAPPE V1.1 compliant metadata excel file