Plant Data Management for Phenotyping Experiments

Data standards and use cases for plant scientists

Cyril Pommier, Célia Michotey, Anne Françoise Adam Blondon

GenRes Bridge and AGENT have received funding from the European Union's Horizon 2020 research and innovation programme respectively under grant agreements No 817580 and 862613

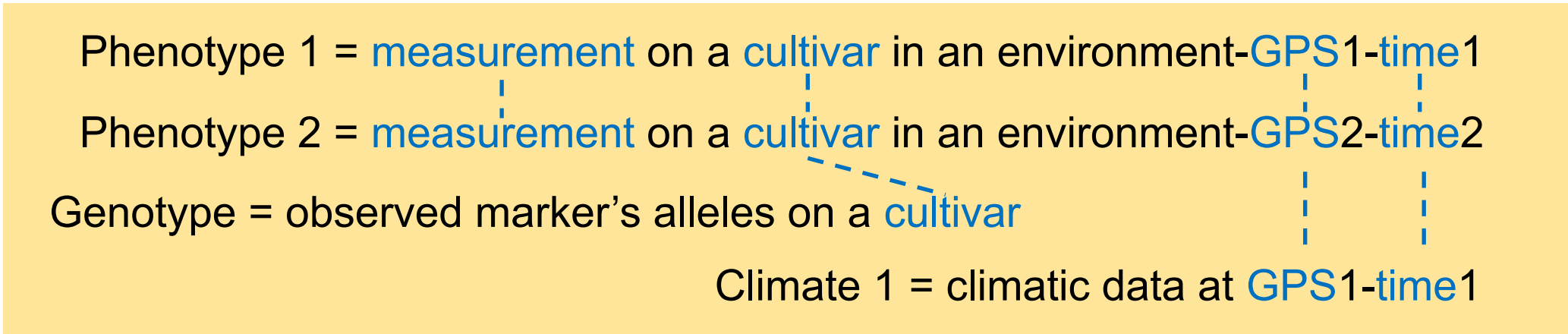
PLANT PHENOTYPING STANDARDS : WHY




Why should we standardize research data?

- To enable anyone (including yourself) to reuse it: metadata about the experiment (who did it, for what purpose, where and how)

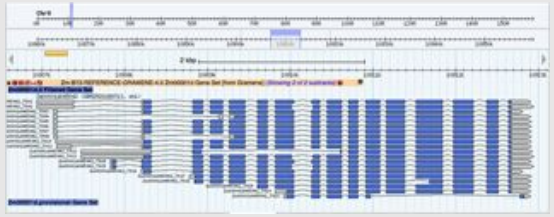
- To enable the (automatic) integration with other types of data: unique identification of the concepts used to link different data sets



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

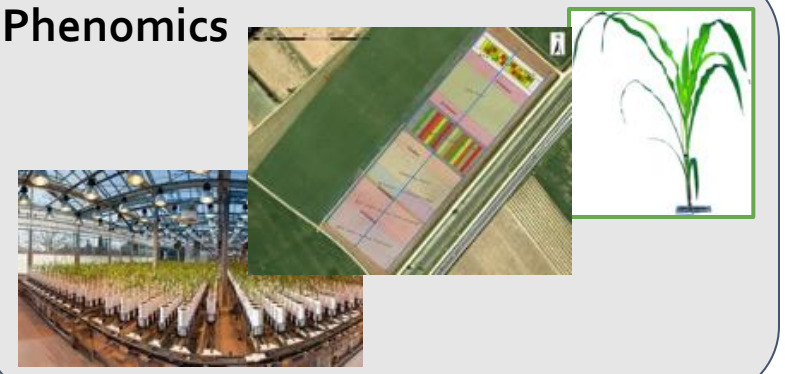
Genome / Environment / Phenome

Genetics & Genomics



Homogenous (Mostly)
Central Access (Mostly)

Phenomics



Heterogenous
Distributed

Plant Breeding
Genetic variations by Traits

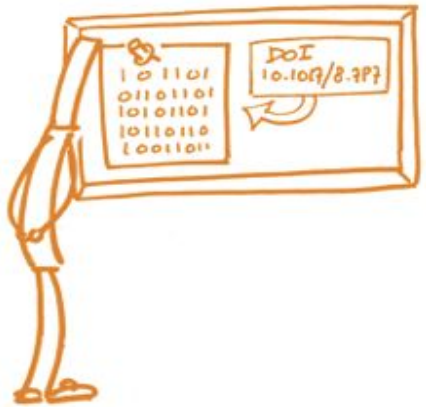
Climate Change Studies
Genotype by Environment
Phenology

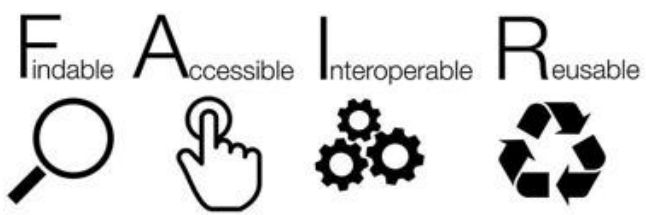
Environment

Heterogenous
Distributed

FAIR Data Principles

Wilkinson et al., **The FAIR Guiding Principles for scientific data management and stewardship.**
Scientific Data 3 (2016)





FINDABLE

Unique identifiers and metadata are used to allow data to be located quickly and efficiently

Ids
Metadata
Index

ACCESSIBLE

Data is open, free and universally available for research discovery efforts

Open Protocole
Permanent Metadata

INTER-OPERABLE

A common programming language is used to allow use in a broad range of applications

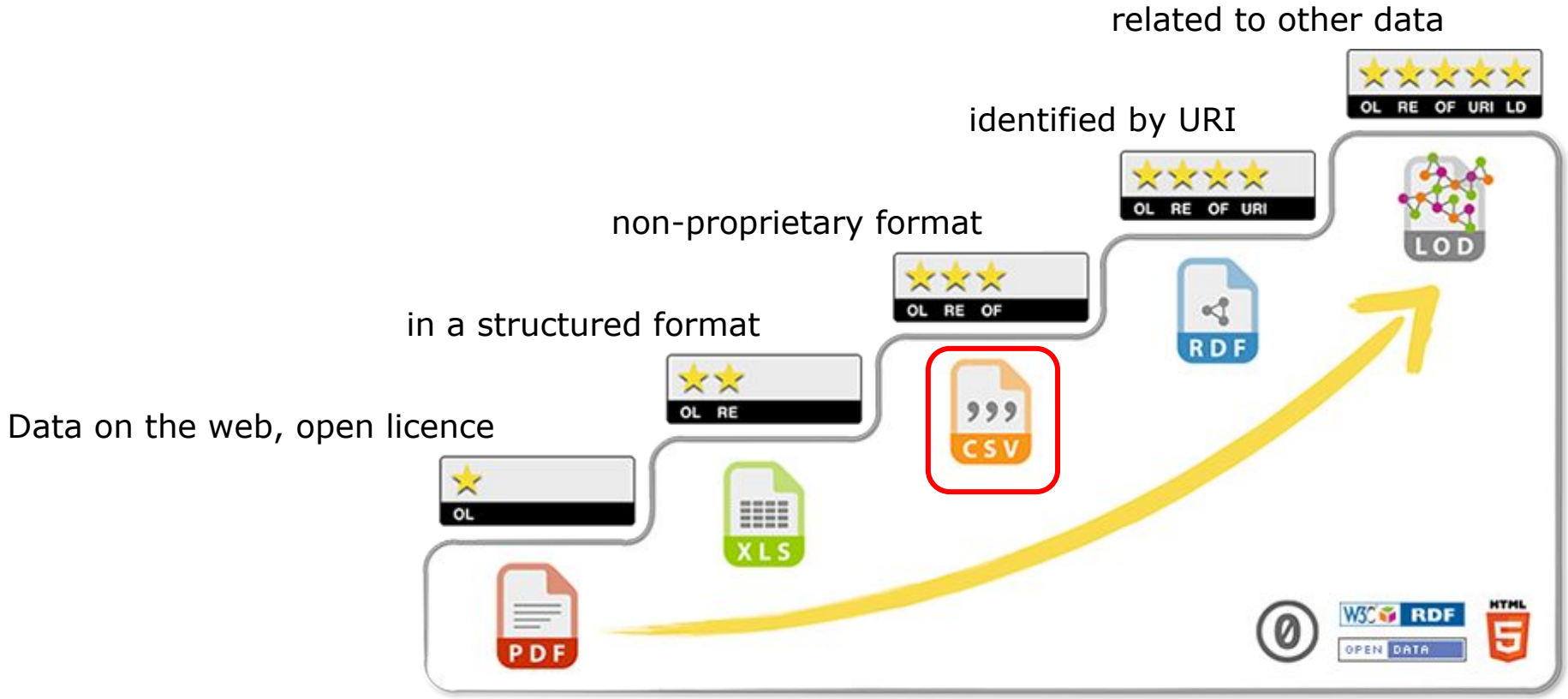
Semantics
Linked Data
Vocabularies

REUSABLE

All data is clearly described and outlines associated data-use standards

License
Well described
Provenance
Standards

5 stars Open Data



Progressing towards FAIR and Open Data requires a multidisciplinary cooperation :

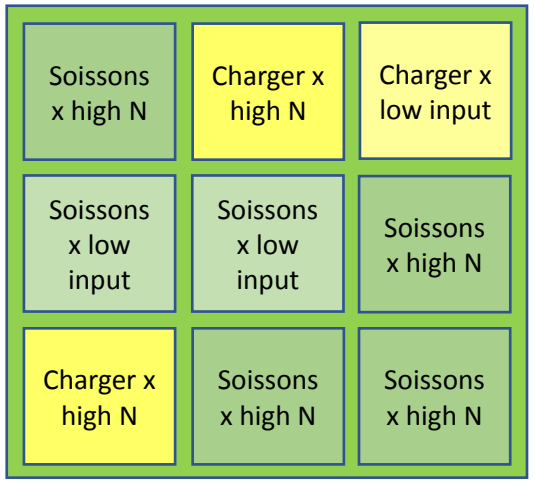
- Biologists
- Bioinformaticians
- Specialists of ontologies/semantics

Plant phenomics Data

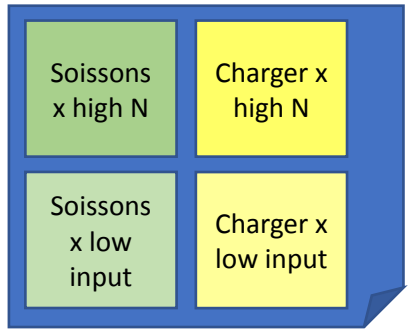
- Heterogenous data
- Different spatial scales
  - Metabolites
  - organ (leaf)
  - group of group of plants
  - whole experimental field
- Different time scales
  - Single measure
  - Time series : every 15 minutes
- Environment
- Complex life cycle

Phenotyping data life cycle

Raw data

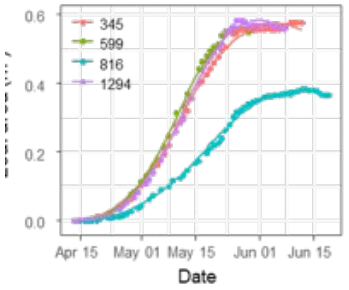


Elaborated data



Expe	Genotype	Treatment	Nitrogen	Date	Rep	Height
De	Soissons	low input	15,32	15/11/2011	1	5
De	Soissons	low input	15,31	16/11/2011	2	7
Cz	Soissons	low input	10	16/11/2011	5	6

Expe	Genotype	Treatment	Height
De	Soissons	low input	6
Cz	Soissons	low input	6



Plant Phenotyping Life cycle

Raw data long term conservation

Data acquisition

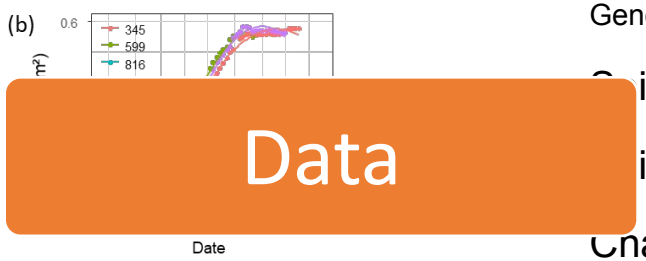
- Traceability
- Raw measures
- Data Cleaning
- Platform IS (Emphasis IS, PHIS, …)
- Reproducibility
- Traceability
- Provenance

Data computation

- New computation for each scientific question
- One raw dataset □ many computed datasets

Data publication

- One Data Publication by datasets.
- **Platform IS**
  - Raw
- **FAIR Data Repositories**
  - Computed



Genotype	traitement	height
isson	low input	5
isson	h	
Charger	low input	1
Charger	high N	2

Data

Knowledge

Variety charger

practice

PLANT DATA STANDARDS : WHO



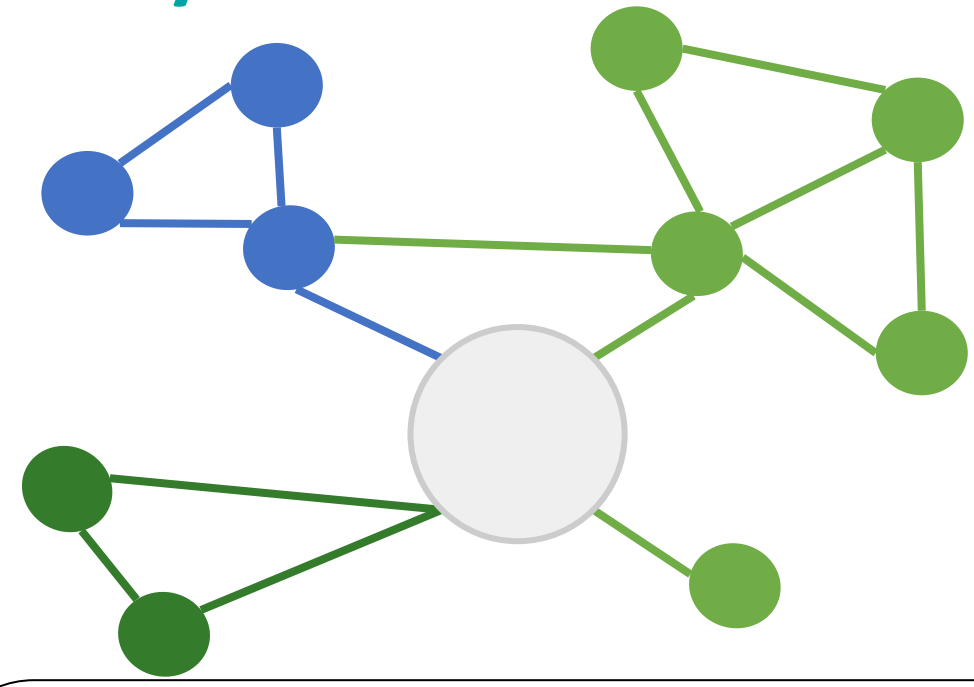

Interoperability is an International endeavour

National Networks

ifb de NBI
GERMAN NETWORK FOR BIOINFORMATICS INFRASTRUCTURE
PHENOME EMPHASIS
AgroMaizing
FRANCE GRILLES
Breed Wheat

Global Networks

WheatIS
WHEAT INITIATIVE
CGIAR



European Networks

elixir
EMPHASIS
European Plant Phenotyping Network
EPPN

International data standards

BrAPI
Web services

RDA
RESEARCH DATA ALLIANCE

miappe
Minimal information

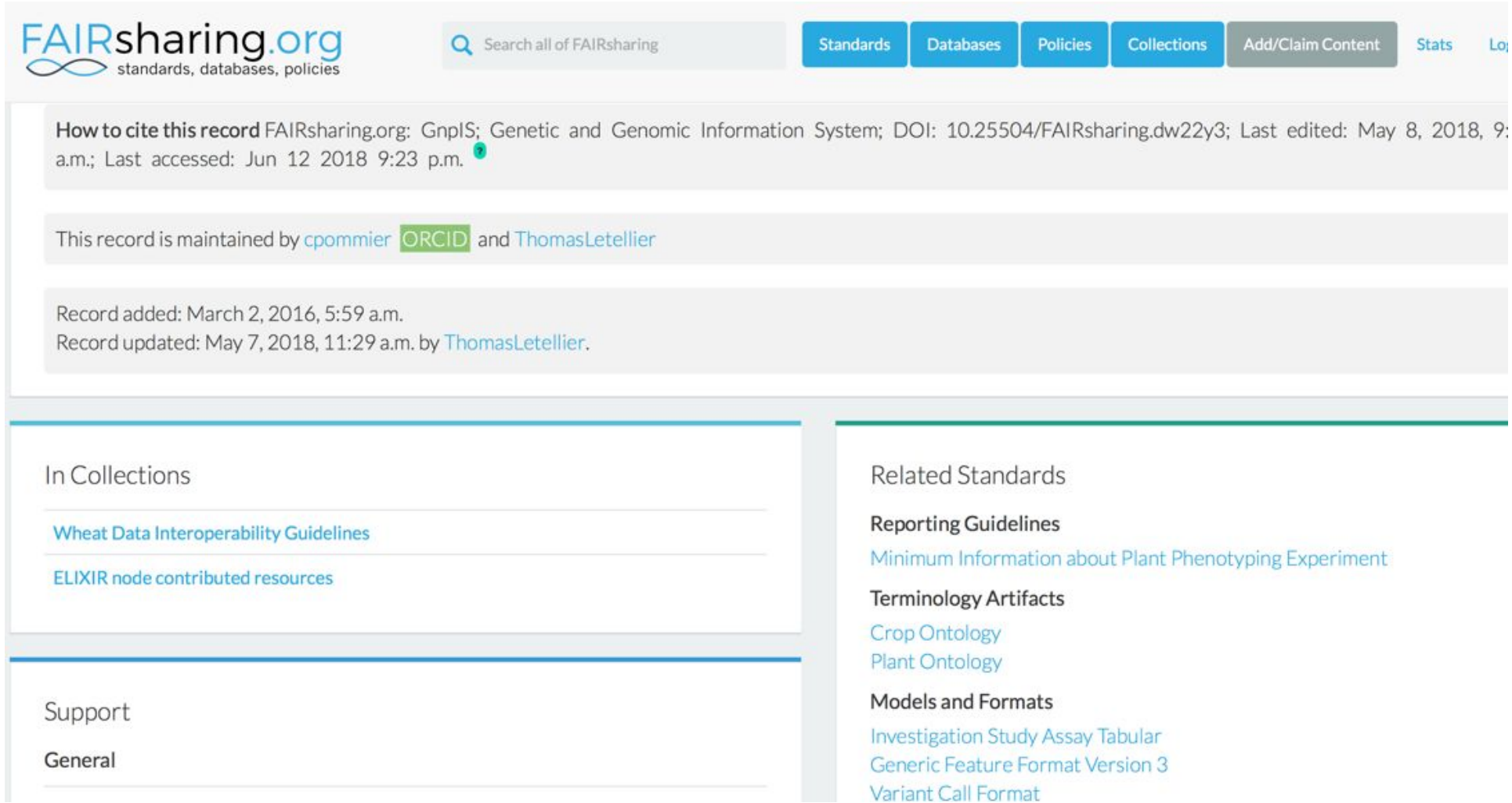
Crop Ontology
for agricultural data
Controlled vocabularies
Trait dictionaries

FAO
Bioversity International
MCPD

GenRes Bridge
ECPGR EUFORGEN ERFP
Genetic resources for food-secure and forested Europe

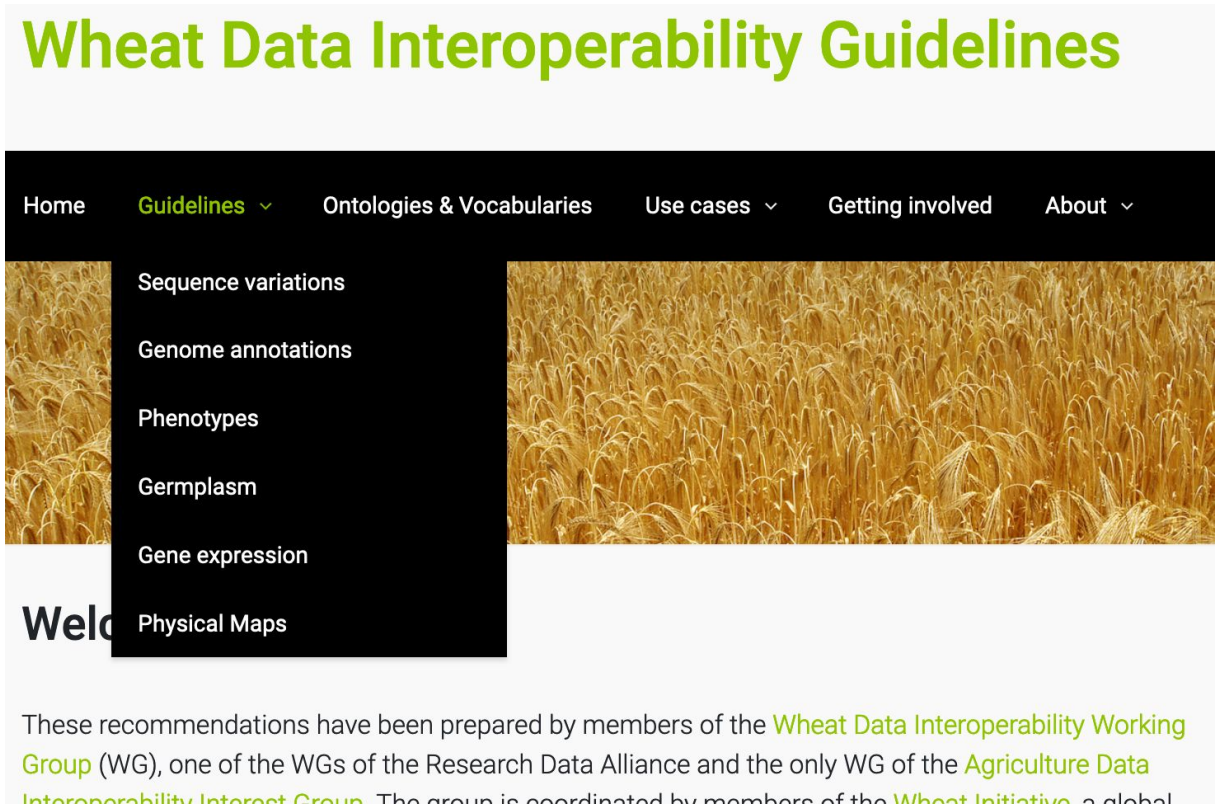
AGENT

Sharing standards: standards registries

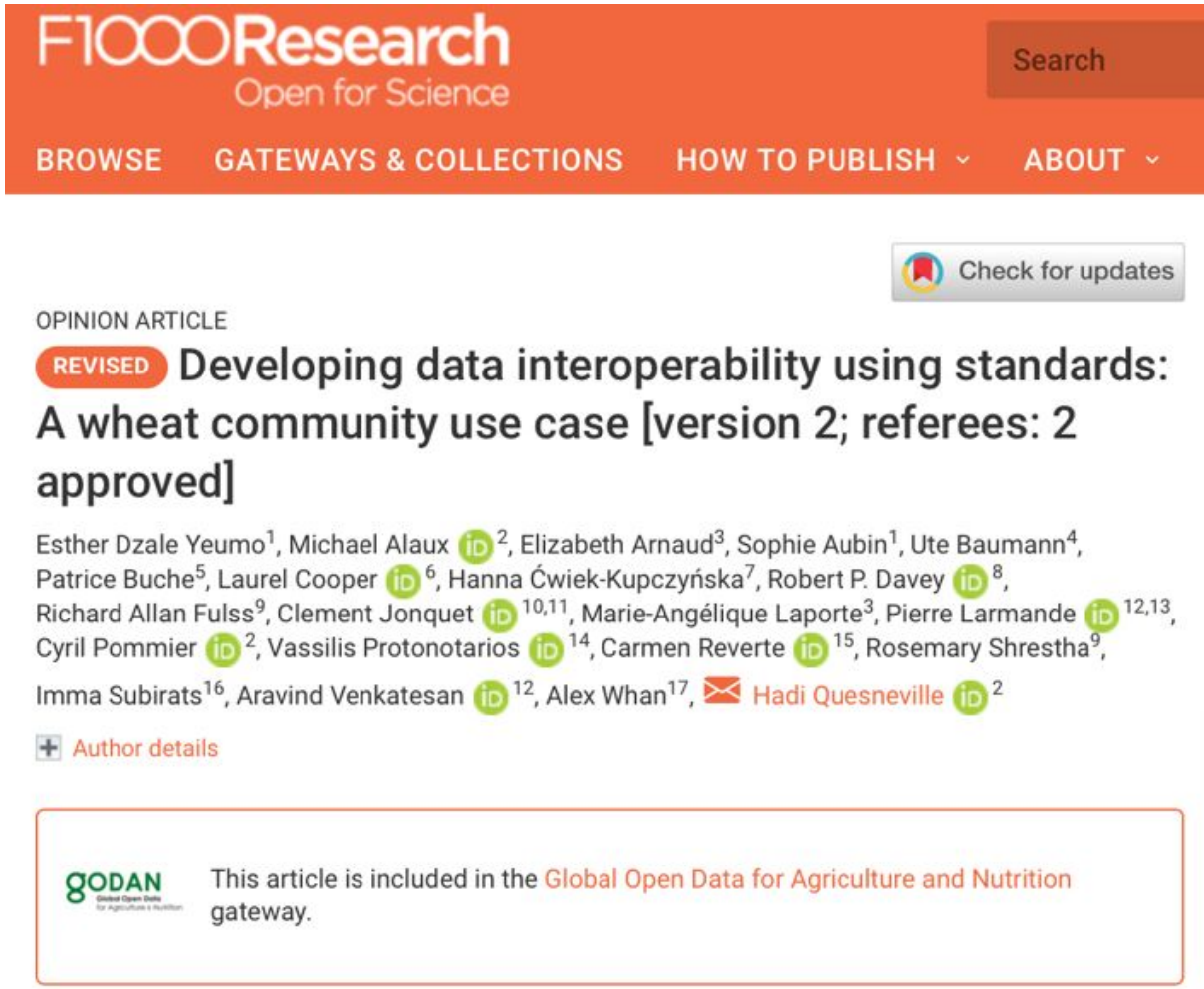


Community driven recommendations

- WheatIS: http://wheatis.org/DataStandards.php



- Community story



PLANT DATA STANDARDS




Data standards for FAIR

Semantic

- Description of the data
- Controlled vocabularies: term name and definitions
- Ontologies: semantic links between terms
- *Biologist* driven

Structure

- Formatting and Organizing the data
- Data Models
- Standards : CSV, VCF, GFF, MIAPPE (www.miappe.org) , etc…
- *Biologist & Computer scientist* driven

Persistent Unique Identifiers
URI, gene ID, accessions ID, Trait ID, DOI,…

Technical

- Data integration and sharing
- Interoperability : tools and systems
  - GA4GH
  - Breeding API www.brapi.org
- *Computer scientist* driven

Phenotype Semantic Standard: Ontologies




Phenotype Semantic Standard: Ontologies

- Describing traits/features in specific plant species
- Crop Ontology Trait + Method + Scale Semantic model
- Dedicated presentation

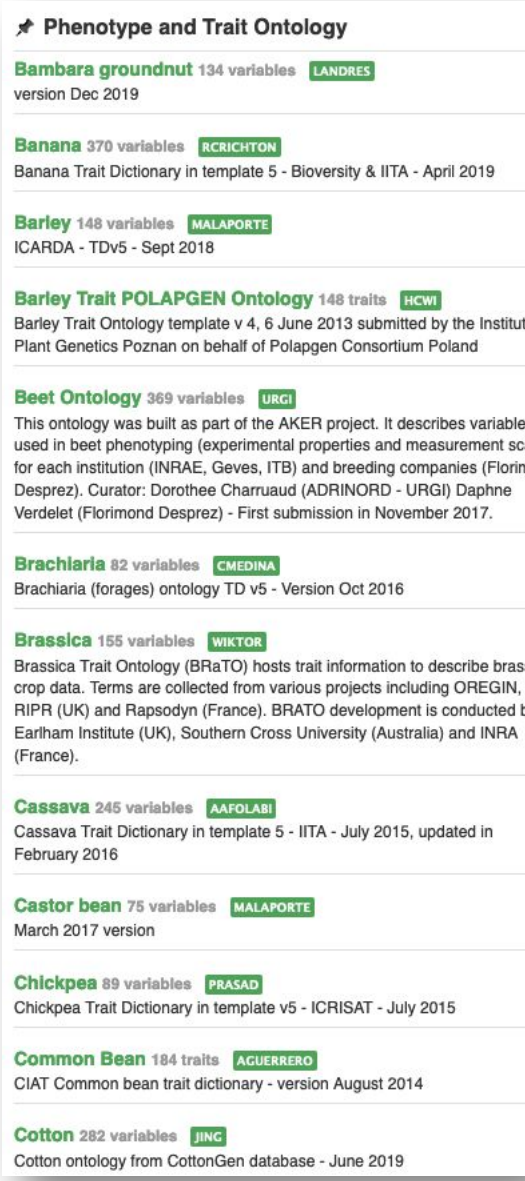
Context

- Need
  - Data annotation and FAIRness (reusability, findability)
  - Must capture crop specificities

Study	Genotype	CIR	HT	SUR
Domaine de Valcros	6579	600	17	1
Domaine de Valcros	6580	482	14	1

Study	Genotype	H	CIR
La Nerthe	6579	23	650
La Nerthe	6580	28	514

- Solution
  - Controlled vocabularies & ontologies to annotate phenotypes: “Crop Ontology”
  - Crop-specific vocabulary, community established
  - ⇒ 31 ontologies crop-specific available (including woody plants): www.cropontology.org



Phenotyping variable model

Variable = trait + method + scale

Trial	Genotype	CIR	HT	FLO
Domaine de Valcros	6579	600	17	2002-05-27
Domaine de Valcros	6580	482	14	2002-04-27

The phenotypic observation

Variable	**Variable ID**
	Variable name
	Variable synonyms
	Context of use
	Growth stage
	Variable status
	Variable Xref
	Institution
	Scientist
	Date
	Language
	Crop

Phenotyping variable model

Variable = **trait** + method + scale

Trial	Genotype	CIR	HT	FLO
Domaine de Valcros	6579	600	17	2002-05-27
Domaine de Valcros	6580	482	14	2002-04-27

Circumference of the tree

Total height

Flowering time

What is the studied character?

Trait	**Trait ID**
	Trait
	Trait class
	Trait description
	Trait synonyms
	Main trait abbreviation
	Alternative trait abbreviations
	Entity
	Attribute
	Trait status
	Trait Xref

Phenotyping variable model

Variable = trait + **method** + scale

Trial	Genotype	CIR	HT	FLO
Domaine de Valcros	6579	600	17	2002-05-27
Domaine de Valcros	6580	482	14	2002-04-27

Measured at breast height with a graduated ribbon

Measured from soil to crown with a pole or a clinometer

Interpolation after visual assessment

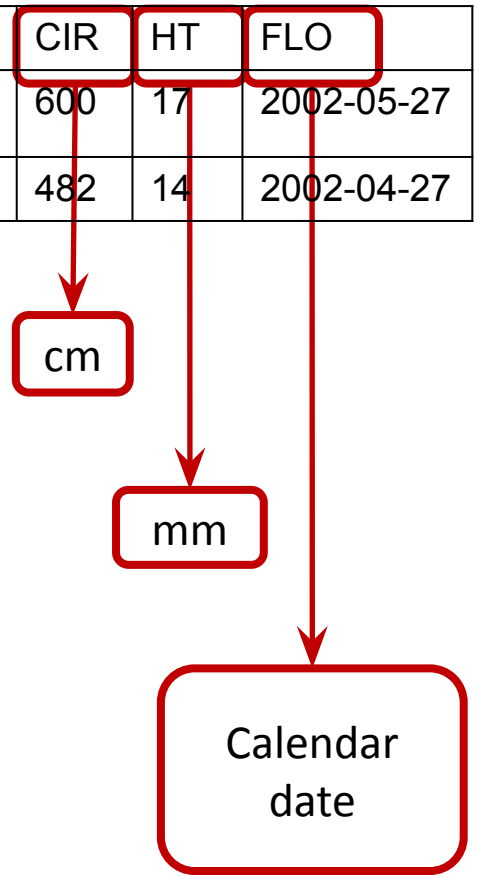
How is it observed?

	Method ID
Method	Method
	Method class
	Method description
	Formula
	Method reference

Phenotyping variable model

Variable = trait + method + **scale**

Trial	Genotype	CIR	HT	FLO
Domaine de Valcros	6579	600	17	2002-05-27
Domaine de Valcros	6580	482	14	2002-04-27



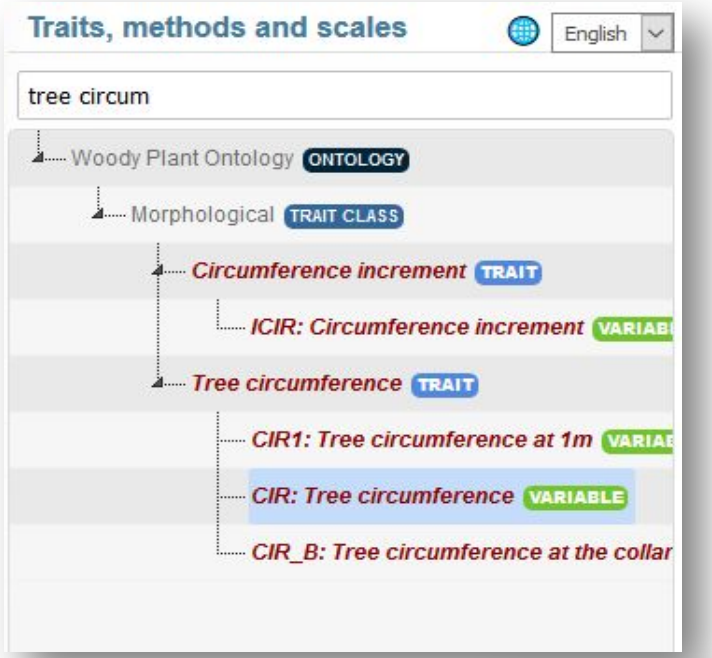
How is it expressed (unit or scale)?

Scale	**Scale ID**
	Scale name
	Scale class
	Decimal places
	Lower limit
	Upper limit
	Scale Xref
	Category 1
	Category 2
	…
	Category n

Tree Variable example1

https://urgi.versailles.inra.fr/ontologyportal#termIdentifier=CO_357:0000019

1 Trait
3 Variables
3 methods



CIR: Tree circumference VARIABLE

Ontology name	Woody Plant Ontology
Identifier	CO_357:0000019
Name	CIR
Synonyms	Tree circumference Tree girth C130 Ci130 CIR[Adonis]
Context of use	Research-intensive characterization Trial evaluation Breeding criterion
Status	Standard for INRA
Institution	INRA
Scientist	Célia Michotey
Date	13/03/2017

Tree circumference TRAIT

Identifier	CO_357:1000012
Name	Tree circumference
Description	Circumference of the tree main stem
Synonyms	Tree girth
Main abbreviation	CIR
Alternative abbreviations	CIRC Circ CI Girth
Entity	trunk
Attribute	circumference
Status	Standard for INRA
Class	Morphological

Ribbon 130cm protocol METHOD

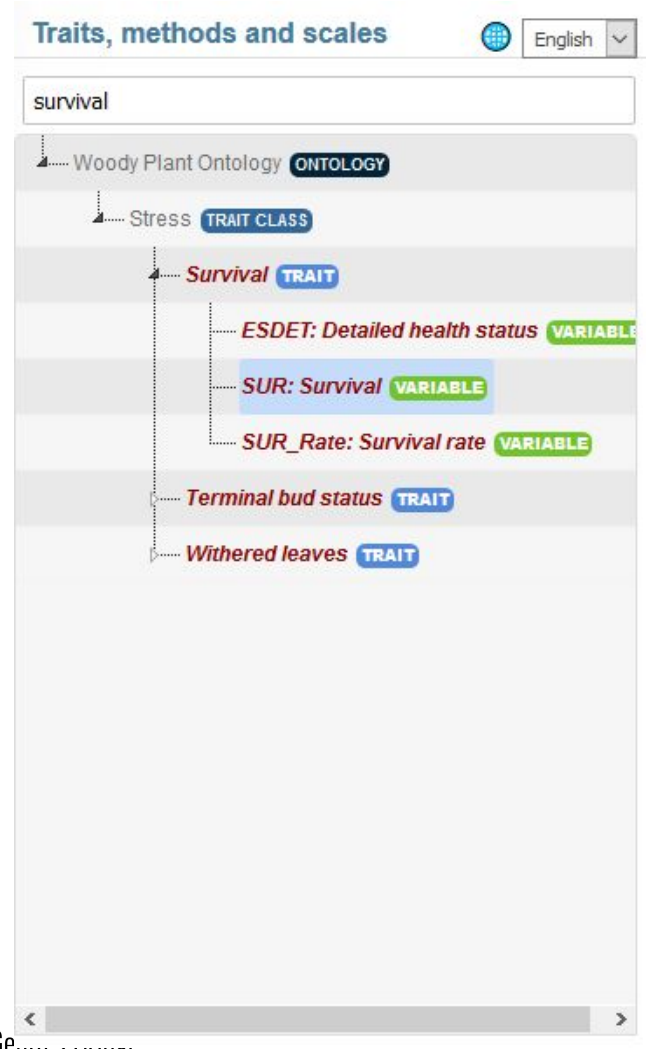
Identifier	CO_357:2000017
Name	Ribbon 130cm protocol
Description	Measured at 130cm from the ground (approximately breast height) with a graduated ribbon
Class	Measurement

mm SCALE

Identifier	UO:0000016
Name	mm
Data type	Numerical
Xref	http://purl.obolibrary.org/obo/...

Tree Variable example2

https://urgi.versailles.inra.fr/ontologyportal#termIdentifier=CO_357:0000082



SUR: Survival VARIABLE

Ontology name	Woody Plant Ontology
Identifier	CO_357:0000082
Name	SUR
Synonyms	Survival Health status ES S MOR SUR[Adonis]
Context of use	Trial evaluation
Status	Standard for INRA
Institution	INRA/IBET
Scientist	Célia Michotey
Date	13/03/2017
Crop	WoodyPlant

Survival TRAIT

Identifier	CO_357:1000070
Name	Survival
Description	Assessment of the survival state of the tree
Synonyms	Health status
Main abbreviation	SUR
Alternative abbreviations	S SR Survival Survie Etat sanitaire ES Vivants MOR
Entity	tree
Attribute	survival
Status	Standard for INRA
Class	Stress

Visual scoring METHOD

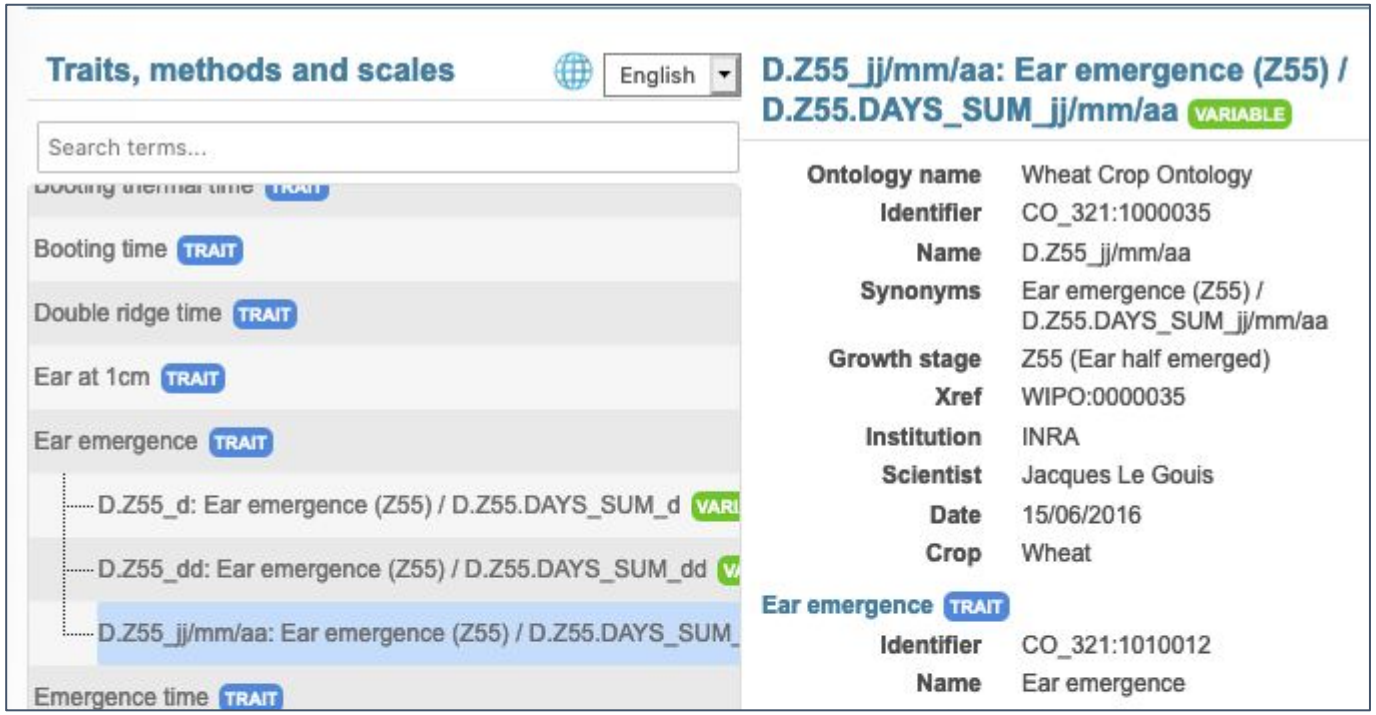
Identifier	CO_357:2000003
Name	Visual scoring
Description	Visual assessment with a reference scoring scale
Class	Estimation

Survival scoring scale SCALE

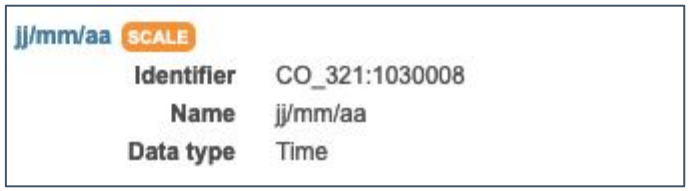
Identifier	CO_357:3000036
Name	Survival scoring scale
Data type	Nominal
Decimal places	0
Min	0
Max	2
Categories	0 = Alive 1 = Dead 2 = Doubtful

Wheat phenology : Ear emergence

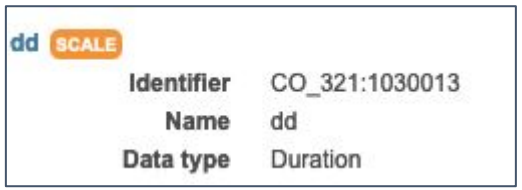
https://urgi.versailles.inra.fr/ontologyportal#termIdentifier=CO_321:1000035



Raw data: Emergence date

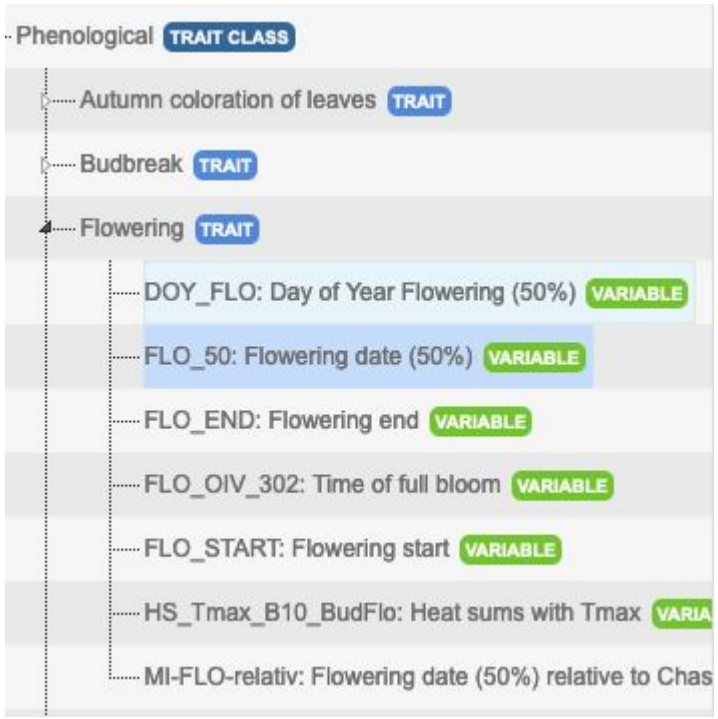


Computed: degree days



Vitis Phenology : Flowering

https://urgi.versailles.inra.fr/ontologyportal#termIdentifier=CO_356:1000011



Raw data: Flowering date

Calendar date SCALE

Identifier	CO_356:4000003
Name	Calendar date
Data type	Time

Computed: degree days

Tmax_Heatsummations METHOD

Identifier	CO_356:3000172
Name	Tmax_Heatsummations
Description	Heat summation with maximum tempŽratures
Formula	=sum(max(Tmax-tb),0)
Class	Measurement

Degree.days SCALE

Identifier	CO_356:4000038
Name	Degree.days
Data type	Numerical

Phenotype Structure Standard : MIAPPE
MINIMUM INFORMATION ABOUT A PLANT PHENOTYPING EXPERIMENT




New Phytologist

Methods | Open Access

Enabling reusability of plant phenomic datasets with MIAPPE 1.1

Evangelia A. Papoutsoglou, Daniel Faria, Daniel Arend, Elizabeth Arnaud, Ioannis N. Athanasiadis, Inês Chaves, Frederik Coppens, Guillaume Cornut, Bruno V. Costa, Hanna Ćwiek-Kupczyńska, Bert Droesbeke, Richard Finkers, Kristina Gruden, Astrid Junker, Graham J. King, Paweł Krajewski, Matthias Lange, Marie-Angélique Laporte, Célia Michotey, Markus Oppermann, Richard Ostler, Hendrik Poorter, Ricardo Ramírez-Gonzalez, Živa Ramšak, Jochen C. Reif, Philippe Rocca-Serra, Susanna-Assunta Sansone, Uwe Scholz, François Tardieu, Cristobal Uauy, Björn Usadel, Richard G. F. Visser, Stephan Weise, Paul J. Kersey, Célia M. Miguel, Anne-Françoise Adam-Blondon, Cyril Pommier ... See fewer authors

First published: 14 March 2020 | https://doi.org/10.1111/nph.16544 | Citations: 10

MIAPPE Overview and use cases

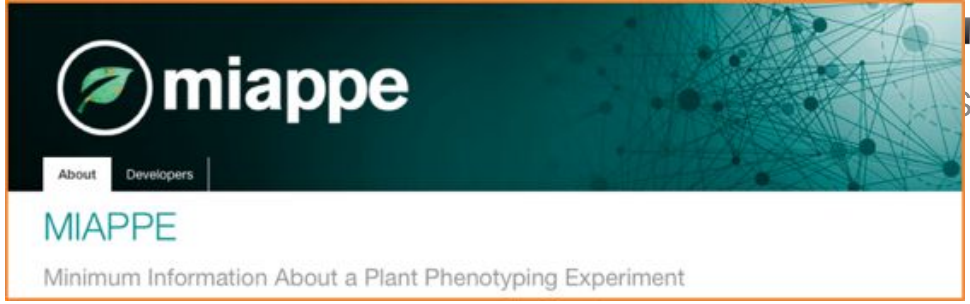
Phenotype Structure Standard

Minimum Information for Biological and Biomedical Investigations

A collection of the historical MIBBI foundry reporting guidelines. The minimum information standard is a set of guidelines for reporting data derived by relevant methods in biosciences. If followed, it ensures that the data can be easily verified, analysed and clearly

- Biologist Friendly
- Minimal and sufficient list of metadata:
  - The objective of the experiment
  - Who contributed to the experiment
  - What were the experimental procedures
  - What was the biological material experimented
  - ...

Phenotype Structure Standard



Minimal Information About Plant Phenotyping Experiment : version 1.1

www.miappe.org

- Many stakeholders
  - **Elixir, Emphasis, Bioversity, North American PPN**
- Open Community:
  - **Request for comments**
  - **Github Feature requests**
  - **Mailing lists**
  - **Meetings & Workgroups**
- Crops and woody plants

line #	MIAPPE Check list	Definition	Example	Format	Cardinality
		MIAPPE			
DM-1	Investigation	Investigations are research programmes with defined aims. They can exist at various scales (for example, they could encompass a grant-funded programme of work, the various components comprising a peer-reviewed publication, or a single experiment).			1 per MIAPPE submission
DM-2	Investigation unique ID	Identifier comprising the unique name of the institution/database hosting the submission of the investigation data, and the accession number of the investigation in that institution.	EBI:12345678	Unique identifier	0-1
	Investigation title	Human-readable string summarising the investigation.	Adaptation of Maize to Temperate Climates: Mid-Density Genome-Wide Association Genetics and Diversity Patterns Reveal Key Genomic Regions, with	Free text (short)	1

line#	Environment			
ENV-1	Non-exhaustive list of Environment Parameters.			
ENV-2	Environment parameters	Definition	Example	Format
ENV-3	Growth facility			
ENV-4	Air temperature	List of hourly air temperature throughout the experiment.	22 °C	Numeric
ENV-5	Organ temperature	List of hourly organ temperatures throughout the experiment	18 °C	Numeric

line #	Experimental Factors			
TR-1	Non-exhaustive list of Experimental Factors that can be applied.			
TR-2	Factor type	Definition	Example factor values	Format
TR-3	Seasonal environment	A plant treatment (EO:0001001) involving an exposure to a given conditions of regional seasons.	Spring season; dry season	Plant Environment Ontology:'EO_0007038'
TR-4	Air treatment regime	The treatment involving an exposure to wind/air with varying degree of temperature, which may depend on the study type or the regional environment.	28/25°C (Day/Night)	Plant Environment Ontology:'EO_0007161'
TR-5	Soil temperature regime	A physical plant treatment (EO:0007316) involving an exposure to varying degree of temperature, which may depend on regional environment.	27/25°C (Day/Night)	Plant Environment Ontology:'EO_0007161'

Papoutsoglou *et al*. (2020) Enabling reusability and interoperability of plant phenomic datasets with MIAPPE 1.1. New Phytol, 227:260-273; **https://doi.org/10.1111/nph.16544**

Capturing important information about phenotyping experiments using the MIAPPE standard

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

- ***MAIZE:*** [1] Millet *et al.* 2019 (https://doi.org/10.15454/IASSTN): A multi-site experiment in a network of European fields for assessing the maize yield response to environmental scenarios.

- ***POPLAR:*** [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

Millet et al 2019 [1] - Material&methods section on phenotyping experiments

- A panel of **256 maize hybrids** was grown with **two water regimes** (irrigated or rainfed).
- **Location**: seven fields in 2012 and 2013, plus one site in Chile in 2013
- This resulted in 29 experiments defined as the combination of one year, one site and one **water regime**, with two and three **repetitions** for rainfed and irrigated treatments, respectively.
- A detailed **environmental characterisation** was carried out, with hourly records of micrometeorological data and soil water status, and associated with precise measurement of phenology.
- « **grain.yield**»: yield adjusted at 15% grain moisture, in ton per hectare (t ha-1). «**grain.number**»: number of grain per square meter. «**grain.weight**»: individual grain weight (mg). «**anthesis**»: male flowering (pollen shed), in thermal time cumulated since emergence (d20°C). «**silking**»: female flowering (silking emergence), in thermal time cumulated since emergence (d20°C). «**plant.height**»: plant height, from ground level to the base of the flag leaf (highest) leaf (cm). «**tassel.height**»: plant height including tassel, from ground level to the highest point of the tassel (cm). «**ear.height**»: ear insertion height, from ground level to ligule of the highest ear leaf (cm).

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

- **3 Field trial** established in April 2003 **located** in France (Ardon, 47°49’41"N, 1°54’39"E, 110 m),Italy Cavallermaggiore ((44° 43' 0" N) , (7° 41' 0" E)), UK Headley ((51° 7' 0" N) , (-1° 10' 0" W))

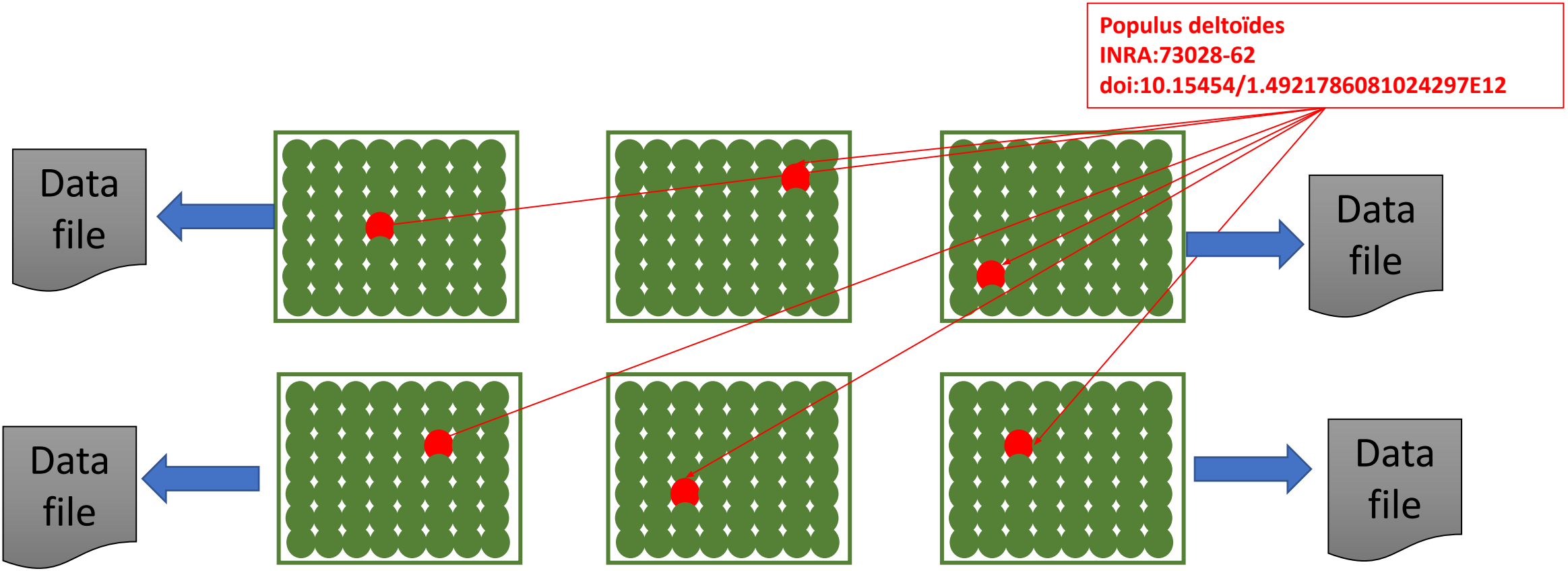
- The **biological material** consisted of a cloned 336 F1 progeny from an interspecific cross between the female *Populus deltoides* (Bartr. Ex Marsh.) ‘73028-62’ from Illinois and the male *P. trichocarpa* (Torr. and Gray) ’101-74’ from Washington State.

- 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** where 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 °C and weighed, and specific leaf area (SLA, $cm^2\ g^{-1}$) was computed. Leaf discs were ground to fine powder for analysis of leaf carbon isotope composition ($\delta^{13}C$), carbon (C_M) and nitrogen (N_M) contents. One-milligram subsamples of ground material were used for measuring the CO_2 produced by combustion and its $^{13}CO_2/^{12}CO_2$ ratio by a continuous flux isotope ratio mass spectrometer. The discrimination between atmospheric CO_2 and plant material was calculated.

MIAPPE 1.1 Overview

Typical Dataset: Monclus *et al., 2012,* [2]

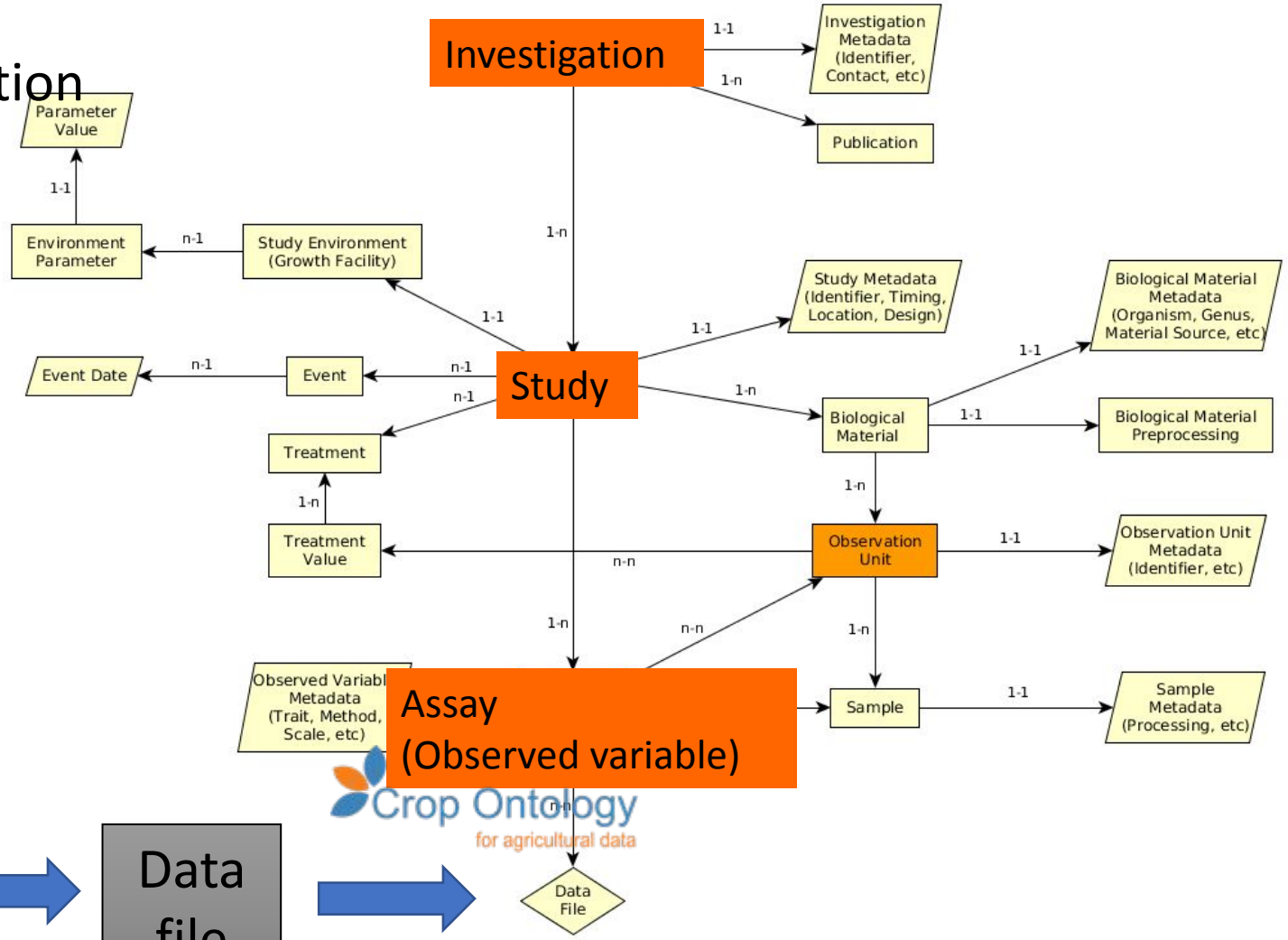


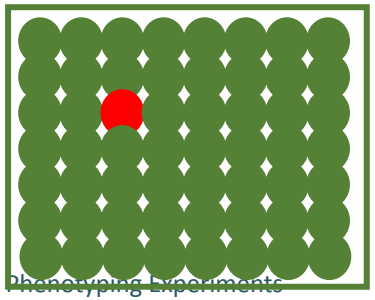
6 randomized blocks
1 observation unit = one tree
No treatment
6 replicates defined by their position in each block: row and column

MIAPPE V1.1 Overview: the (ISA) backbone

- **Investigation**: whole dataset
- **Study** : one <u>experiment</u> in one location for one to several year
- **Assay**: Level + Trait + Method + Scale/Unit

- Level:
  - Plant
  - Microplot
  - Block
  - Trial
  - …





Data file

MIAPPE V1.1 Overview: Data file content

- Any format (Near Infra Red Spectrum, Images, Image Archives references, ….)
- Mostly tabular
- Metadata on each column header

A	B	C	D	E	F	G	H	I
Accession Number	Trial Site	Campaign	Circum1: Tree circumference at 1 year	Date [Circum1]	Height1: Tree total height at 1 year	Date [Height1]	Shoots3: Number of resprouts at 3 years	Date [Shoots3]
661300270	Ardon	2004	45.645632645603683	12/01/2004	284.3	12/01/2004		
661300270	Ardon	2005					14.630625	12/05/2005
661300444	Ardon	2004	38.96112577281653	12/01/2004	228.8	12/01/2004		
661300444	Ardon	2005					8.5030559999999991	12/05/2005
661300312	Cavallermaggiore	2004	52.4	01/01/2004	249.9	01/01/2004		
661300312	Cavallermaggiore	2005					12.981609000000001	01/05/2005
661300371	Cavallermaggiore	2004	45.74	01/01/2004	230.2	01/01/2004		
661300371	Cavallermaggiore	2005					10.3041	01/05/2005
661300487	Cavallermaggiore	2004	72.52	01/01/2004	309.8	01/01/2004		
661300487	Cavallermaggiore	2005					10.679823999999998	01/05/2005
661300585	Cavallermaggiore	2004	71.739999999999995	01/01/2004	305.7	01/01/2004		
661300585	Cavallermaggiore	2005					10.956100000000001	01/05/2005
661300468	Headley	2004	45.27	01/01/2004	247	01/01/2004		
661300468	Headley	2005					15.888196000000002	01/05/2005
661300469	Headley	2004	70.930000000000007	01/01/2004	313	01/01/2004		
661300469	Headley	2005					13.271448999999999	01/05/2005
661300533	Headley	2004	57.67	01/01/2004	258.8	01/01/2004		

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**
  - **Similar to Archive deposition**
  - **DOI, title, description, associated publications/people, ...**

- **Examples Investigation :**
  - **MAIZE [1]: the whole set of multilocal and pluriannual phenotyping experiments**
  - **POPLAR [2]: the whole set of multilocal measurements over three years**

MIAPPE main sections – Study

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

- **Metadata**
  - **experiment as a whole**
  - **timing, location, statistical design, cultural pratices (but not event lists), etc...**

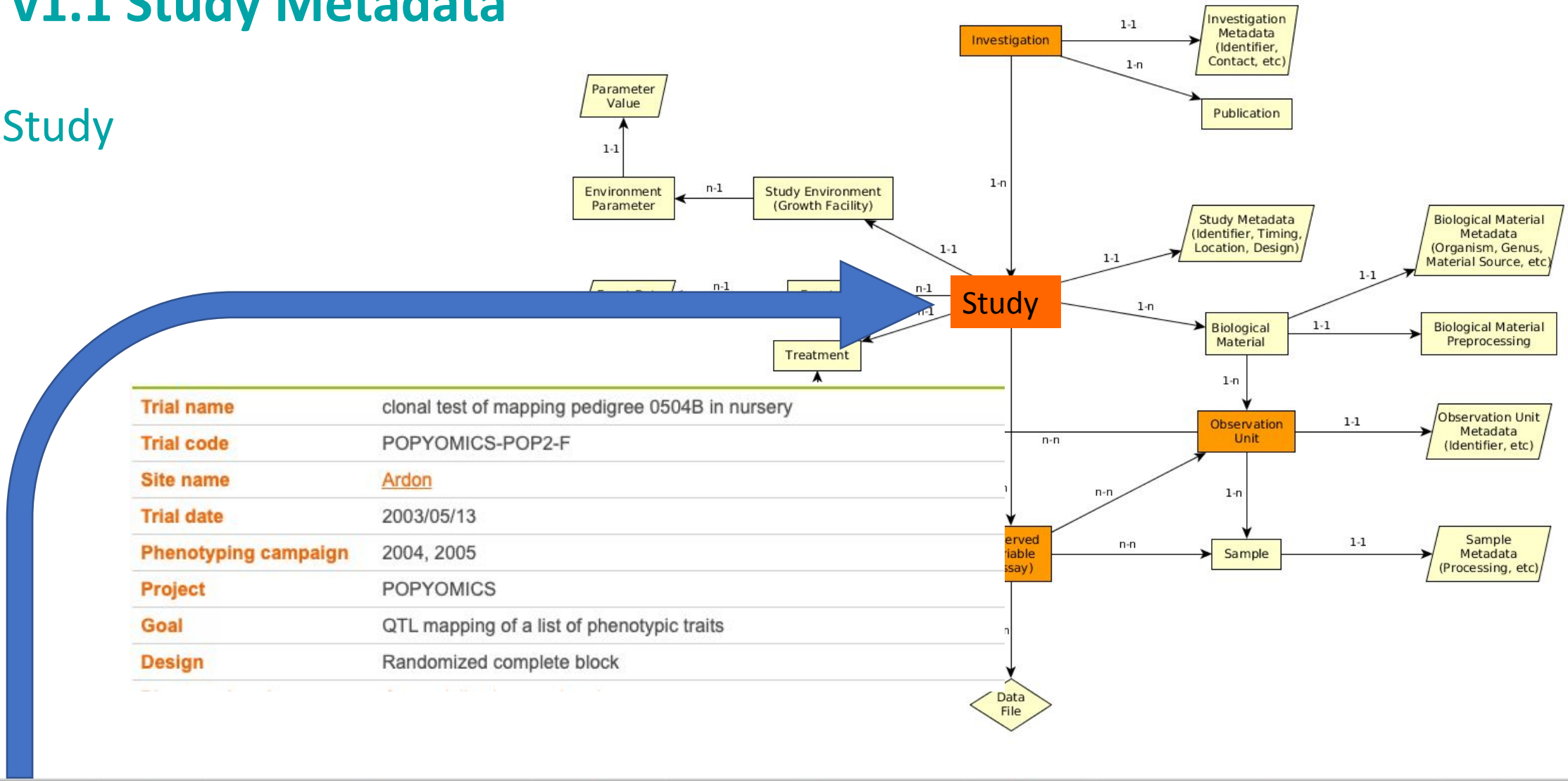
- **Examples of Studies:**

  - **MAIZE [1] : 37 studies: year x location x treatment (Gaillac 2012 rain, Gaillac 2013 watered, …)**

  - **POPLAR [2] : 3 locations over one to several years (Ardon_2003) (Ardon_2003-2005)**

MIAPPE v1.1 Study Metadata

- Poplar Study



Trial name	clonal test of mapping pedigree 0504B in nursery
Trial code	POPYOMICS-POP2-F
Site name	Ardon
Trial date	2003/05/13
Phenotyping campaign	2004, 2005
Project	POPYOMICS
Goal	QTL mapping of a list of phenotypic traits
Design	Randomized complete block

A	B	C	D	E	F	G	H	I
Accession Number	Expe location	Campaign	Circum1: Tree circumference at 1 year	Date [Circum1]	Height1: Tree total height at 1 year	Date [Height1]	Shoots3: Number of resprouts at 3 years	Date [Shoots3]
66130027	Ardon	2004	45.645632645603683	12/01/2004	284.3	12/01/2004		
66130027	Ardon	2005					14.630625	12/05/2005
66130044	Ardon	2004	38.96112577281653	12/01/2004	228.8	12/01/2004		
66130044	Ardon	2005					8.5030559999999991	12/05/2005

• MIAPPE main sections – Biological material

- **Biological material being studied**
- Plus its source (stock center, gen bank, etc…).
- ***Crucial*** for integrating phenotyping data with genomic or genetic data.

- **Metadata**
  - **Minimal fields from Multicrop Passport Descriptor (MCPD) standard**
  - **GPS location for forest tree / in situ material provenance**

Source of the material used: accession, cultivar/variety, region of provenance, laboratory cross, ...

MCPD identification system:
- Genebank/Lab + Species + accession number (mandatory)
- DOI

Biological material used in the study: seed lot, cuttings…

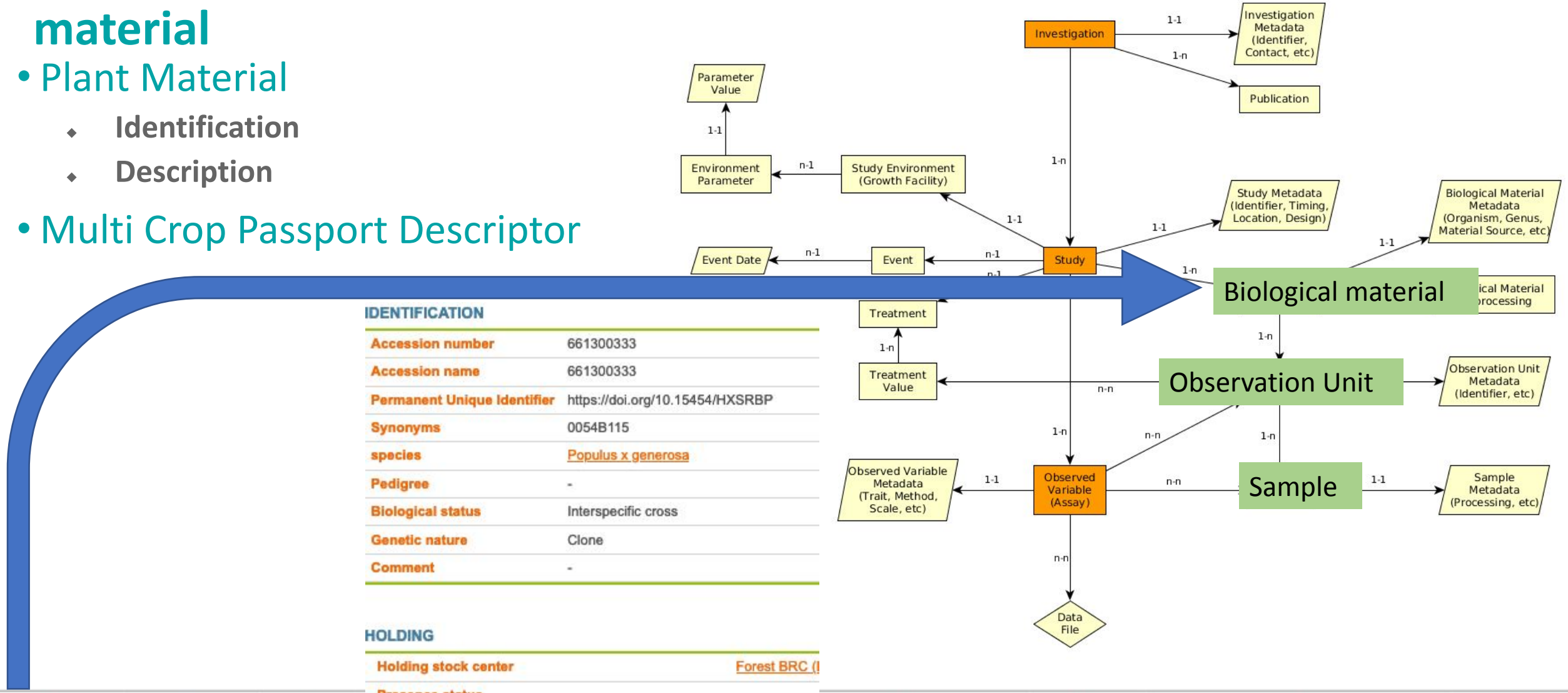
- Lab + internal accession number (mandatory)
- URI

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

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

MIAPPE main sections – Biological material

- Plant Material
  - Identification
  - Description
- Multi Crop Passport Descriptor



IDENTIFICATION

Accession number	661300333
Accession name	661300333
Permanent Unique Identifier	https://doi.org/10.15454/HXSRBP
Synonyms	0054B115
species	Populus x generosa
Pedigree	-
Biological status	Interspecific cross
Genetic nature	Clone
Comment	-

HOLDING

Holding stock center	Forest BRC (I
Presence status	

A	B	C	D	E	F	G	H	I
Accession Number	Trial Site	Campaign	Circum1: Tree circumference at 1 year	Date [Circum1]	Height1: Tree total height at 1 year	Date [Height1]	Shoots3: Number of resprouts at 3 years	Date [Shoots3]
661300270	Ardon	2004	45.645632645603683	12/01/2004	284.3	12/01/2004		
661300270	Ardon	2005					14.630625	12/05/2005
661300444	Ardon	2004	38.96112577281653	12/01/2004	228.8	12/01/2004		
661300444	Ardon	2005					8.5030559999999991	12/05/2005

• MIAPPE main sections – Observed Variable (Assay)

- **An Observed Variable (assay)**
  - **specific measurement.**
  - **Targets a trait**
  - **with a method and a scale.**
  - **Trait: Phenotype or Environment**

- **Metadata**
  - **Crop Ontology Trait Dictionary:**
  - **trait, method, scale**

- **Example:**
- In each study of Millet *et al.* [1] :
  - **Plant level: 20 variables**
  - **Genotype/ Study level: 19 variable both phenotype and environment**
  - **Plot level: 9 variables**
  - ***E.g. Female flowering days to silking D20deg***
  - **E.g. Plant height (cm)**

FFLW_D20deg: Female flowering days to silking D20deg VARIABLE

Synonyms	Female flowering days to silking D20deg FFLW
Growth stage	Flowering
Crop	Maize

Silking time TRAIT

Identifier	CO_322:0000031
Name	Silking time
Description	Silking time
Synonyms	Female flowering time
Main abbreviation	Silk
Alternative abbreviations	S FFlw
Entity	Flower
Attribute	Silking time
Class	Phenological

Thermal time between emergence and silking – Computation METHOD

Identifier	MIPO:0000027
Name	Thermal time between emergence and silking – Computation
Description	Calculated as equivalent days at 20 °C unit between emergence and 50% anthesis.
Reference	B. Parent, O. Turc, Y. Gibon, M. Stitt and F. Tardieu (2010) Modelling temperature-compensated physiological rates, based on the co-ordination of responses to temperature of developmental processes. Journal of Experimental Botany
Class	Computation

D20deg: days equivalent time at 20 °C SCALE

Identifier	MIPO:0000030
Name	D20deg: days equivalent time at 20 °C
Data type	Numerical

PTHT: Plant height (cm) VARIABLE

Ontology name	Maize Traits
Identifier	MIPO:0000006
Name	PTHT
Synonyms	Plant height (cm)
Xref	CO_322:0000994
Crop	Maize

Plant height TRAIT

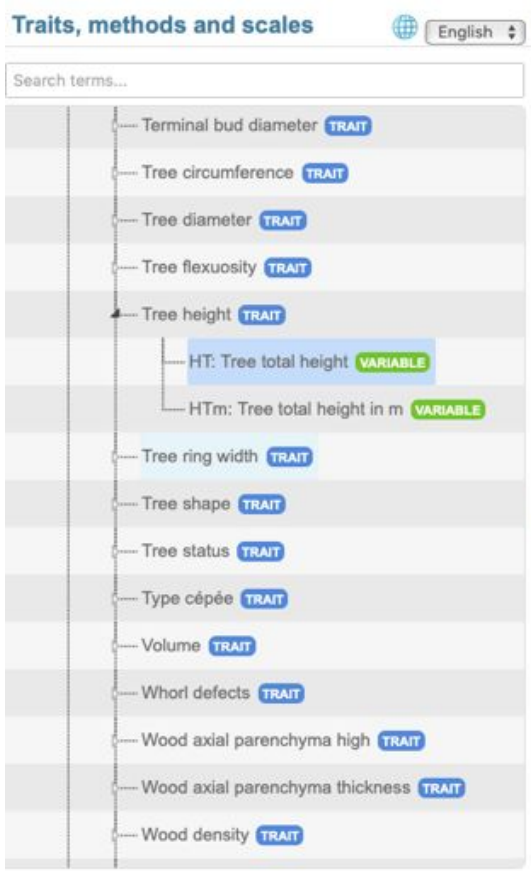
Identifier	CO_322:0000994
Name	Plant height
Description	Plant height from the base to the top part (in reproductive stages to the top of the tassel).
Main abbreviation	PH
Entity	Plant
Attribute	height
Class	Agronomical

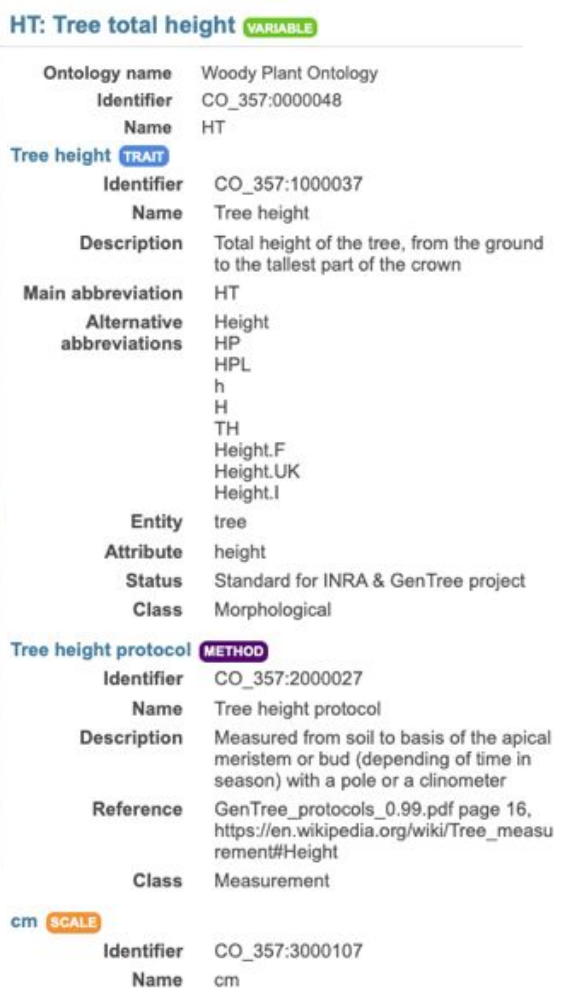
PH - Measurement METHOD

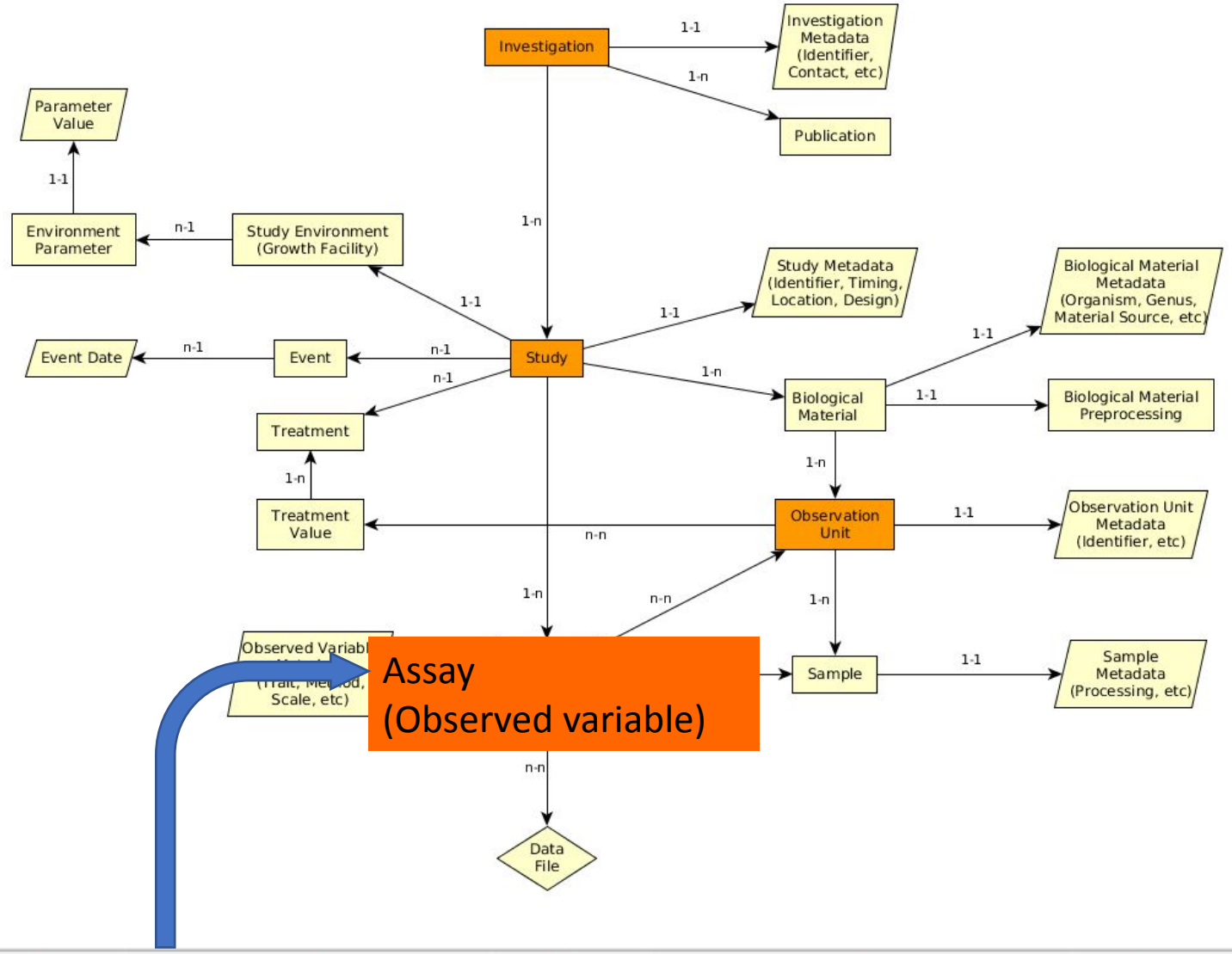
Identifier	CO_322:0000995
Name	PH - Measurement
Description	Recommended to take multiple plants and measure the height from the base of a plant to the top of the tassel, enter the data individually in the FieldBook and calculate the average.
Reference	DTMA drought phenotyping protocol. 2009. CIMMYT. Magorokosho et al. 2010. Characterization of maize germplasm grown in eastern and southern Africa: Results of the 2009 regional trials coordinated by CIMMYT. Zimbabwe

• MIAPPE v1.1 Metadata

Observed Variable: Trait + Method + Scale







A	B	C	D	E	F	G	H	I
Accession Number	Trial Site	Campaign	Circum1: Tree circumference at 1 year	Date [Circum1]	Height1: Tree total height at 1 year	Date [Height1]	Shoots3: Number of resprouts at 3 years	Date [Shoots3]
661300270	Ardon	2004	45.645632645603683	12/01/2004	284.3	12/01/2004		
661300270	Ardon	2005					14.630625	12/05/2005
661300444	Ardon	2004	38.96112577281653	12/01/2004	228.8	12/01/2004		
661300444	Ardon	2005					8.5030559999999991	12/05/2005

• MIAPPE main sections – Observation Unit, Samples

- **Observation units**
  - **objects in the study**
  - **Observations/measures are made on observation units**
  - **Treatments values are made on observation units**
  - **Also used for environmental variables.**

- **Metadata are specific to MIAPPE: identifiers, location, replication, treatments, …**

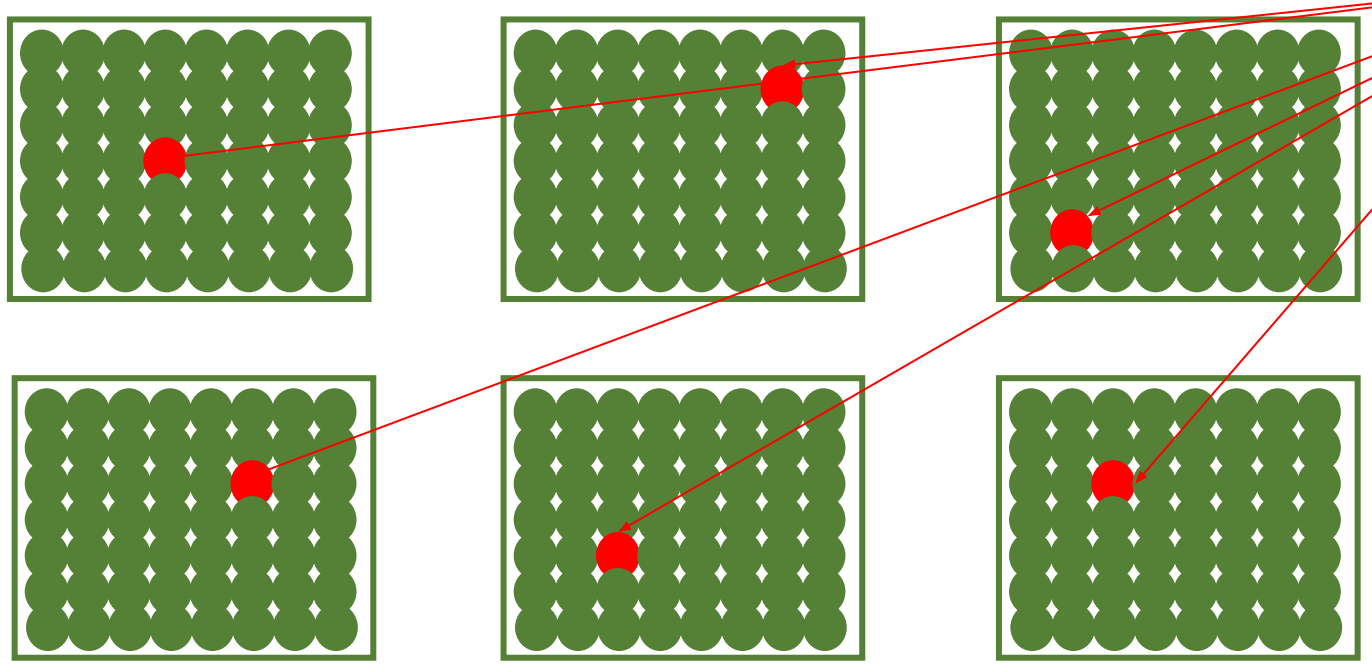
- **A sample**
- portion of plant tissue extracted from an observation unit
- sub-plant observations and/or molecular studies.

- **Metadata : 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

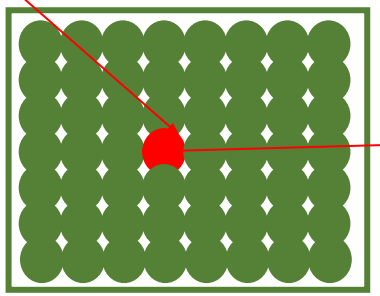


6 randomized blocks
1 observation unit = one tree
No treatment
6 replicates defined by their position in each block: row and column

• MIAPPE main sections – Observation Unit, Samples

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

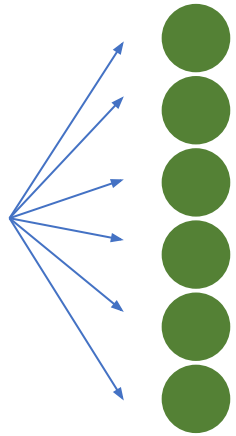
Populus deltoïdes
INRA:73028-62
doi:10.15454/1.4921786081024297E12
[observation unit] : Block1-Row4-Col4



For each tree of each block
= for each observation unit

One leaf

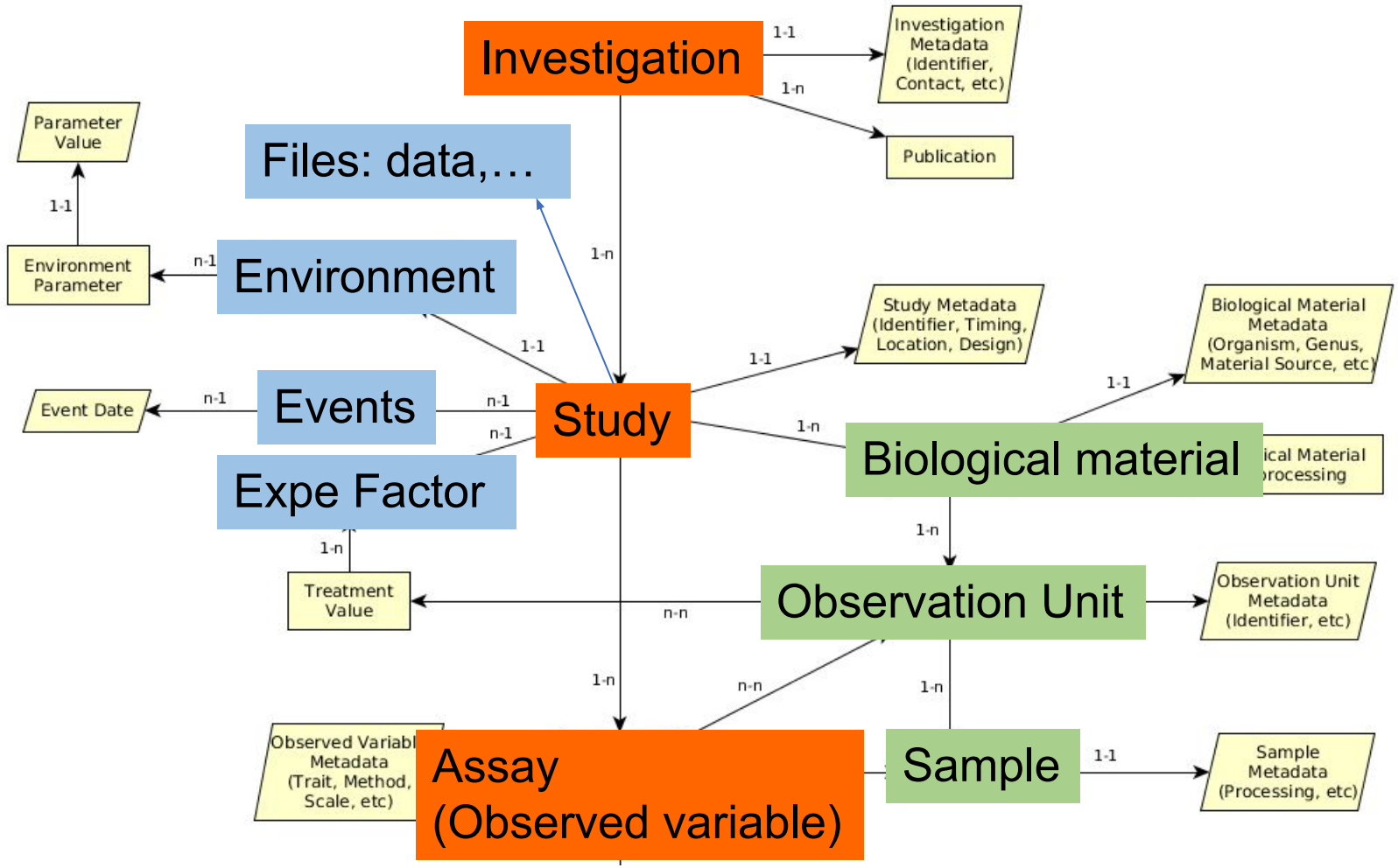
Set of samples



6 leaf disks

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 data model – Other Important sections



• MIAPPE main sections – Experimental Factor

- **Biotic or abiotic experimental factor**
  - **Its effects are evaluated in the study.**
  - **Takes different values in the observation units**
- **Metadata: name, description and value**
- **Example :** Maize dataset [2]
  - **Two Experimental factors/treatments: Rainfed, Watered**
  - **Identified by dedicated metadata**
  - **Block organisation**
  - **Study the effect of Environment/Drought on the Biological material**

GENOTYPE ID		TREATMENT			LEVEL	LEVEL	LEVEL
Accession Number	Accession Name	water_regime	Trial Name	Trial Site	BLOCK	PLOT	REPLICATE
FR19_H	FR19_H	rainfed	KWS Karlsruhe 2011	Karlsruhe	5	124	1
FR19_H	FR19_H	rainfed	KWS Karlsruhe 2011	Karlsruhe	10	250	3
FR19_H	FR19_H	rainfed	KWS Karlsruhe 2013	Karlsruhe	11	401503	2
FR19_H	FR19_H	watered	KWS Karlsruhe 2012	Karlsruhe	23	400353	1
FR19_H	FR19_H	watered	KWS Karlsruhe 2011	Karlsruhe	2	6	1
FR19_H	FR19_H	rainfed	KWS Karlsruhe 2012	Karlsruhe	16	401451	2

• MIAPPE main sections – Event

- **Event**
  - **Discrete occurrence at a particular time**
  - **Natural, e.g. rain, unwanted pathogen attack**
  - **Cultural practice, e.g. planting, watering, etc.**
  - **Whole study level or at the observation unit level**
  - **It is not the studied Factor but an additional information**
- **Metadata: name, description, time/date**

- **Examples:** In Poplar, *[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 *et al., 2012*	Rain	15mm of rain on the orchard	2012-03-15

• MIAPPE main sections – Environment

- **Environment parameters**
  - **Constant throughout the study**
  - **Did not change between observation units or assays.**
  - **Environment characteristics that vary over time, i.e. environmental variables, should be recorded as Observed Variables**

- **Metadata :name, value**

- **Example:**

Environment parameter	Environment parameter Value
Sowing density	20 seeds/m2
Rooting medium composition	Ca (XEO:00058): 5 mg/L; ...

Take home message

Semantic

- Description of the data
- Controlled vocabularies: term name and definitions
- Ontologies: semantic links between terms
- *Biologist* driven

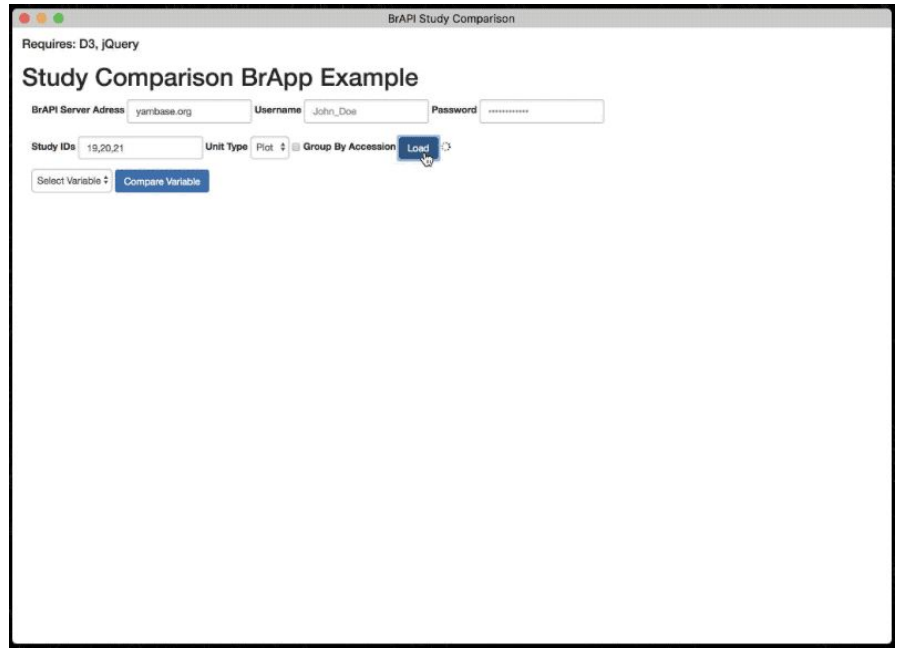
Structure

- Formatting and Organizing the data
- Data Models
- Standards : CSV, VCF, GFF, MIAPPE (www.miappe.org) , etc…
- *Biologist & Computer scientist* driven

Persistent Unique Identifiers
accessions ID, Trait ID, DOI,…

Technical

- Data integration and sharing
- Interoperability : tools and systems
  - GA4GH
  - Breeding API www.brapi.org
- *Computer scientist* driven



Take Home message

Standardizing your data is

-Good for you

-Good for others

-Getting easier

-Contribution Welcomed !

Thank you!