EMPHASIS

Data Management
S4: Advanced data management (1)
Variables & Measurements

Thursday 07/12/2023 - 14:00-16:00 (CET)

Isabelle Alic, Farzaneh Kazemipour-Ricci - INRAE

Sylvain Poque - University of Helsinki

General objectives: Overview of data management for plant phenotyping - focus on FAIR data

Session 4

Advanced data management (1):

Variables & Measurements

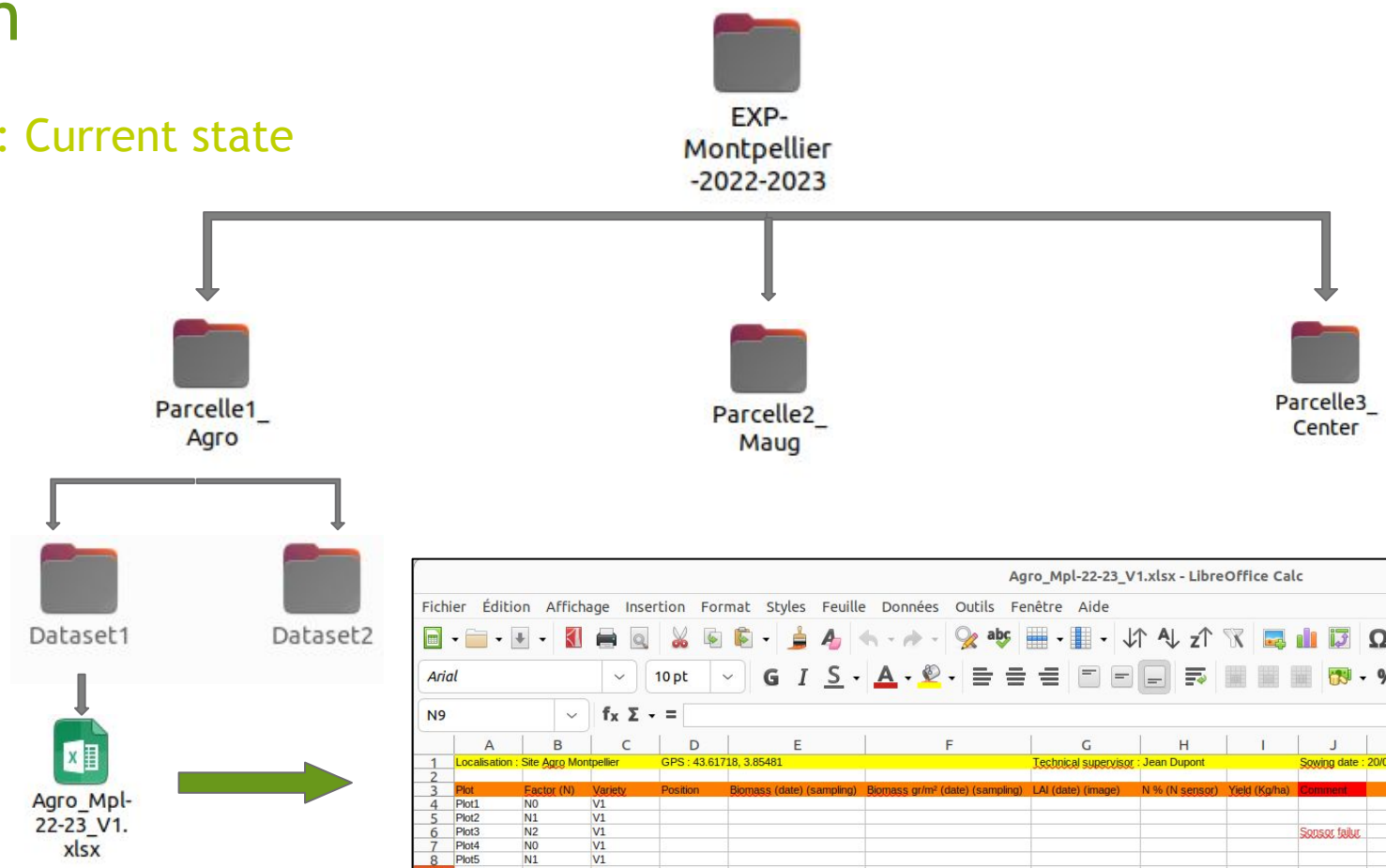
Overview

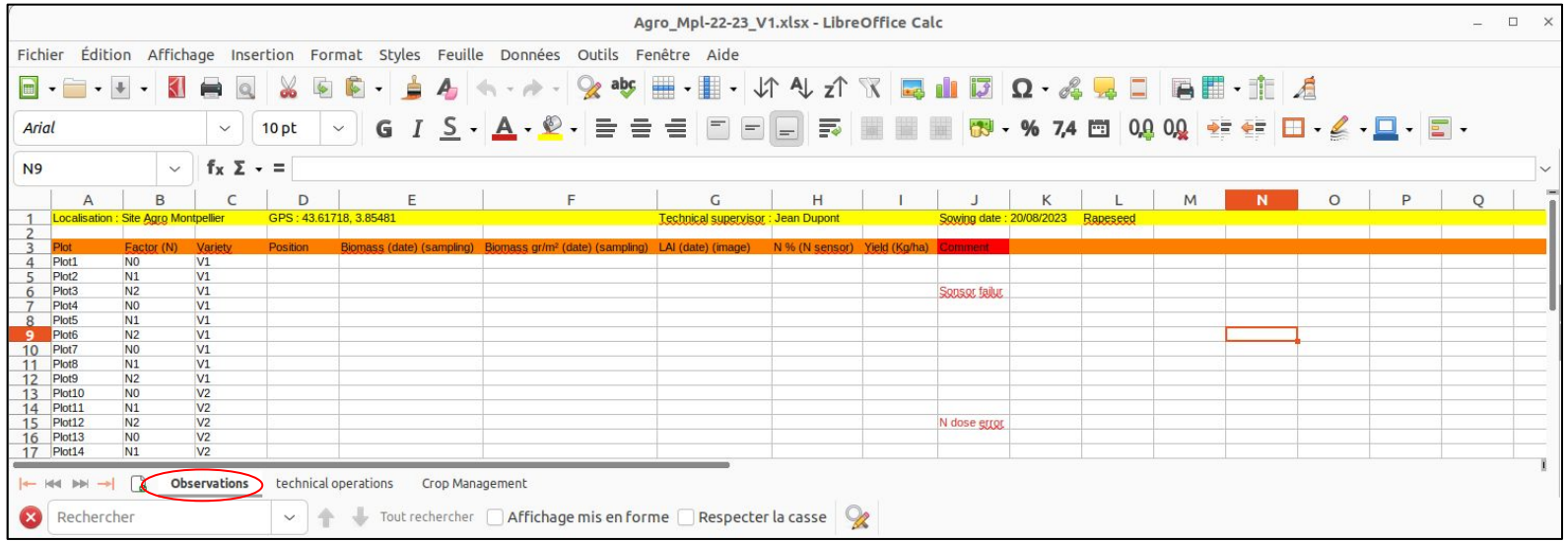
Advanced data management Variables & Measurements (1)

- Reminder of concepts
- Measured data, Variables & Provenance
- Data Pipeline, Automatized data insertion
- Use Case: Sending data to PHIS

Introduction

Experimental Data: Current state

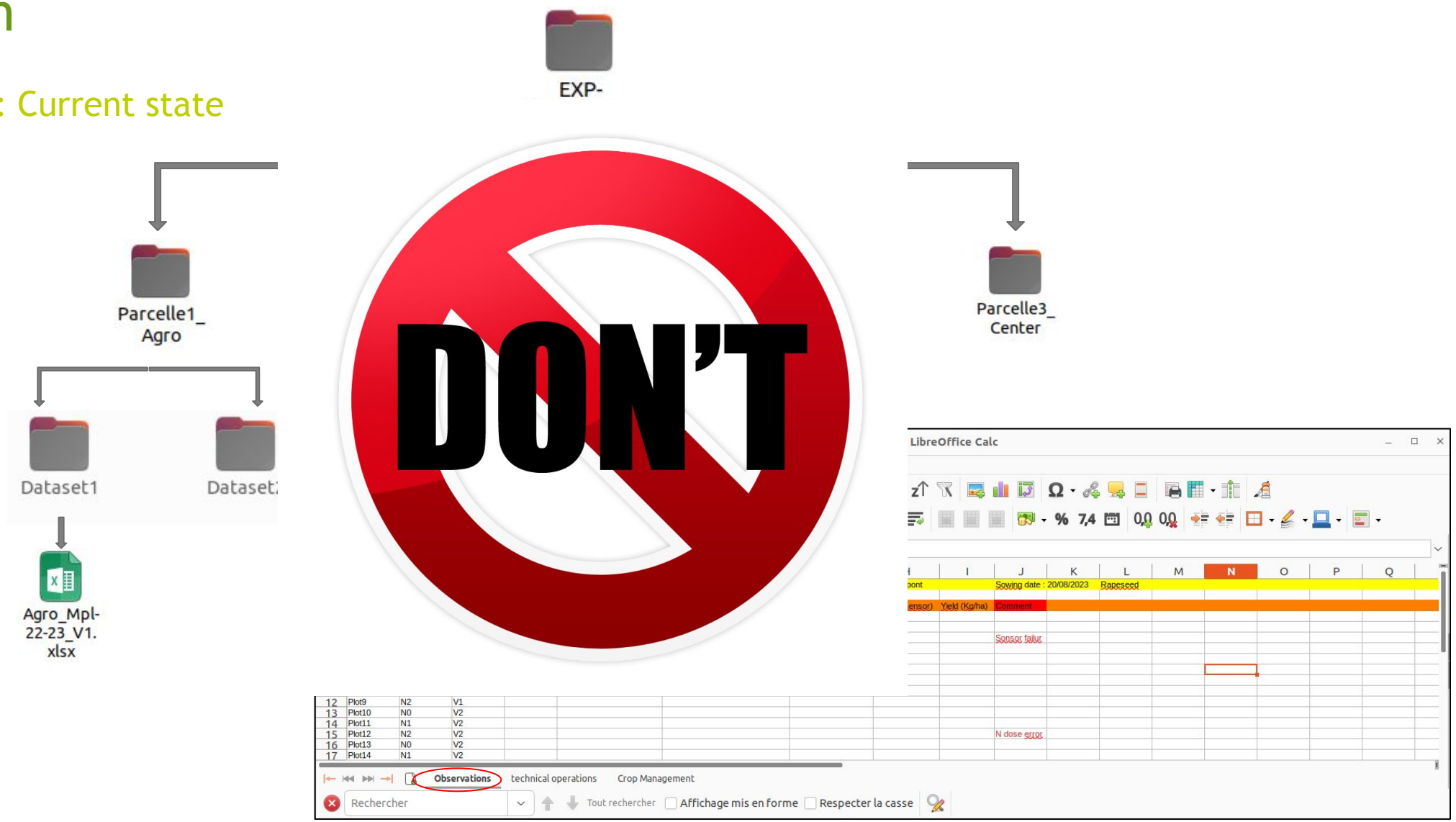




	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Localisation : Site Agro Montpellier			GPS : 43.61718, 3.85481			Technical supervisor : Jean Dupont			Sowing date : 20/08/2023	Rapeseed						
2																	
3	Plot	Factor (N)	Variety	Position	Biomass (date) (sampling)	Biomass gr/m² (date) (sampling)	LAI (date) (image)	N % (N sensor)	Yield (Kg/ha)	Comment							
4	Plot1	N0	V1														
5	Plot2	N1	V1														
6	Plot3	N2	V1							Sensor failur							
7	Plot4	N0	V1														
8	Plot5	N1	V1														
9	Plot6	N2	V1														
10	Plot7	N0	V1														
11	Plot8	N1	V1														
12	Plot9	N2	V1														
13	Plot10	N0	V2														
14	Plot11	N1	V2														
15	Plot12	N2	V2							N dose error							
16	Plot13	N0	V2														
17	Plot14	N1	V2														

Introduction

Experimental Data: Current state



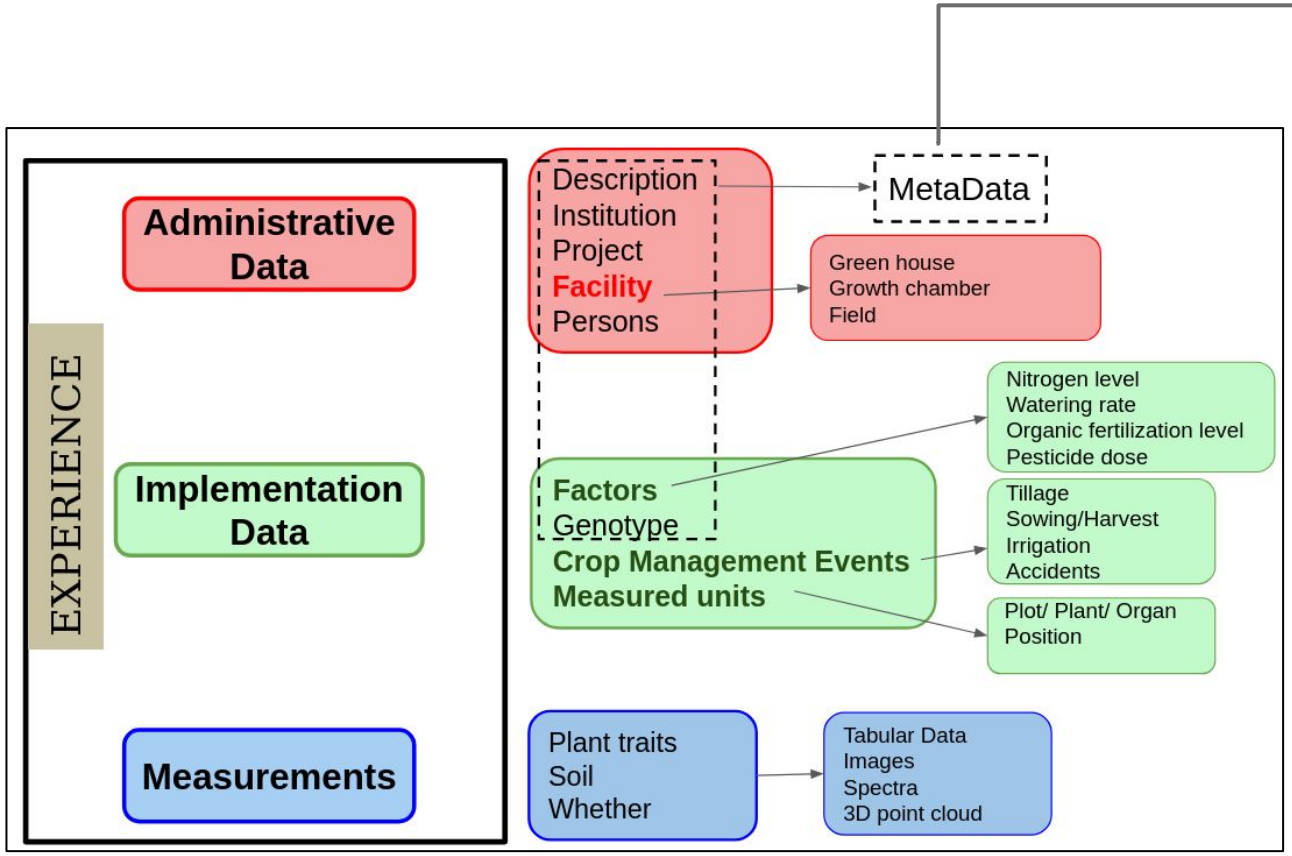
Introduction

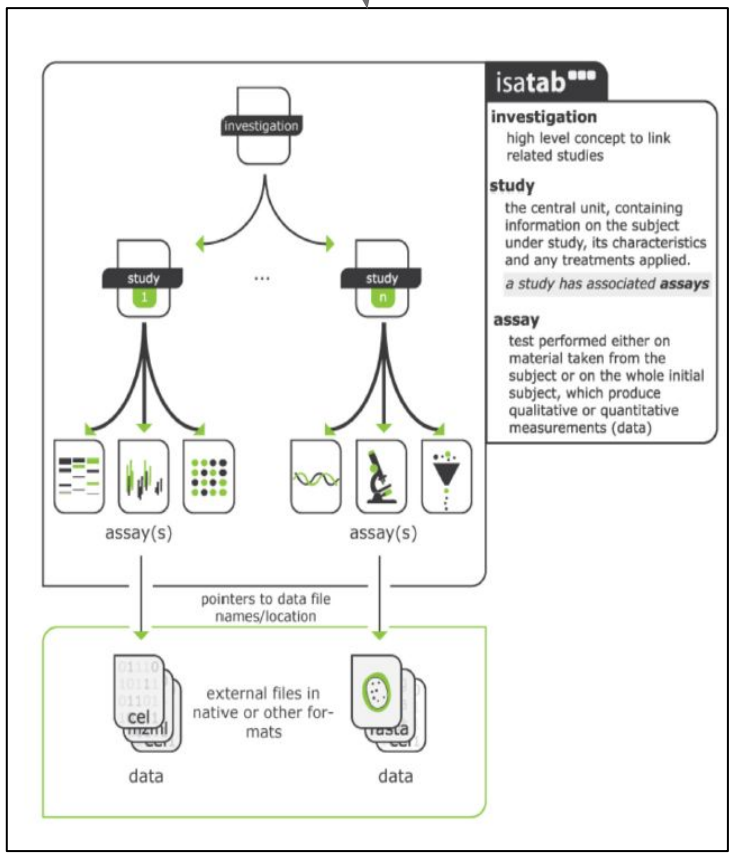
Reproducible science levels

- **Re-Run:** same lab, same experiments, redo analytics (checking)
- **Repeat:** same lab, redo experiments and analytics
- **Transparency:** same experiments, other lab redo analytics (proof)
- **Replicate, Reproduce:** other lab redo experiments and analytics (robustness)
- **Reuse:** for different scientific results and/or other contexts

Introduction

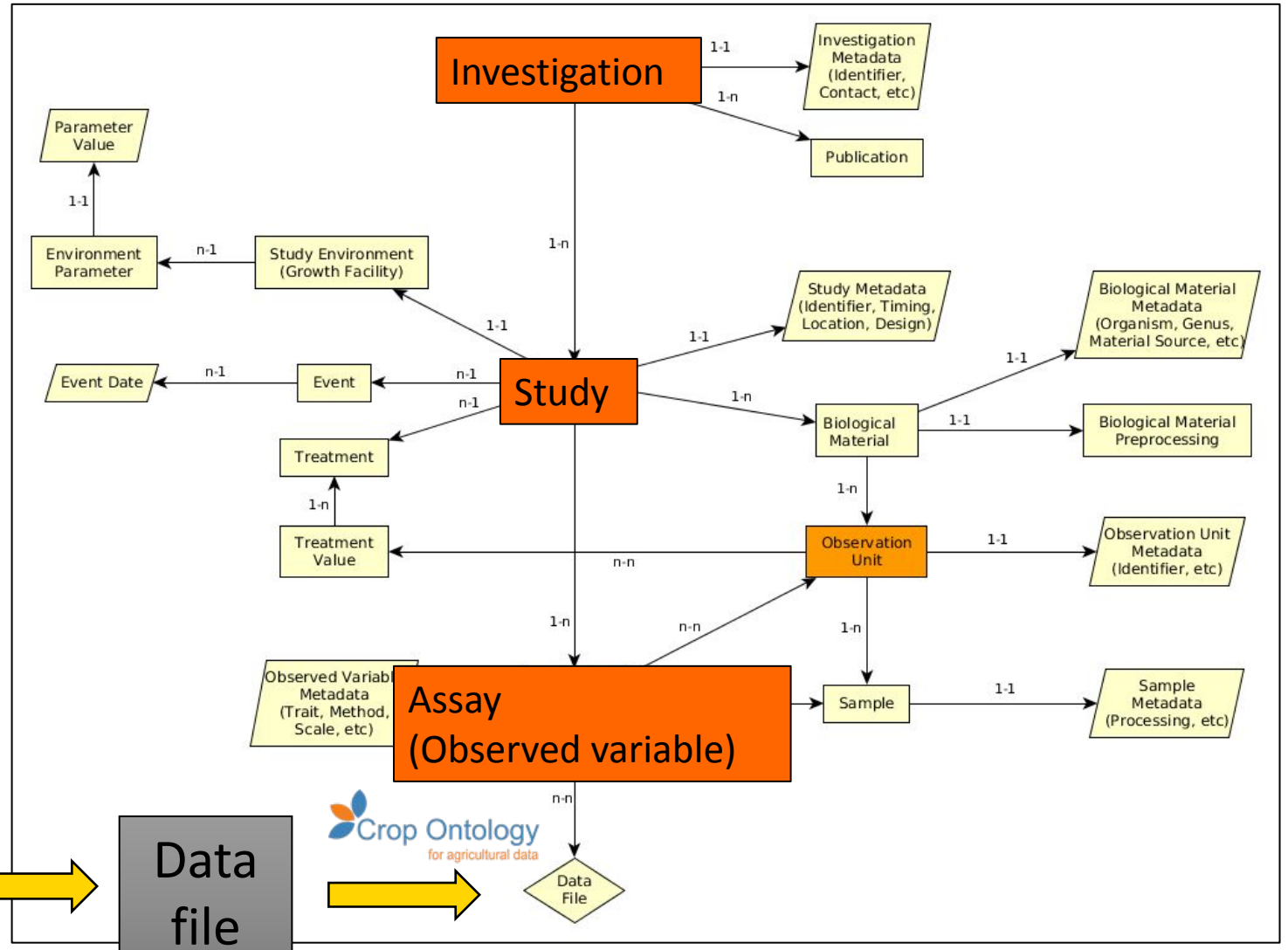
Data structure





Introduction

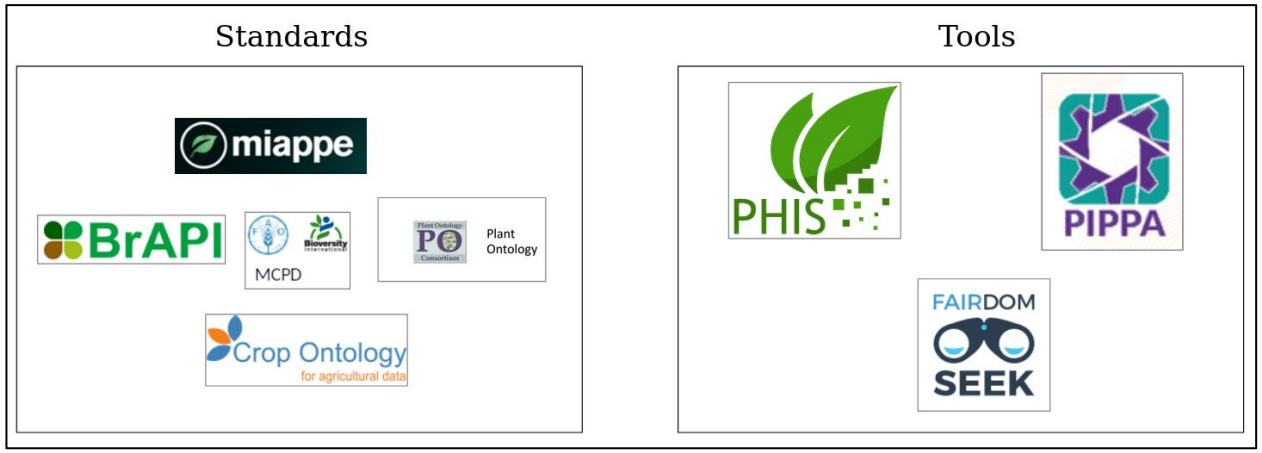
Data structure & Standards

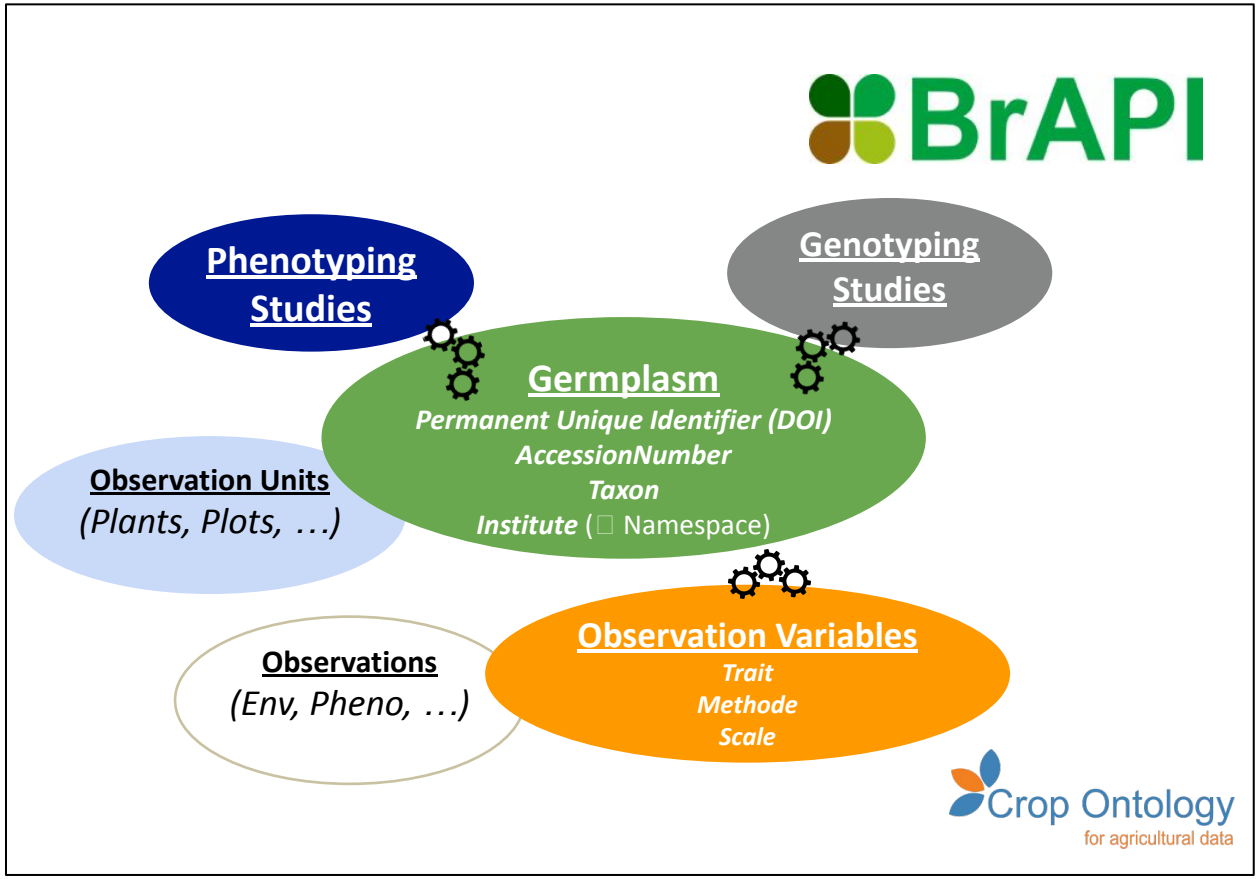


Introduction

Use Standards, Compatible tools and APIs

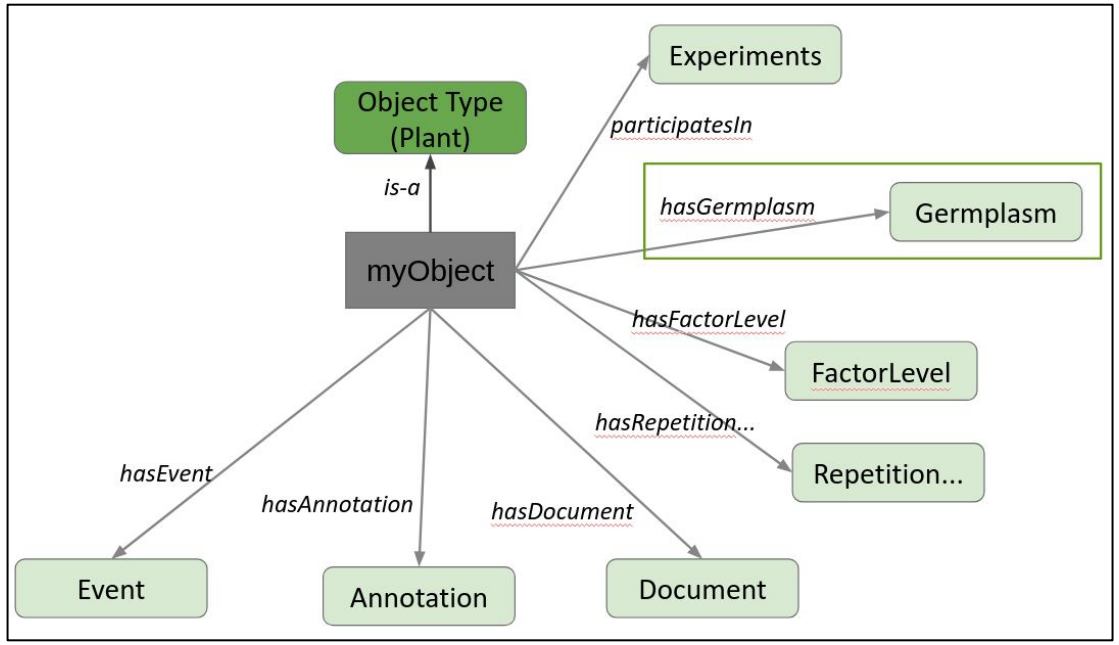
- Standard Open Web Service API
- Information Exchange, Main target: Breeding

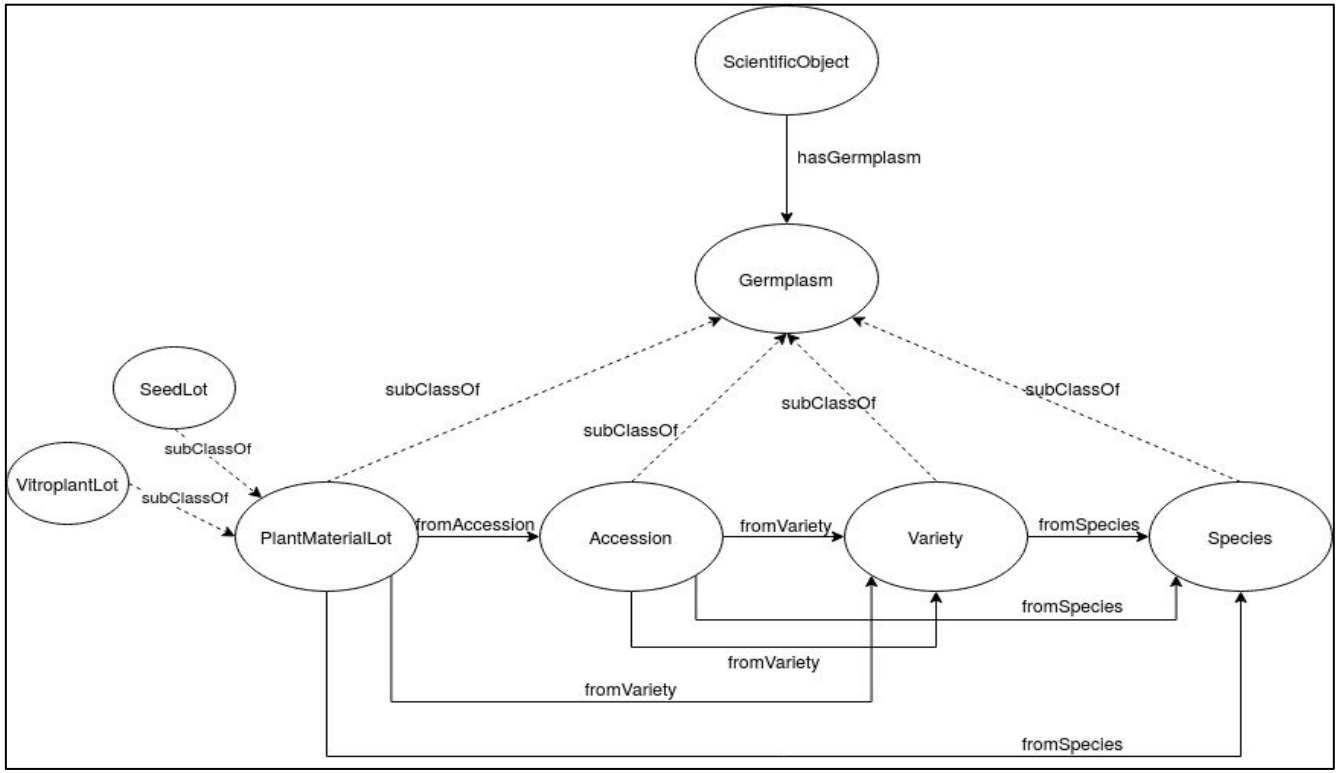




Introduction

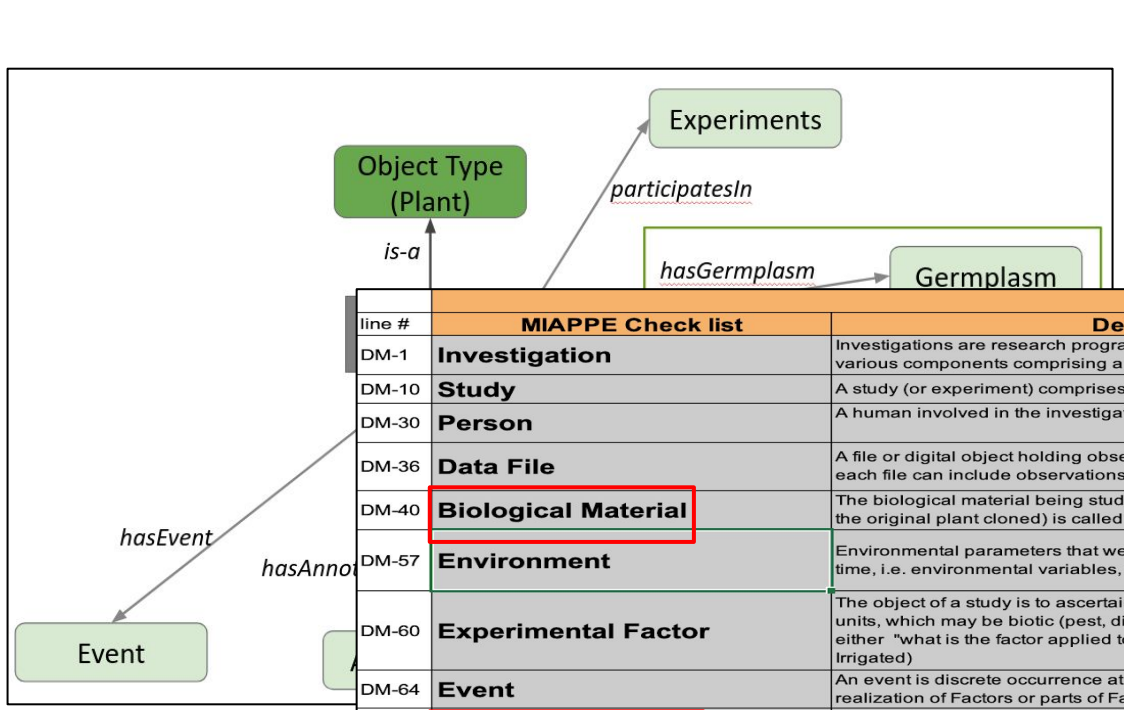
MIAPPE/BrAPI Compliant S.I.

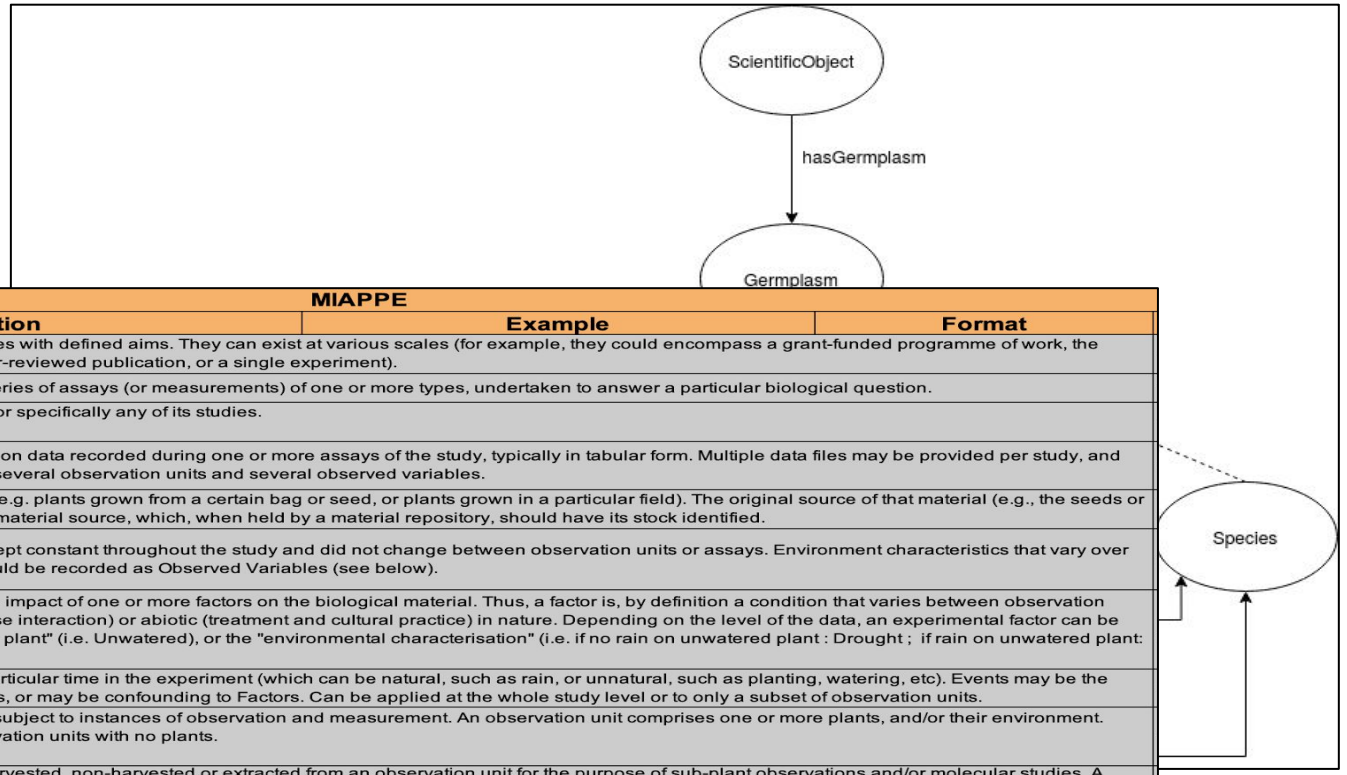




Introduction

MIAPPE/BrAPI Compliant S.I.





	MIAPPE			
line #	MIAPPE Check list	Definition	Example	Format
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).		
DM-10	**Study**	A study (or experiment) comprises a series of assays (or measurements) of one or more types, undertaken to answer a particular biological question.		
DM-30	**Person**	A human involved in the investigation or specifically any of its studies.		
DM-36	**Data File**	A file or digital object holding observation data recorded during one or more assays of the study, typically in tabular form. Multiple data files may be provided per study, and each file can include observations for several observation units and several observed variables.		
DM-40	**Biological Material**	The biological material being studied (e.g. plants grown from a certain bag or seed, or plants grown in a particular field). The original source of that material (e.g., the seeds or the original plant cloned) is called the material source, which, when held by a material repository, should have its stock identified.		
DM-57	**Environment**	Environmental parameters that were kept constant throughout the study and did not change between observation units or assays. Environment characteristics that vary over time, i.e. environmental variables, should be recorded as Observed Variables (see below).		
DM-60	**Experimental Factor**	The object of a study is to ascertain the impact of one or more factors on the biological material. Thus, a factor is, by definition a condition that varies between observation units, which may be biotic (pest, disease interaction) or abiotic (treatment and cultural practice) in nature. Depending on the level of the data, an experimental factor can be either "what is the factor applied to the plant" (i.e. Unwatered), or the "environmental characterisation" (i.e. if no rain on unwatered plant : Drought ; if rain on unwatered plant: Irrigated)		
DM-64	**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). Events may be the realization of Factors or parts of Factors, or may be confounding to Factors. Can be applied at the whole study level or to only a subset of observation units.		
DM-69	**Observation Unit**	Observation units are objects that are subject to instances of observation and measurement. An observation unit comprises one or more plants, and/or their environment. There can be pure environment observation units with no plants. Synonym: Experimental unit.		
DM-75	**Sample**	A sample is a portion of plant tissue harvested, non-harvested or extracted from an observation unit for the purpose of sub-plant observations and/or molecular studies. A sample must be used when there is a physical sample that needs to be stored and traced. Otherwise, observations made at the sub-plant level should be recorded as plant level observations using the observed variables to characterize the object of the observation (e.g. Berry sugar content, Fruit weight, Grain Protein content, Leaf 1 width, Leaf 2 width, Leaf 2 length).		
DM-82	**Observed Variable**	An observed variable describes how a measurement has been made. It typically takes the form of a measured characteristic of the observation unit (plant or environmental trait), associated to the method and unit of measurement. Multiple variables with the same combination of trait, method and scale can be used in association with different plant parts (leaf 1, leaf 2), when this distinction is necessary for observations referring to different parts of the same observation unit.		

Context

Compared to 20 years ago…

- Number and diversity of data sources has increased
- Stronger complexity of data pipelines to be designed has increased

Increasing the difficulties for a reproducible and open science

Makes hard to understand how and under what conditions the data were produced

We need to know about **data provenance**

We need to know about **method of data production**

Variables - Definition

Variable is the key element of data structuring and analytics

A variable can be represented by a **combination of descriptive components giving meaning to the value** derived from a **data acquisition** event be it **an observation, a measurement, a simulation or a calculation.**

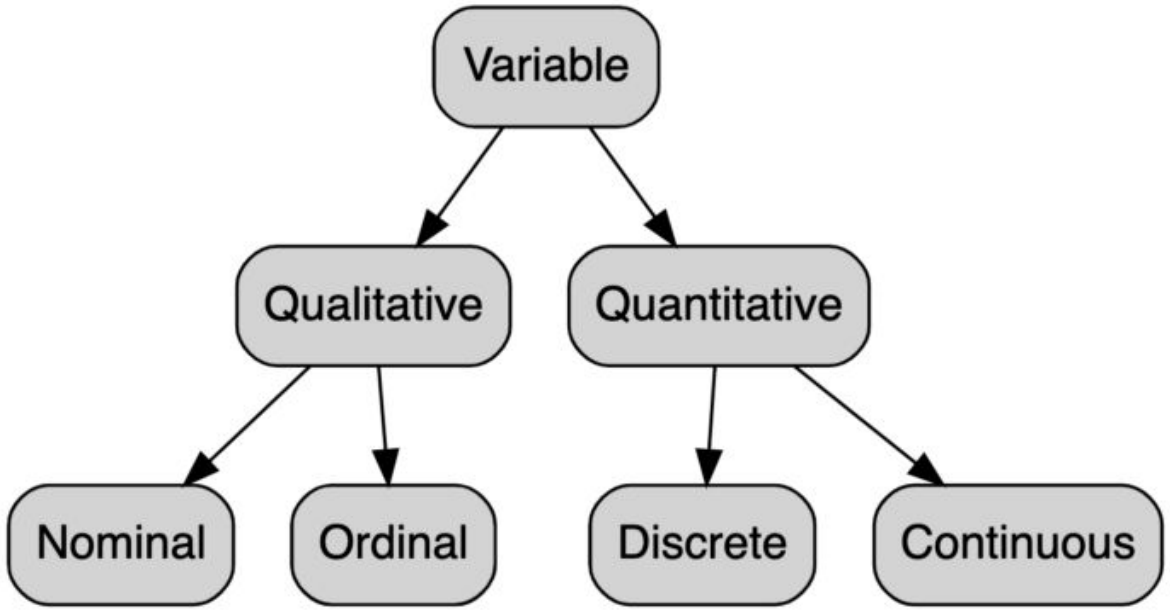


Image: Moira Sheehan

Variables - Definition

Reality is that too often…

- **No variable ID**
- Same name for different variables
- Different names for same variable
- No (or not machine readable) variable description
- Unstable variable

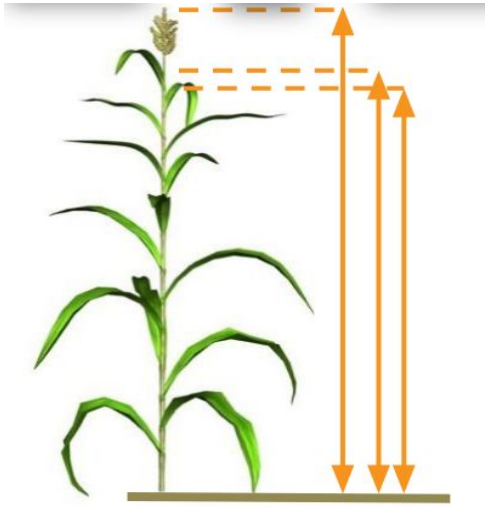


Figure From D. Pot, CIRAD

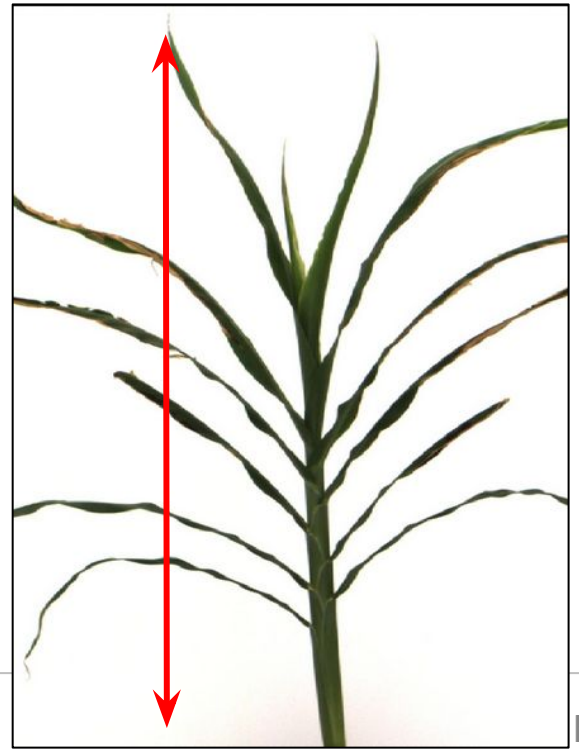


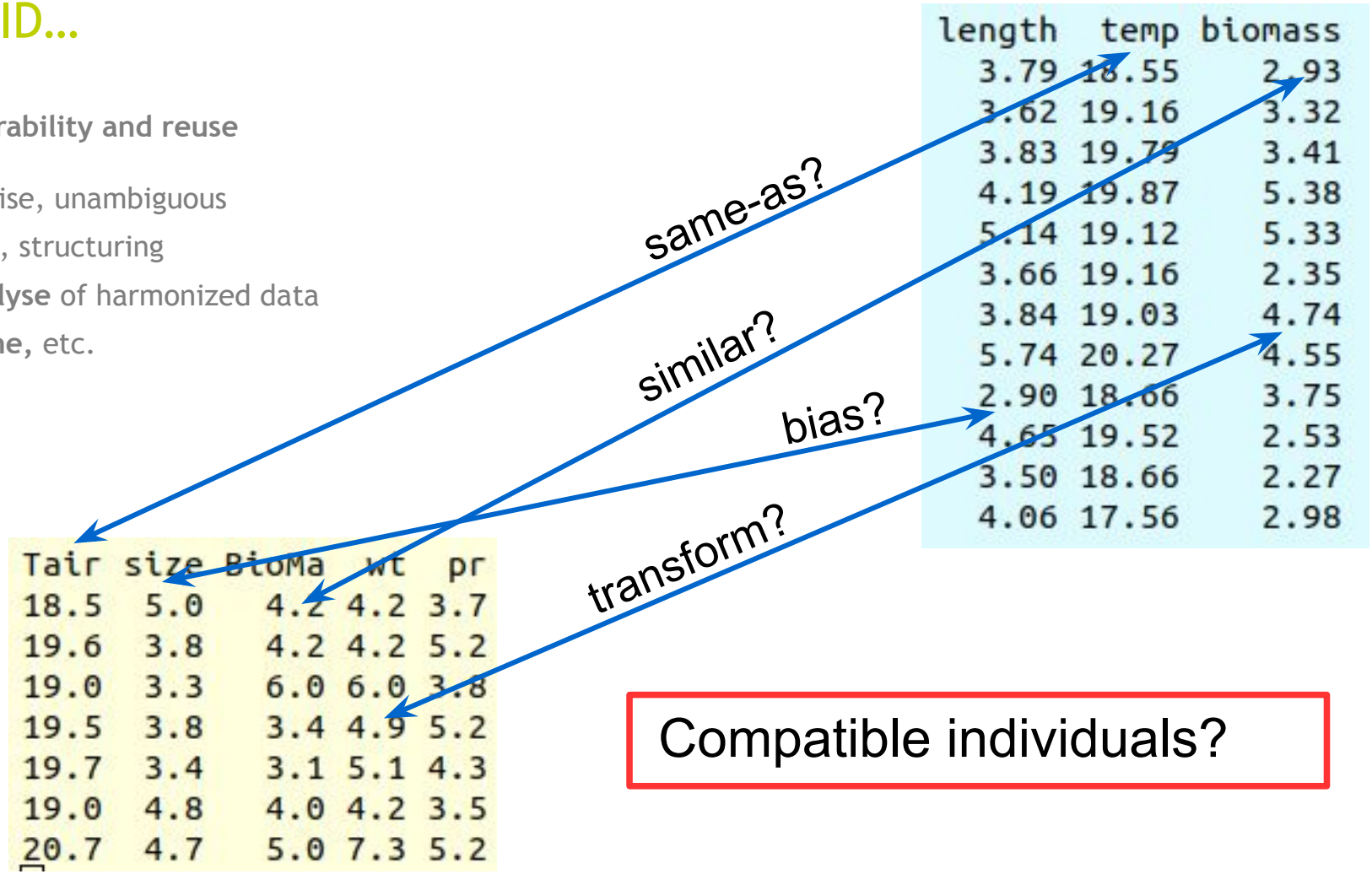
Image: Moira Sheehan

Variables - Definition

Formalized variable with ID…

…Aims FAIR data with a **focus on interoperability and reuse**

- Provides a **description**: simple, precise, unambiguous
- **Share/Reuse variables:** descriptions, structuring
- Make it easier to **aggregate and analyse** of harmonized data
- Be able to **compare, merge, combine,** etc.



Covariables?

Variables that affect a variable

Compatible individuals?

Variable Model

How do you formalise the variables?

Requirements for formalising of variables:

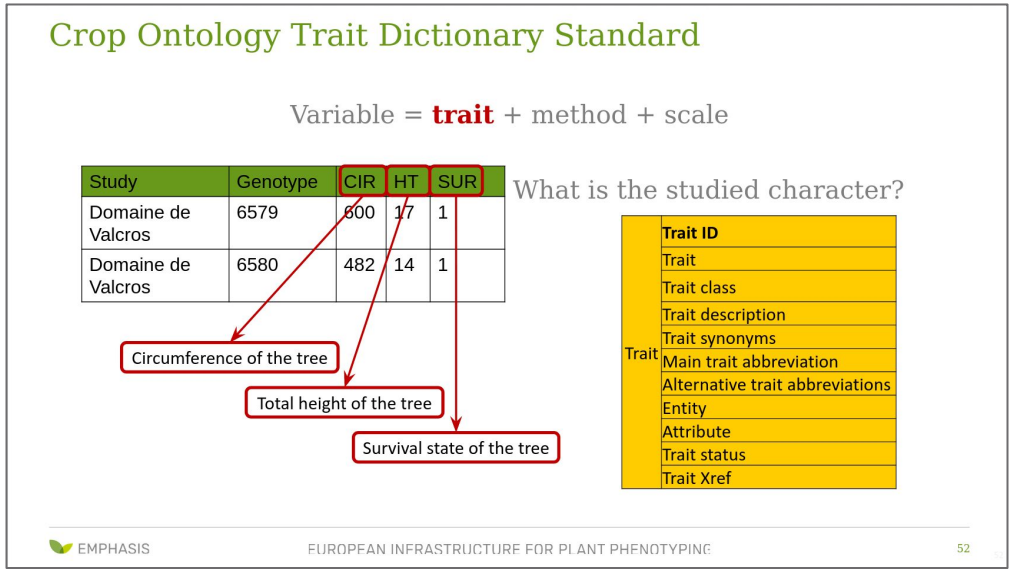
- Focus on **WHAT, HOW, (WHERE, WHEN)**
- **In-depth understanding (area of expertise)**
- Step back from the variable, its use, its usefulness and its potential second use
- Provide a human-understandable definition

Variable Model

We can create an infinite number of variables!

Need to formalise variables:

- that anyone can easily understand
- that can be manipulated with machine



A Common Model in Plant Sciences

- **Trait** characteristic of an organism
- **Method** is the way we have to measure variable. Giving description of protocols and type of instruments used.
- **Unit** is an element from the Ontology of Units of Measurement

Variable Model

We can create an infinite number of variables!

Need to formalise variables:

- that anyone can easily understand
- that can be manipulated with machine

A Common Model in Plant Sciences

Trait

- **Entity** is the target of the variable, what we are measuring on
- **Characteristic** is the observed property or physical quantity or quality.

- **Method** is the way we have to measure variable. Giving description of protocols and type of instruments used.
- **Unit** is an element from the Ontology of Units of Measurement

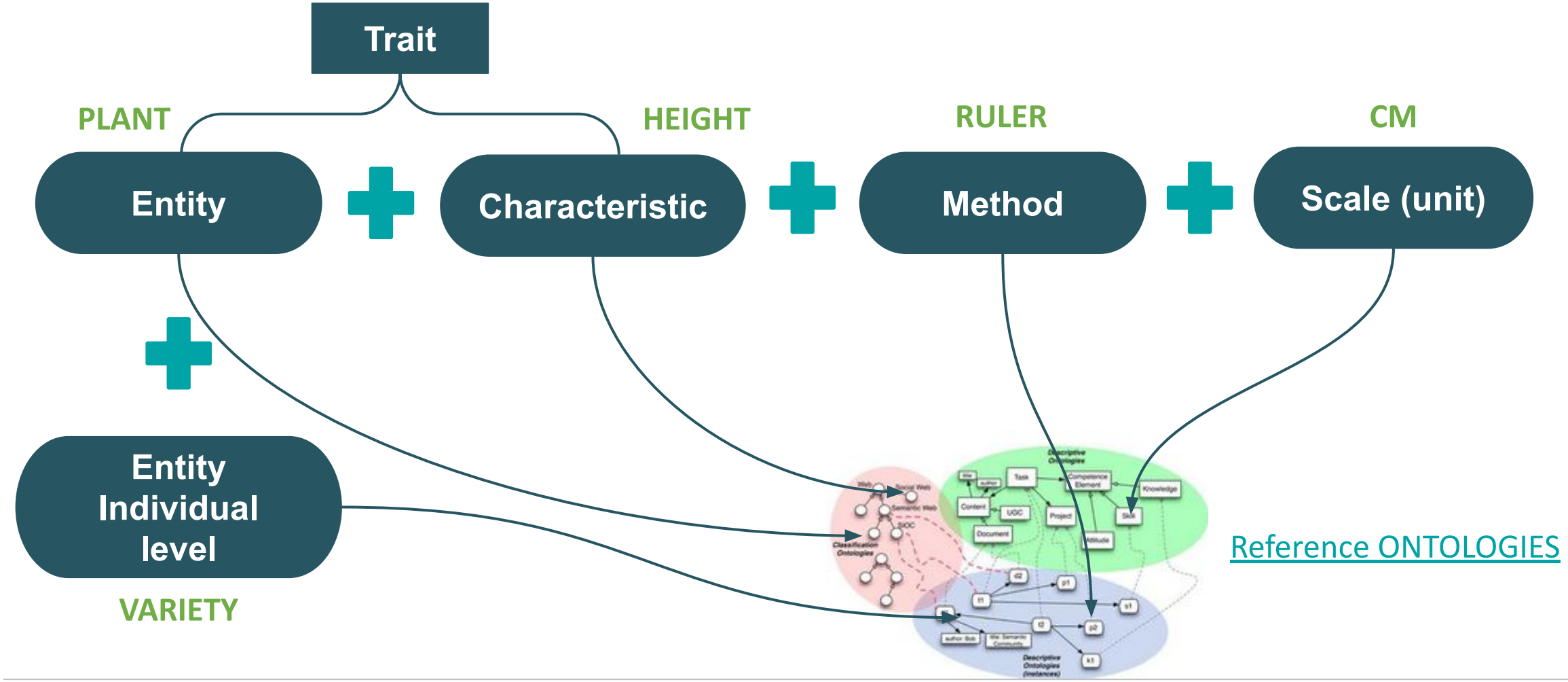
Variable Model

How to implement this model properly

- Enable semantically precise and **FAIR descriptions**
- Break down the description into **standardised atomic elements**
- **Link** these elements to **existing vocabularies/ontologies**
- Make the **description machine-readable**

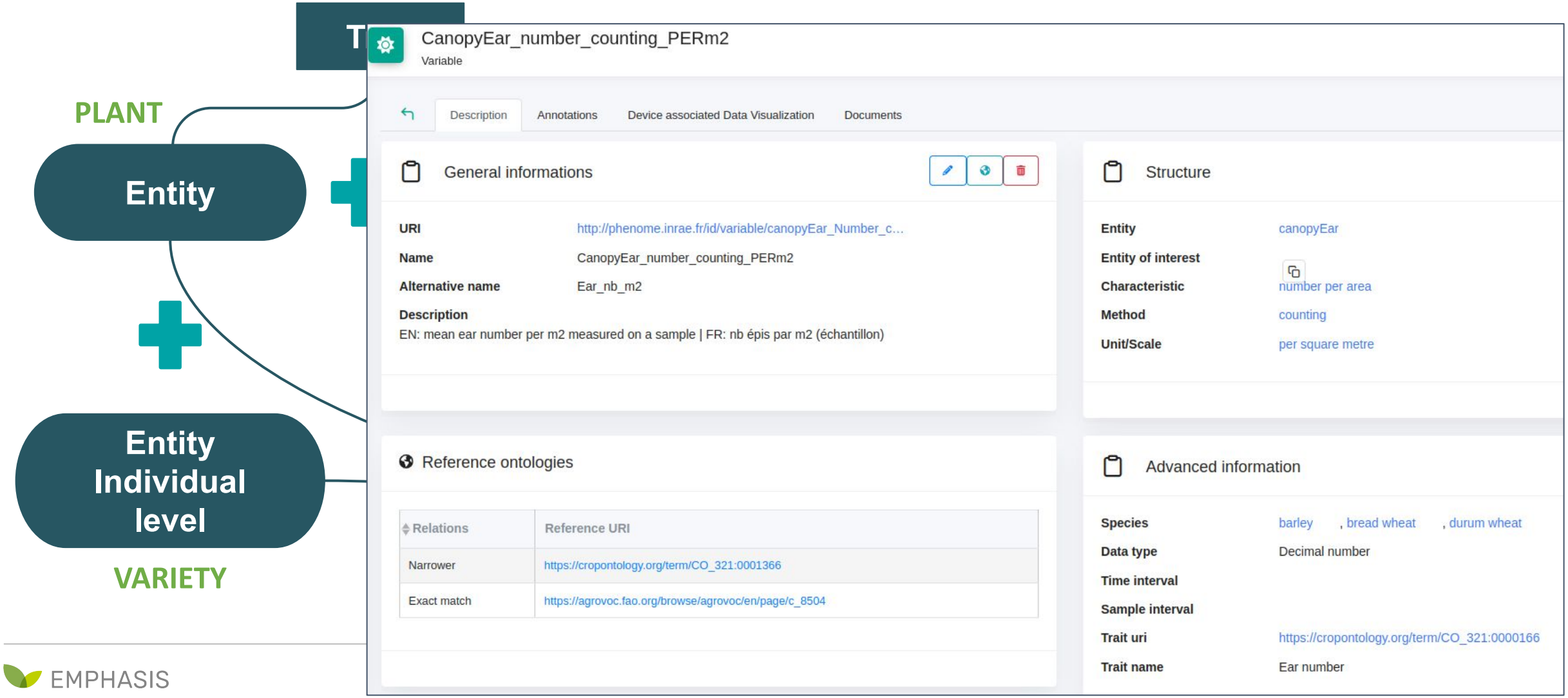
Variable Model

How to implement this model properly



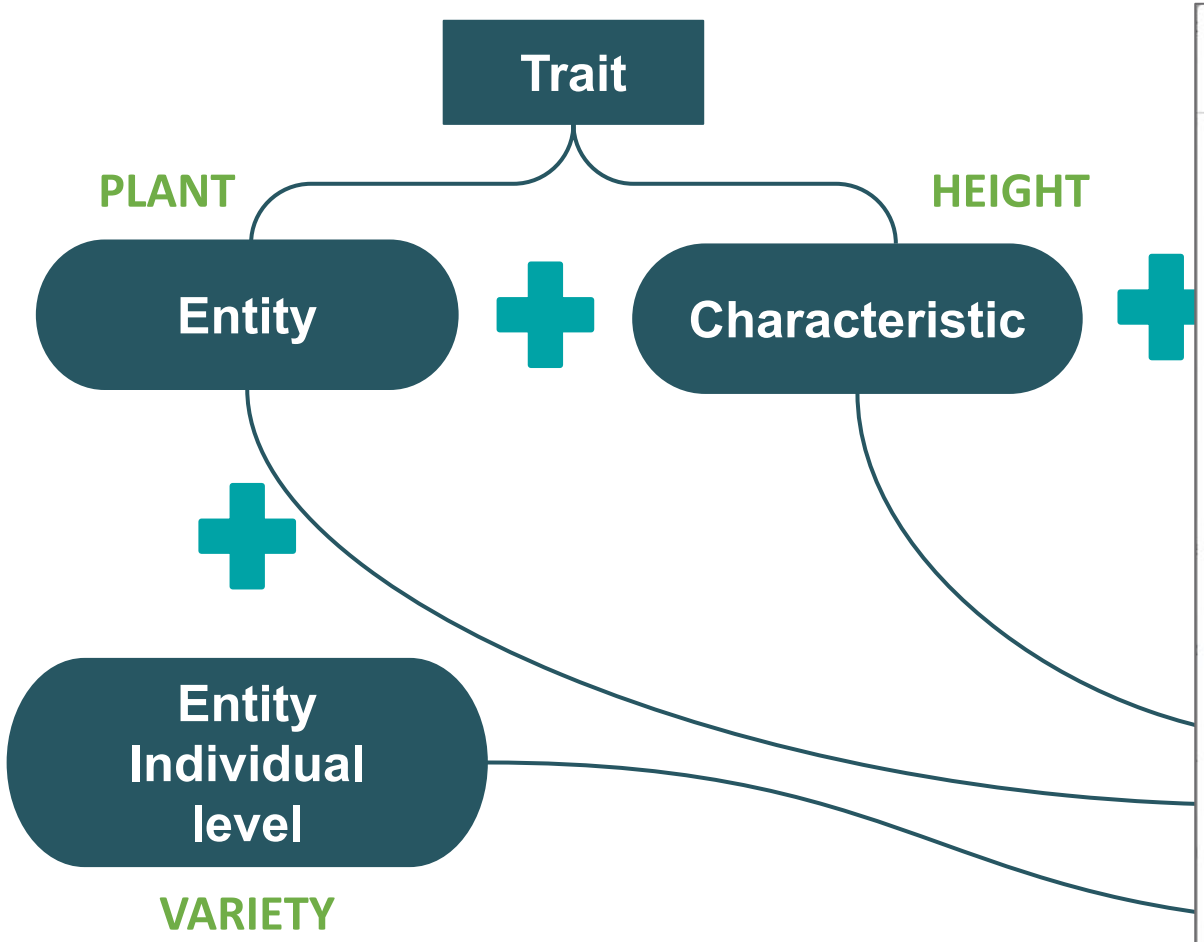
Variable Model

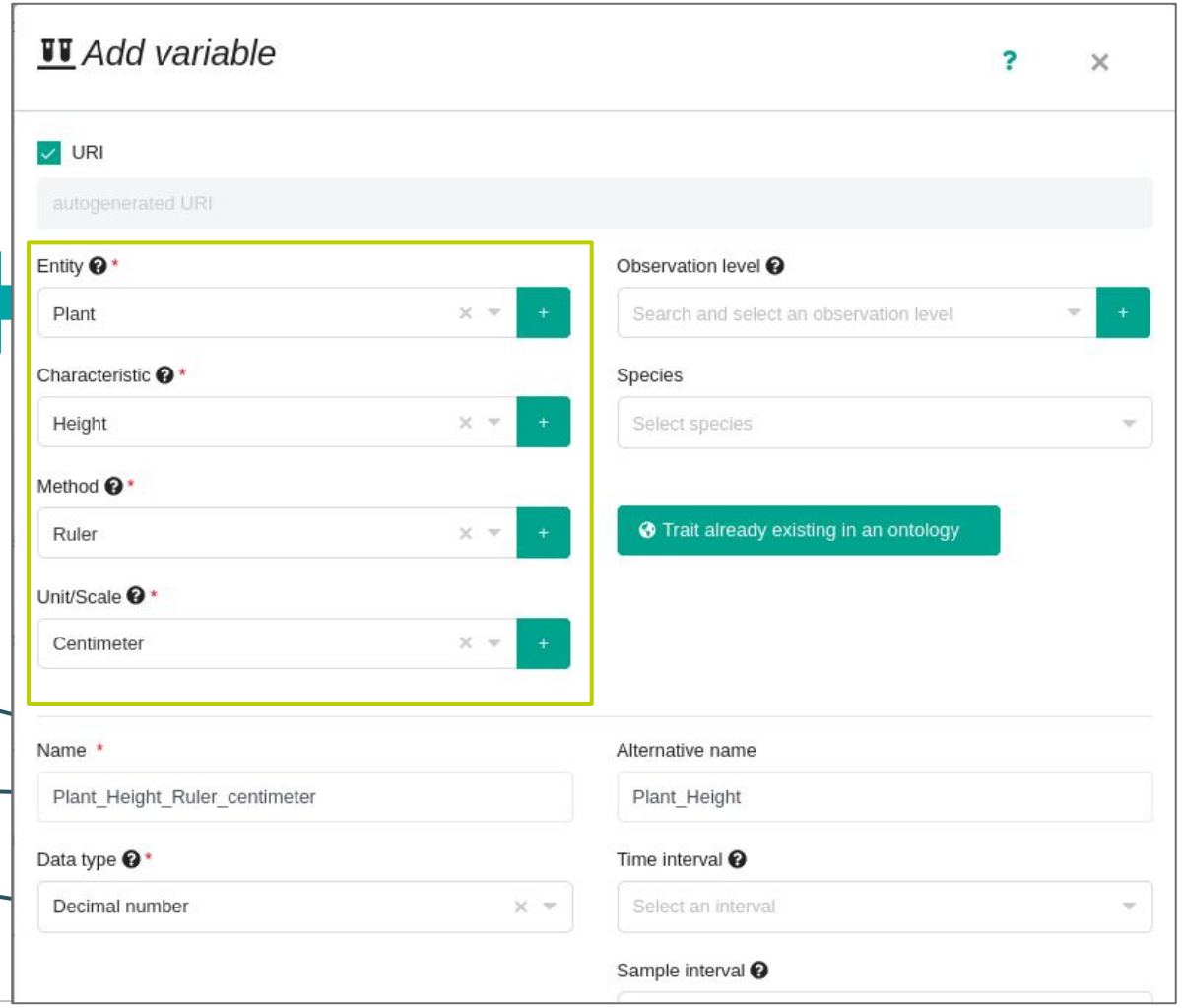
Enable semantically precise and FAIR descriptions - Machine readable



Variable Model

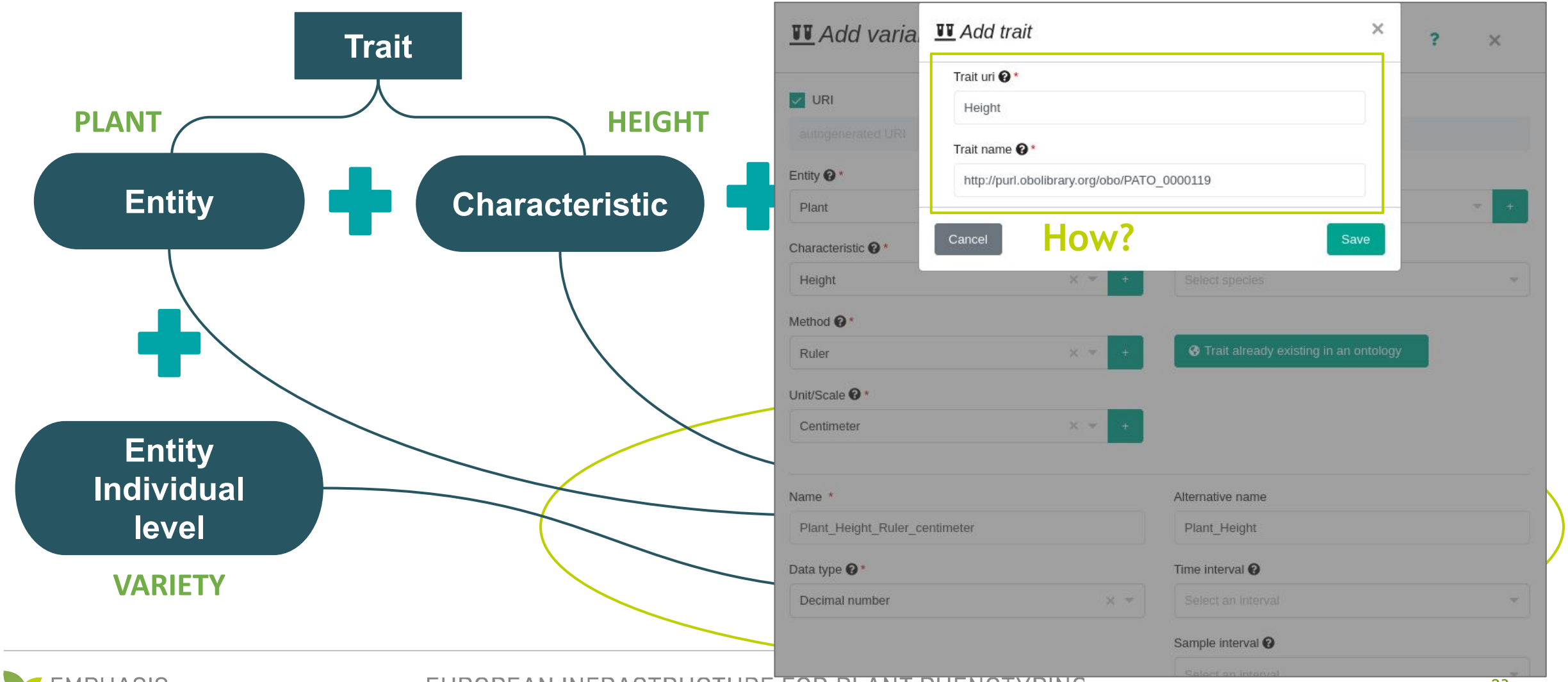
Break down the description into standardised atomic elements





Variable Model

Link these elements to existing vocabularies/ontologies

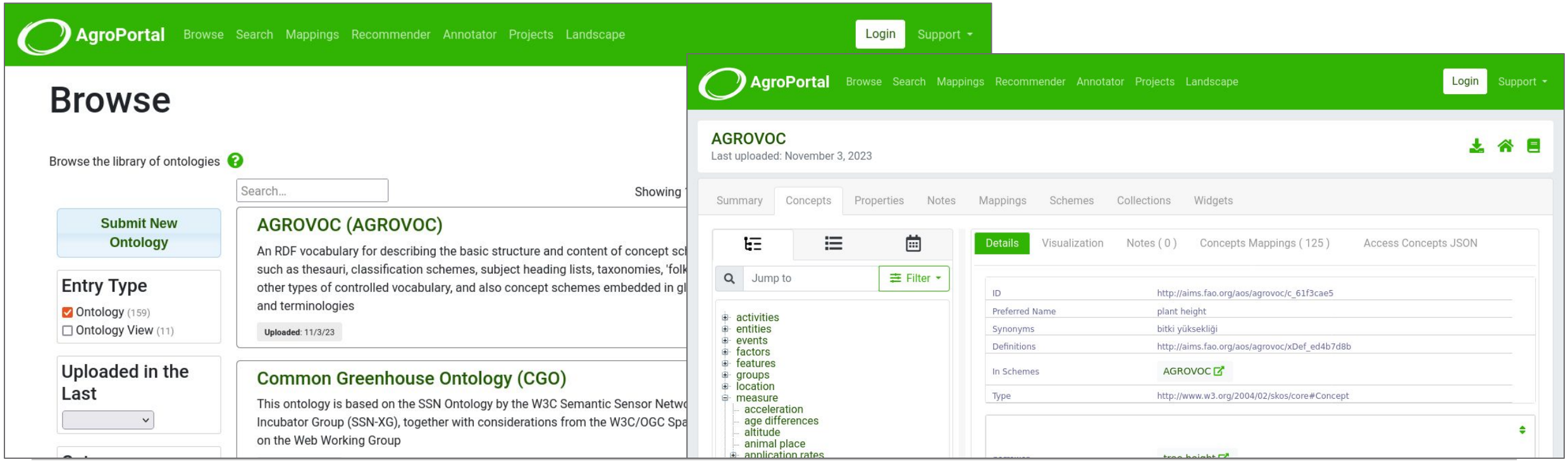


Variable sharing

Link these elements to existing vocabularies/ontologies

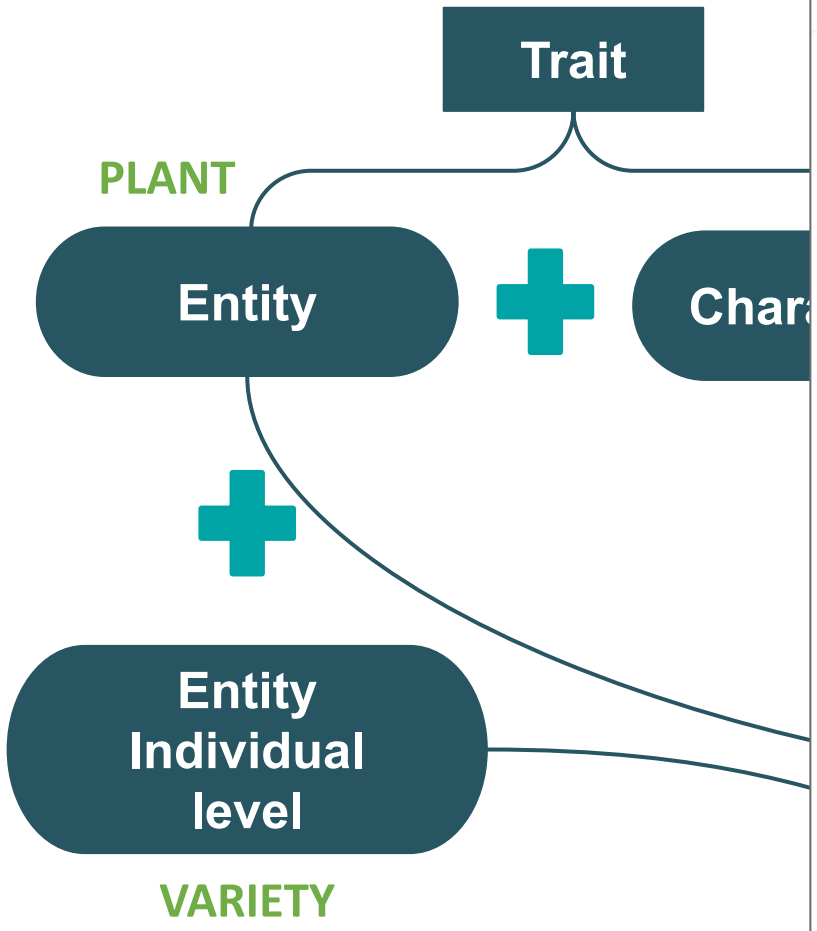
Agroportal, a vocabulary and ontology repository for agronomy and related domains -

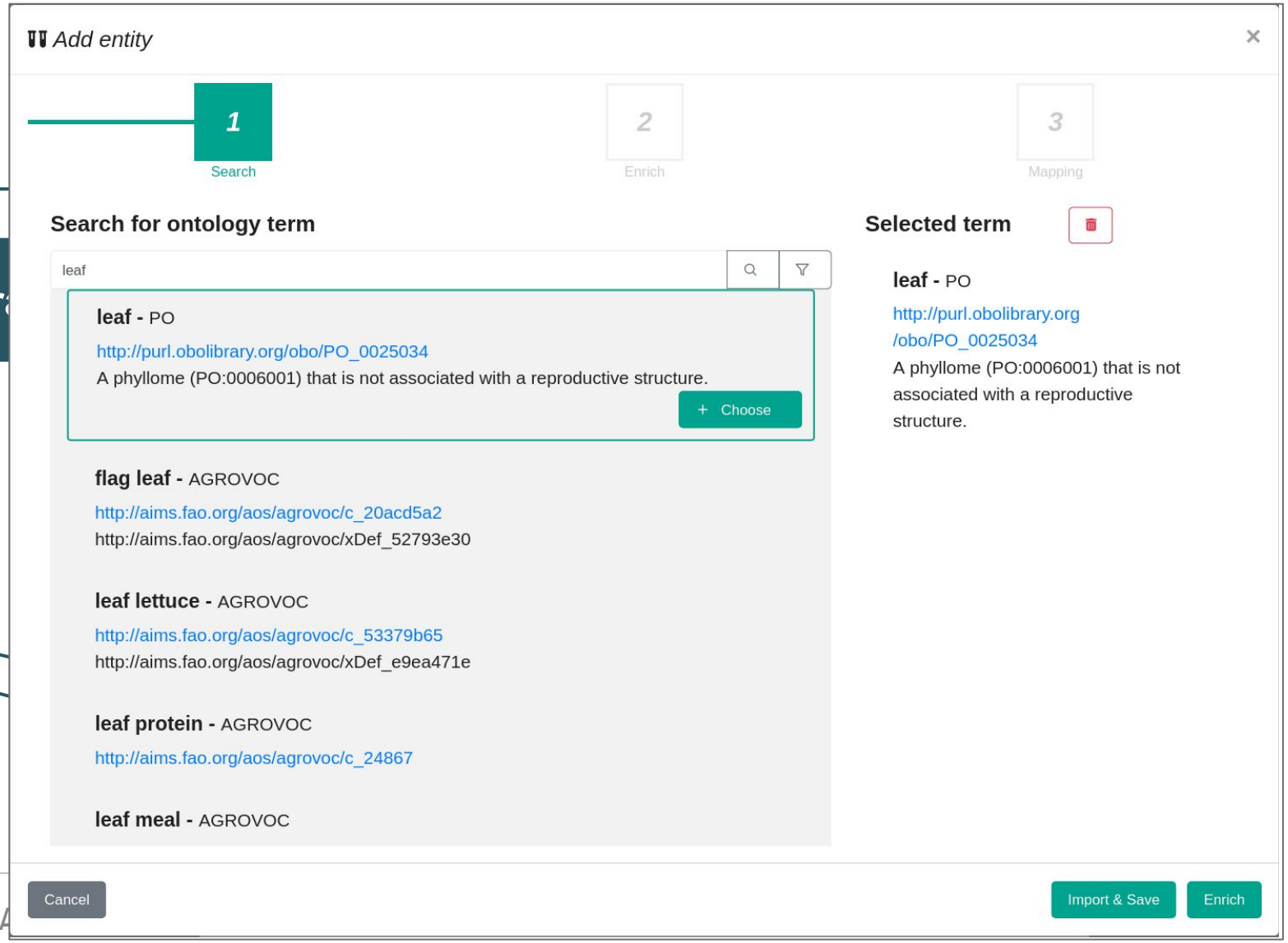
http://agroportal.lirmm.fr/



Describing Variables

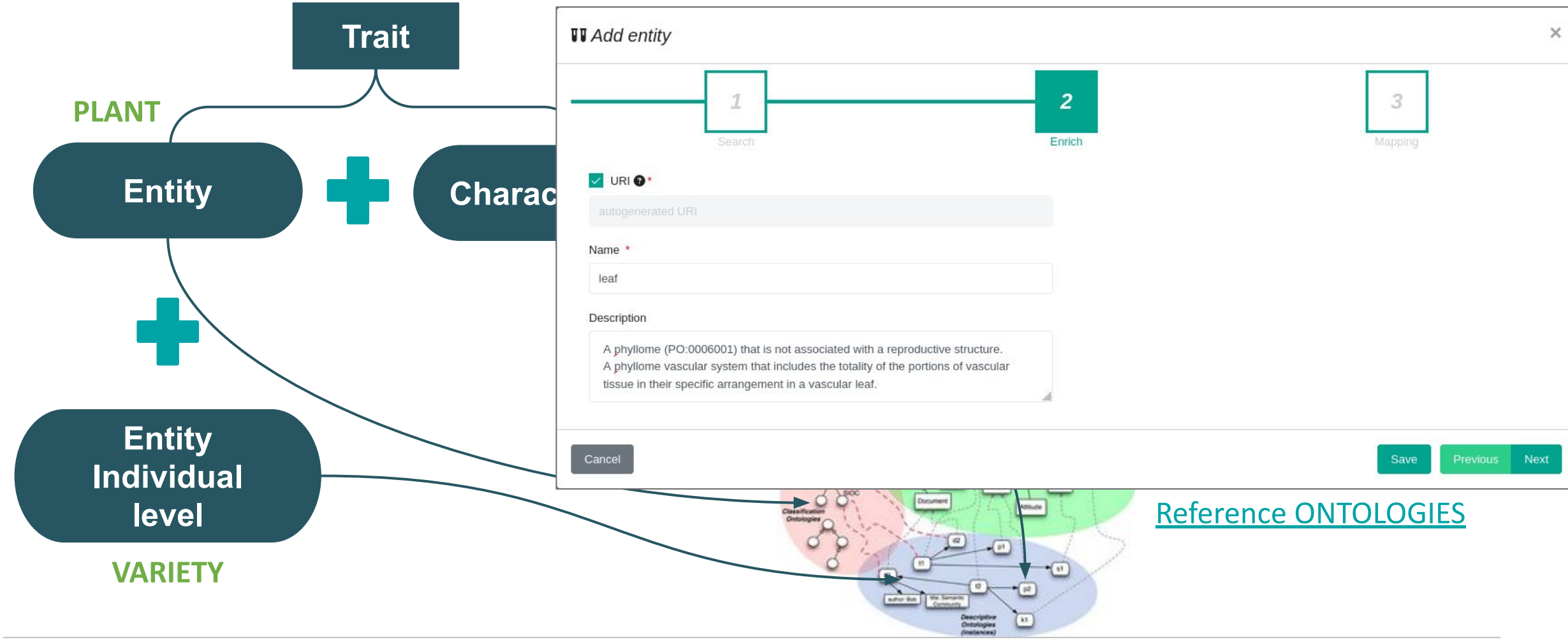
Find an ontology concept





Describing Variables

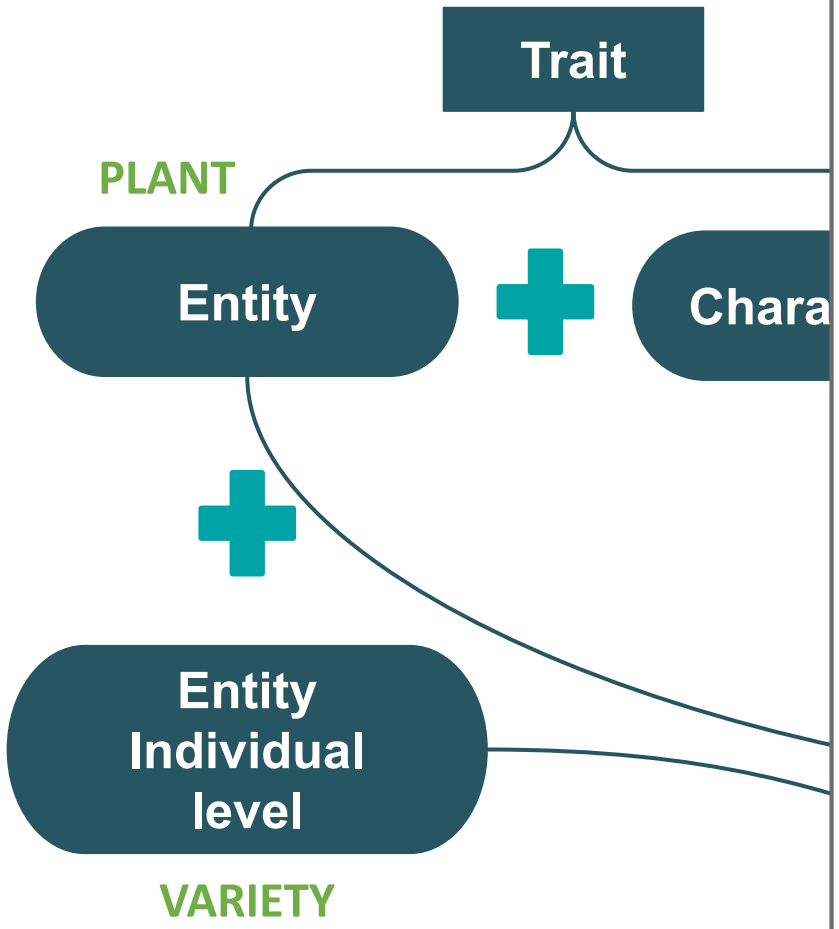
Enrich the concept

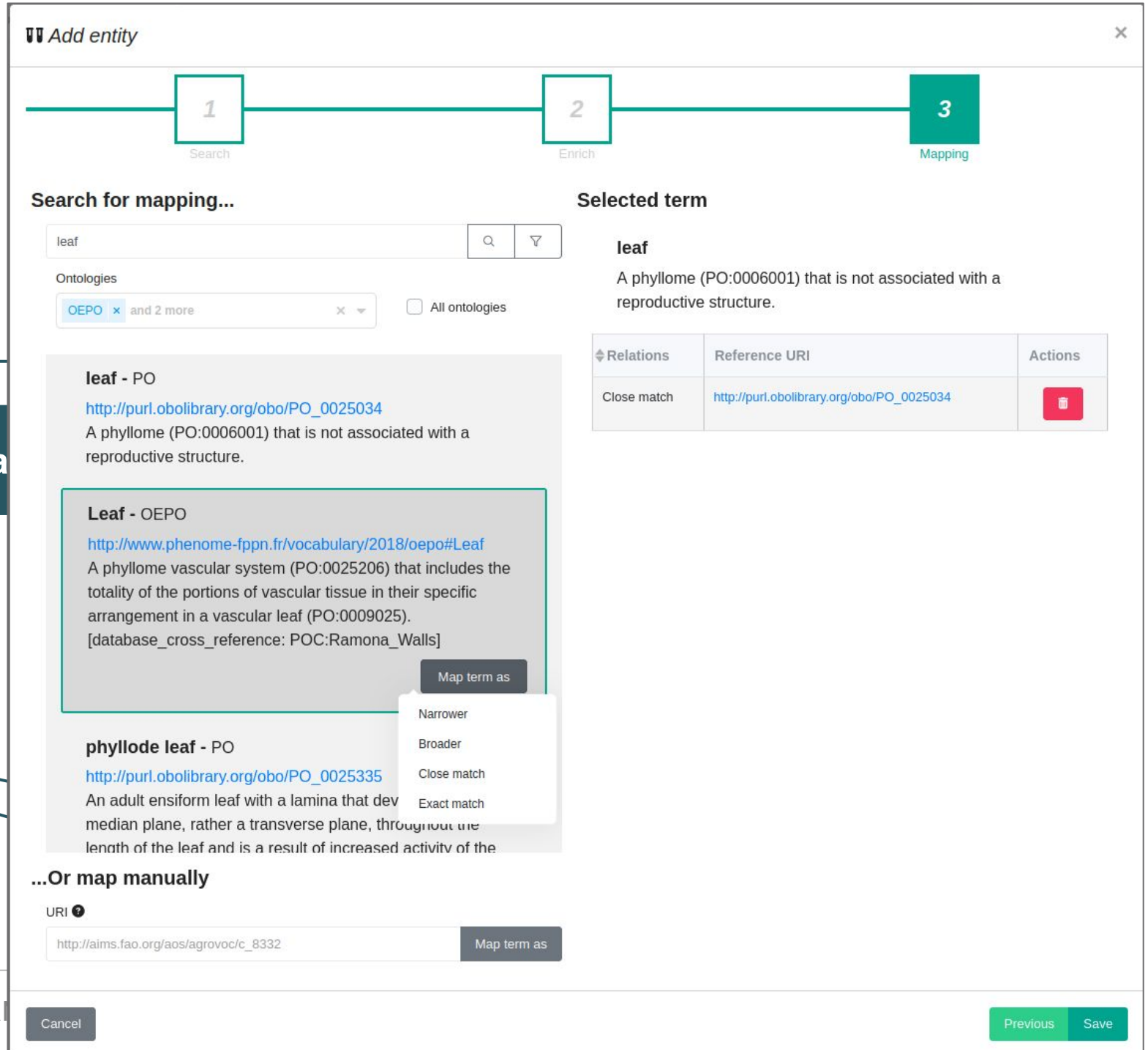


Reference ONTOLOGIES

Describing Variables

Map the concept





Variable sharing

Resource Sharing Tool: Unified documentation and identification

- Centralise descriptions and identifiers
- Avoid duplicate entries and multiple sources
- Pool the work involved in describing variables and genetic resources
- Consistent management of identifiers
- Facilitate the adoption of standards

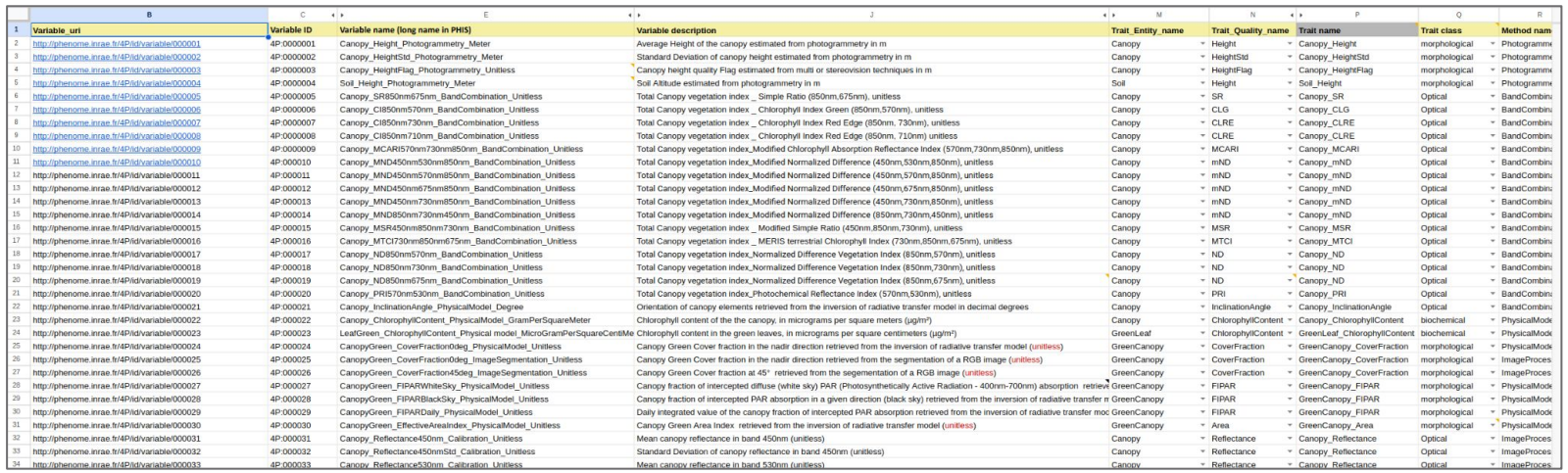
Variable sharing

Resource Sharing Tool: Unified

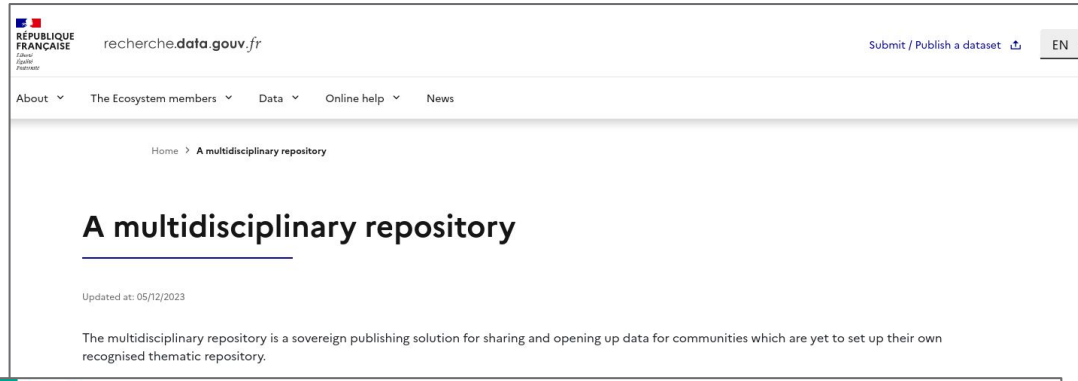
How?

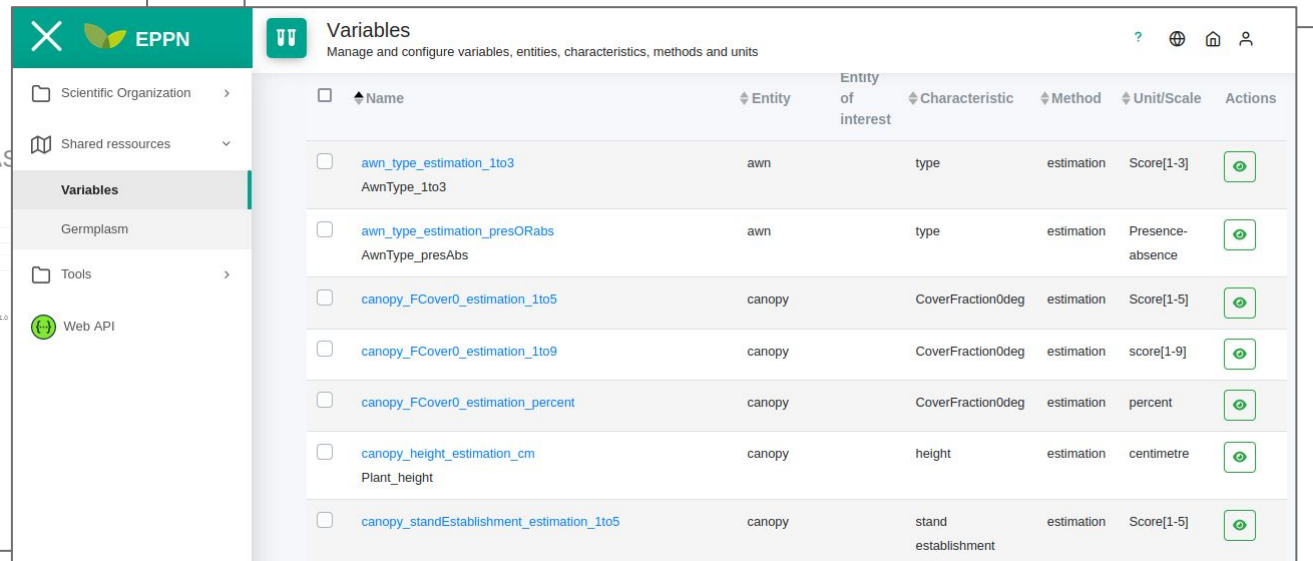
- Shared File (Google drive, Sharepoint, etc.)
- List of variables published on a dataverse
- PHIS instance

http://resources.plant-phenotyping.eu/emphasis/app/



Variable_uri	Variable ID	Variable name (long name in PHIS)	Variable description	Trait_Entity_name	Trait_Quality_name	Trait name	Trait class	Method nam
http://phenome.inrae.fr/4P/id/variable/000001	4P:0000001	Canopy_Height_Photogrammetry_Meter	Average Height of the canopy estimated from photogrammetry in m	Canopy	Height	Canopy_Height	morphological	Photogramm
http://phenome.inrae.fr/4P/id/variable/000002	4P:0000002	Canopy_HeightStd_Photogrammetry_Meter	Standard Deviation of canopy height estimated from photogrammetry in m	Canopy	HeightStd	Canopy_HeightStd	morphological	Photogramm
http://phenome.inrae.fr/4P/id/variable/000003	4P:0000003	Canopy_HeightFlag_Photogrammetry_Unitless	Canopy height quality Flag estimated from multi or stereovision techniques in m	Canopy	HeightFlag	Canopy_HeightFlag	morphological	Photogramm
http://phenome.inrae.fr/4P/id/variable/000004	4P:0000004	Soil_Height_Photogrammetry_Meter	Soil Altitude estimated from photogrammetry in m	Soil	Height	Soil_Height	morphological	Photogramm
http://phenome.inrae.fr/4P/id/variable/000005	4P:0000005	Canopy_SR850nm675nm_BandCombination_Unitless	Total Canopy vegetation index _ Simple Ratio (850nm,675nm), unitless	Canopy	SR	Canopy_SR	Optical	BandCombin
http://phenome.inrae.fr/4P/id/variable/000006	4P:0000006	Canopy_CI850nm570nm_BandCombination_Unitless	Total Canopy vegetation index _ Chlorophyll Index Green (850nm,570nm), unitless	Canopy	CLG	Canopy_CLG	Optical	BandCombin
http://phenome.inrae.fr/4P/id/variable/000007	4P:0000007	Canopy_CI850nm730nm_BandCombination_Unitless	Total Canopy vegetation index _ Chlorophyll Index Red Edge (850nm, 730nm), unitless	Canopy	CLRE	Canopy_CLRE	Optical	BandCombin
http://phenome.inrae.fr/4P/id/variable/000008	4P:0000008	Canopy_CI850nm710nm_BandCombination_Unitless	Total Canopy vegetation index _ Chlorophyll Index Red Edge (850nm, 710nm) unitless	Canopy	CLRE	Canopy_CLRE	Optical	BandCombin
http://phenome.inrae.fr/4P/id/variable/000009	4P:0000009	Canopy_MCARI570nm730nm850nm_BandCombination_Unitless	Total Canopy vegetation index_Modified Chlorophyll Absorption Reflectance Index (570nm,730nm,850nm), unitless	Canopy	MCARI	Canopy_MCARI	Optical	BandCombin
http://phenome.inrae.fr/4P/id/variable/000010	4P:0000010	Canopy_MND450nm530nm850nm_BandCombination_Unitless	Total Canopy vegetation index_Modified Normalized Difference (450nm,530nm,850nm), unitless	Canopy	mND	Canopy_mND	Optical	BandCombin
http://phenome.inrae.fr/4P/id/variable/000011	4P:000011	Canopy_MND450nm570nm850nm_BandCombination_Unitless	Total Canopy vegetation index_Modified Normalized Difference (450nm,570nm,850nm), unitless	Canopy	mND	Canopy_mND	Optical	BandCombin
http://phenome.inrae.fr/4P/id/variable/000012	4P:000012	Canopy_MND450nm675nm850nm_BandCombination_Unitless	Total Canopy vegetation index_Modified Normalized Difference (450nm,675nm,850nm), unitless	Canopy	mND	Canopy_mND	Optical	BandCombin
http://phenome.inrae.fr/4P/id/variable/000013	4P:000013	Canopy_MND450nm730nm850nm_BandCombination_Unitless	Total Canopy vegetation index_Modified Normalized Difference (450nm,730nm,850nm), unitless	Canopy	mND	Canopy_mND	Optical	BandCombin
http://phenome.inrae.fr/4P/id/variable/000014	4P:000014	Canopy_MND850nm730nm450nm_BandCombination_Unitless	Total Canopy vegetation index_Modified Normalized Difference (850nm,730nm,450nm), unitless	Canopy	mND	Canopy_mND	Optical	BandCombin
http://phenome.inrae.fr/4P/id/variable/000015	4P:000015	Canopy_MSR450nm850nm730nm_BandCombination_Unitless	Total Canopy vegetation index _ Modified Simple Ratio (450nm,850nm,730nm), unitless	Canopy	MSR	Canopy_MSR	Optical	BandCombin
http://phenome.inrae.fr/4P/id/variable/000016	4P:000016	Canopy_MTCI730nm850nm675nm_BandCombination_Unitless	Total Canopy vegetation index _ MERIS terrestrial Chlorophyll Index (730nm,850nm,675nm), unitless	Canopy	MTCI	Canopy_MTCI	Optical	BandCombin
http://phenome.inrae.fr/4P/id/variable/000017	4P:000017	Canopy_ND850nm570nm_BandCombination_Unitless	Total Canopy vegetation index_Normalized Difference Vegetation Index (850nm,570nm), unitless	Canopy	ND	Canopy_ND	Optical	BandCombin
http://phenome.inrae.fr/4P/id/variable/000018	4P:000018	Canopy_ND850nm730nm_BandCombination_Unitless	Total Canopy vegetation index_Normalized Difference Vegetation Index (850nm,730nm), unitless	Canopy	ND	Canopy_ND	Optical	BandCombin
http://phenome.inrae.fr/4P/id/variable/000019	4P:000019	Canopy_ND850nm675nm_BandCombination_Unitless	Total Canopy vegetation index_Normalized Difference Vegetation Index (850nm,675nm), unitless	Canopy	ND	Canopy_ND	Optical	BandCombin
http://phenome.inrae.fr/4P/id/variable/000020	4P:000020	Canopy_PRI570nm530nm_BandCombination_Unitless	Total Canopy vegetation index_Photochemical Reflectance Index (570nm,530nm), unitless	Canopy	PRI	Canopy_PRI	Optical	BandCombin
http://phenome.inrae.fr/4P/id/variable/000021	4P:000021	Canopy_InclinationAngle_PhysicalModel_Degree	Orientation of canopy elements retrieved from the inversion of radiative transfer model in decimal degrees	Canopy	InclinationAngle	Canopy_InclinationAngle	Optical	BandCombin
http://phenome.inrae.fr/4P/id/variable/000022	4P:000022	Canopy_ChlorophyllContent_PhysicalModel_GramPerSquareMeter	Chlorophyll content of the the canopy, in micrograms per square meters (µg/m²)	Canopy	ChlorophyllContent	Canopy_ChlorophyllContent	biochemical	PhysicalMod
http://phenome.inrae.fr/4P/id/variable/000023	4P:000023	LeafGreen_ChlorophyllContent_Physical model_MicroGramPerSquareCentiMe	Chlorophyll content in the green leaves, in micrograms per square centimeters (µg/m²)	GreenLeaf	ChlorophyllContent	GreenLeaf_ChlorophyllContent	biochemical	PhysicalMod
http://phenome.inrae.fr/4P/id/variable/000024	4P:000024	CanopyGreen_CoverFraction0deg_PhysicalModel_Unitless	Canopy Green Cover fraction in the nadir direction retrieved from the inversion of radiative transfer model (unitless)	GreenCanopy	CoverFraction	GreenCanopy_CoverFraction	morphological	PhysicalMod
http://phenome.inrae.fr/4P/id/variable/000025	4P:000025	CanopyGreen_CoverFraction0deg_ImageSegmentation_Unitless	Canopy Green Cover fraction in the nadir direction retrieved from the segmentation of a RGB image (unitless)	GreenCanopy	CoverFraction	GreenCanopy_CoverFraction	morphological	ImageProces
http://phenome.inrae.fr/4P/id/variable/000026	4P:000026	CanopyGreen_CoverFraction45deg_ImageSegmentation_Unitless	Canopy Green Cover fraction at 45° retrieved from the segementation of a RGB image (unitless)	GreenCanopy	CoverFraction	GreenCanopy_CoverFraction	morphological	ImageProces
http://phenome.inrae.fr/4P/id/variable/000027	4P:000027	CanopyGreen_FIPARWhiteSky_PhysicalModel_Unitless	Canopy fraction of intercepted diffuse (white sky) PAR (Photosynthetically Active Radiation - 400nm-700nm) absorption retrieve	GreenCanopy	FIPAR	GreenCanopy_FIPAR	morphological	PhysicalMod
http://phenome.inrae.fr/4P/id/variable/000028	4P:000028	CanopyGreen_FIPARBlackSky_PhysicalModel_Unitless	Canopy fraction of intercepted PAR absorption in a given direction (black sky) retrieved from the inversion of radiative transfer n	GreenCanopy	FIPAR	GreenCanopy_FIPAR	morphological	PhysicalMod
http://phenome.inrae.fr/4P/id/variable/000029	4P:000029	CanopyGreen_FIPARDaily_PhysicalModel_Unitless	Daily integrated value of the canopy fraction of intercepted PAR absorption retrieved from the inversion of radiative transfer moc	GreenCanopy	FIPAR	GreenCanopy_FIPAR	morphological	PhysicalMod
http://phenome.inrae.fr/4P/id/variable/000030	4P:000030	CanopyGreen_EffectiveAreaIndex_PhysicalModel_Unitless	Canopy Green Area Index retrieved from the inversion of radiative transfer model (unitless)	GreenCanopy	Area	GreenCanopy_Area	morphological	PhysicalMod
http://phenome.inrae.fr/4P/id/variable/000031	4P:000031	Canopy_Reflectance450nm_Calibration_Unitless	Mean canopy reflectance in band 450nm (unitless)	Canopy	Reflectance	Canopy_Reflectance	Optical	ImageProces
http://phenome.inrae.fr/4P/id/variable/000032	4P:000032	Canopy_Reflectance450nmStd_Calibration_Unitless	Standard Deviation of canopy reflectance in band 450nm (unitless)	Canopy	Reflectance	Canopy_Reflectance	Optical	ImageProces
http://phenome.inrae.fr/4P/id/variable/000033	4P:000033	Canopy_Reflectance530nm_Calibration_Unitless	Mean canopy reflectance in band 530nm (unitless)	Canopy	Reflectance	Canopy_Reflectance	Optical	ImageProces





Variables & Provenance: Understanding Data

Building credibility and trust in data, analysis and reporting

- **Method** is associated to **Variable**
- **Provenance** is associated to **Data**

- **Variable method:** protocol and type of thermometer (not compatible values)
- **Provenance:** name of the nurse or the manufacturer of the thermometer

Variable definition - Special focus on method

Building credibility and trust in data, analysis and reporting

Trait = Plant Height

- Method 1 : Manual measurement
- Method 2 : Image processing

=> 2 different and non-comparable variables!

Image: Moira Sheehan

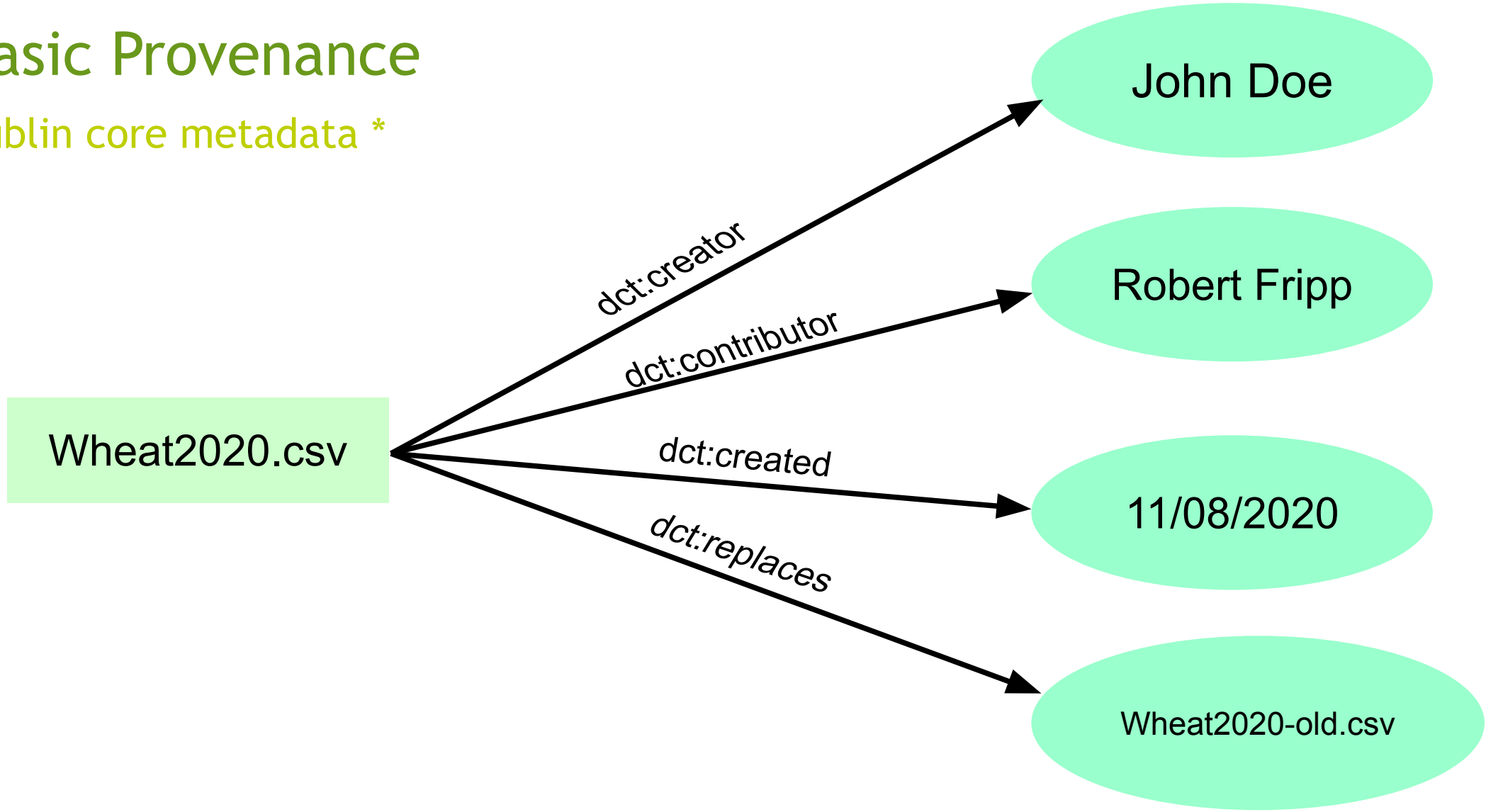
Whats is Provenance ?

That is metadata to know…

- Who played a role when creating the data?
- Who contributed to the data?
- How data was transformed?
- Which tools were used?
- When and how data were produced?
- etc.

Basic Provenance

Dublin core metadata *



* https://www.dublincore.org/specifications/dublin-core/dcmi-terms

Advanced Provenance

Representation model for provenance (PROV-O *)

- **Entity:** data, data sources, documents, results, etc.

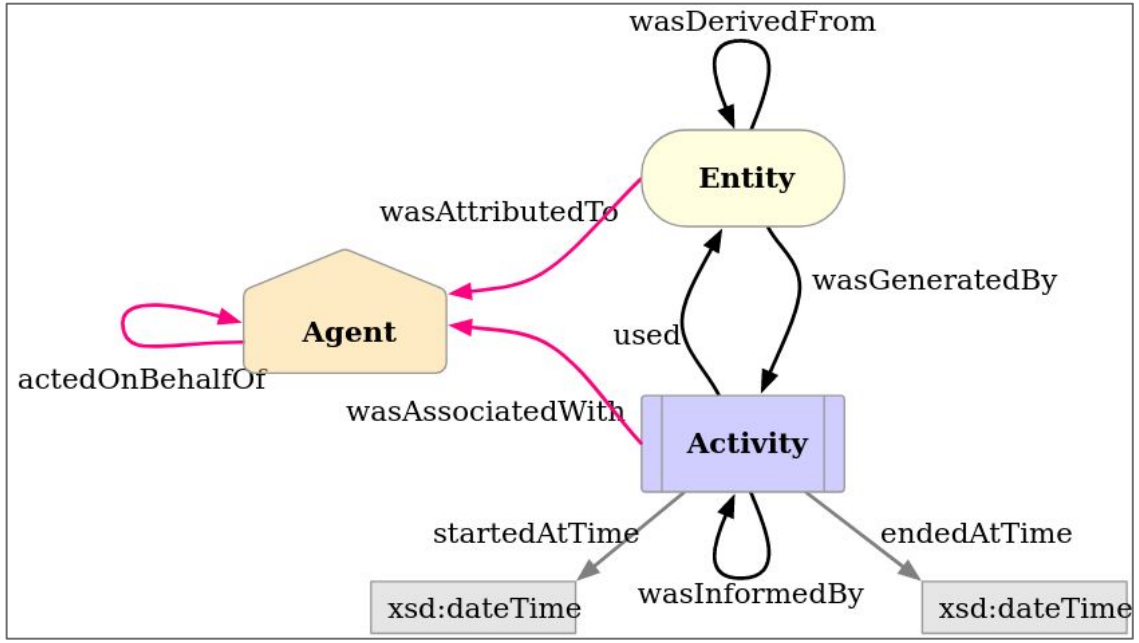
  Entity can be **input** or **output**

- **Agent:** person, software, Web services, institution, company, etc.

  Agent bears the responsibility

- **Activity:** generating, transforming, modifying, processing, etc.

  Activity occurs over a period of time and acts upon or with entities



* *https://www.w3.org/TR/prov-o/*

Provenance: Data Acquisition description

- **Provenance (Activity)**
  - Formalisation of the transformation of raw data into processed data
  - Provenance will describe a process, a step in a data acquisition workflow
- **Datafiles and Data (Entity)**
  - Single item of factual information of variable complexity (e.g. a length or an image) resulting from measurement, observation or processing.

PROVENANCE

- Description
- Start / End Date
- …

DATA

- Variable
- Targets / Scientific Objects
- **Provenance**
- Date
- Value

Provenance: Data Acquisition description

- **Provenance (Activity)**
  - Formalisation of the transformation of raw data into processed data
  - Provenance will describe a process, a step in a data acquisition workflow
- **Datafiles and Data (Entity)**
  - Single item of factual information of variable complexity (e.g. a length or an image) resulting from measurement, observation or processing.
- **Devices (Agent)**
  - Sensors
  - Vectors
  - Softwares

PROVENANCE

- Description
- Start / End Date
- Vector
- Sensor
- Operator
- ...

DATA

- Variable
- Targets / Scientific Objects
- **Provenance**
- Date
- Value

Devices

Description

Device is any equipment to measure or control or calculate a property of an object, including also any equipment holding or carrying another one.

Ex.:

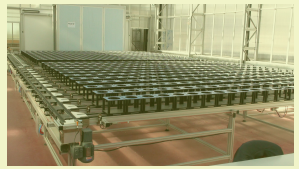
- *Camera, Thermometer, Software, Spectrometer*
- *UAV, Field Robots Conveyor*
- *Weather Station, Acquisition station*

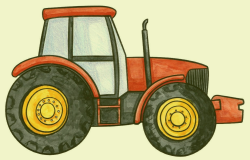
Devices Examples

Phenotyping & envirotyping

Climate
Soil
Canopy
Organism
Leaf
Management

Vector






Sensor








Software/pipeline

python
R
Shiny
MATLAB
ImageJ
Image Processing and Analysis in Java
ENVI
IDL
4P PLANT PHENOTYPING PROCESSING PLATFORM

Datafiles and data

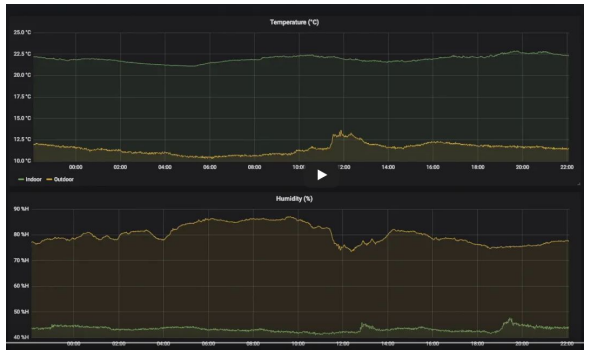
Description

- **Data:** Data is a single item of factual information of variable complexity (e.g. a length or an image) resulting from measurement, observation or processing. This element can be used as input or output by specialised IT tools.
- **Tabular Data:** : Scalar data.
- **Datafile / Complex Data:** Complex data often serialized in files (e.g. image file, spectrum output, orthomosaic, etc.)

Environmental Data Acquisition

Automatized data acquisition from sensors

Data acquisition



Data loggers & monitoring

Provenance
Devices: atp-20
Station: campbell-2018-dl53

Data
Variable: air_temperature_minimumDaily_degC
Date: 2023-07-01
Target: plot2023-BH53
Value: 12

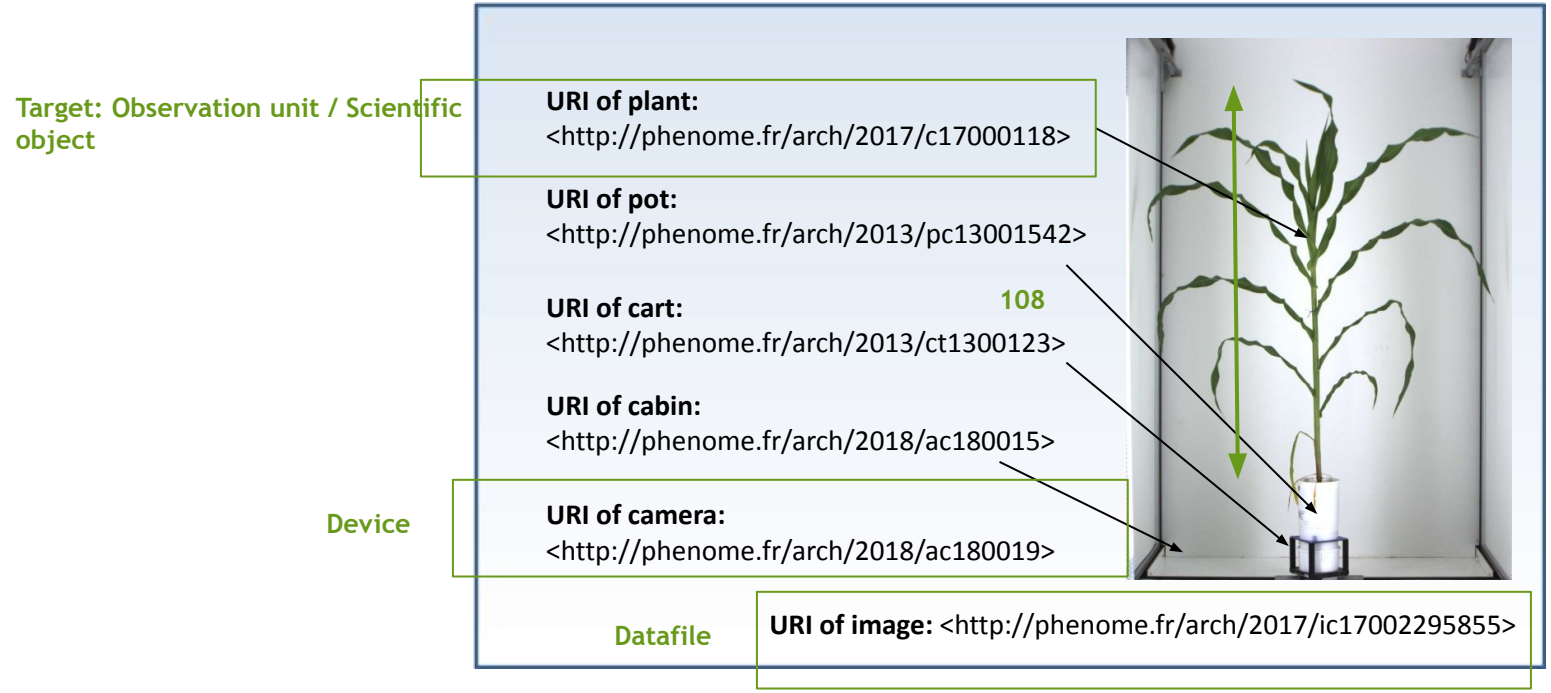
Web API

- Metadata : Experiments, plots, factors, sensors, variables, etc.
- Data

=> Find, Analyze, Visualize

Phenotyping Data Acquisition

Measuring Plant Height



Phenotyping Data Acquisition

Measuring Plant Height

Step 1: Image Acquisition

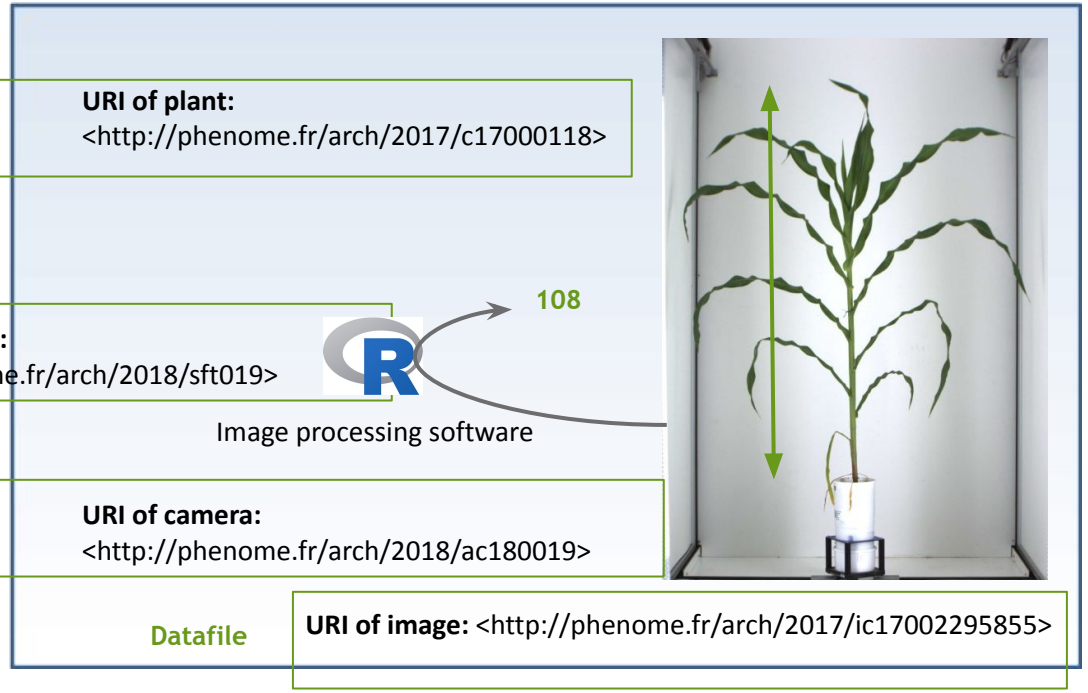
Provenance Datafile

- Image Acquisition
- Acquisition from 2017-06-01 to 2017-07-15
- Device: camera ***ac180019***
  - Type: RGB Camera
  - Settings: Side View

Step 2: Data Acquisition

Provenance Data

- Image Processing
- Acquisition from 2017-06-01 to 2017-07-15
- Device: camera *ac180019*
  - Settings: Side View
- Software: Image processing software
  - Settings: version 1.1.5
- Processing Datafile ***ic17002295855***



Trait = Plant Height

Variable =

- **Plant**
- **Height**
- **Image Processing**
- **Centimetre**

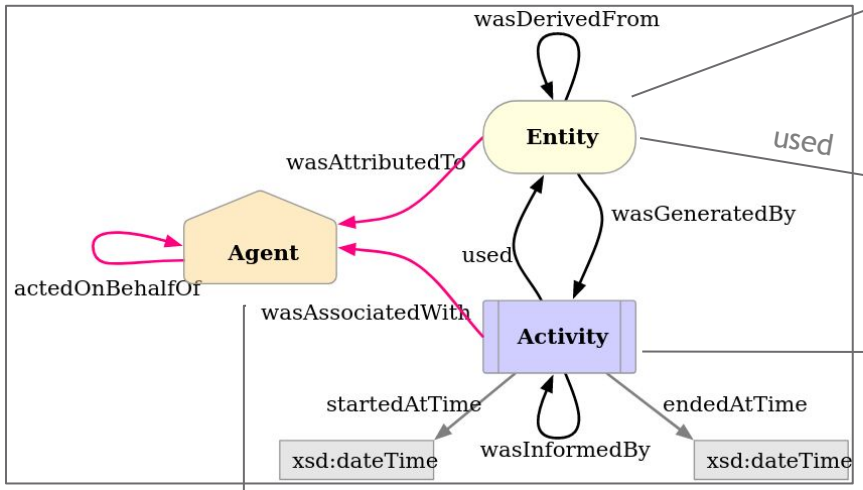
URI of Variable:
<https://emphasis/id/variable/plant_height_imageProcessing_cm>

Data

- Variable: plant_height_imageProcessing_cm
- Target: Plant c17000118
- Date: 2017-06-15
- Value: **108**

Phenotyping Data Acquisition

Measuring Plant Height



wasGeneratedBy

Data

```
{
“variable”: “plant_height_imageProcessing_cm
“target”: “Plant c17000118”,
“Date”: “2017-06-15”,
“Value”: 108
}
```

output

Provenance

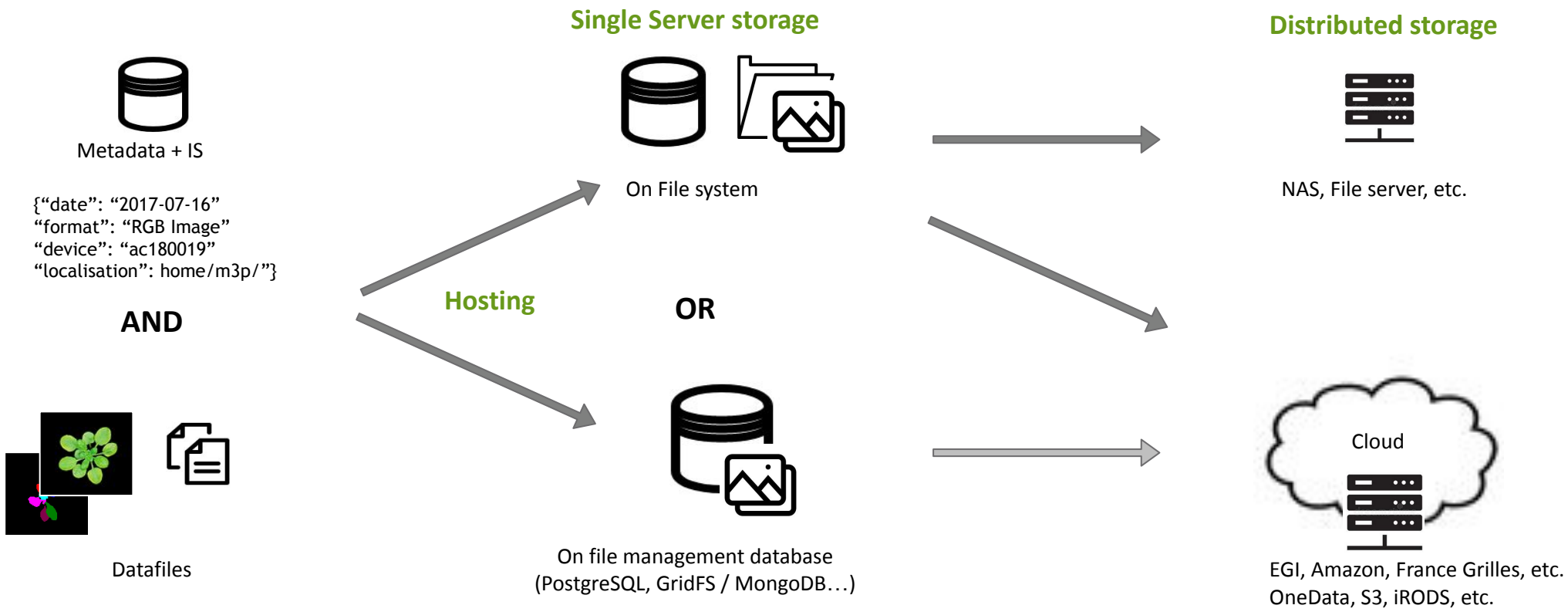
```
{
"name": "Modular_RGB1_MorphoParameters",
"description": "Morphological parameters computed by DataAnalyser",
“prov_entity”: [{
        "uri": "ic17002295855",
        "type": "RGBImage"
      }]
"prov_activity": [{
      "type": "ImageProcessing",
      "start_date": 2017-06-01,
      "end_date": 2017-07-15
    }],
"prov_agent": [
          {"uri": "sft019",
          "type": "Software",
          "settings":
          {"Version": "version 1.1.5",
           "Formula": "4*G-3*B-R",
           “Process”: https://github/OpenSILEX/cli.r” }
          }
          ],
}
```

input

wasAssociatedWith

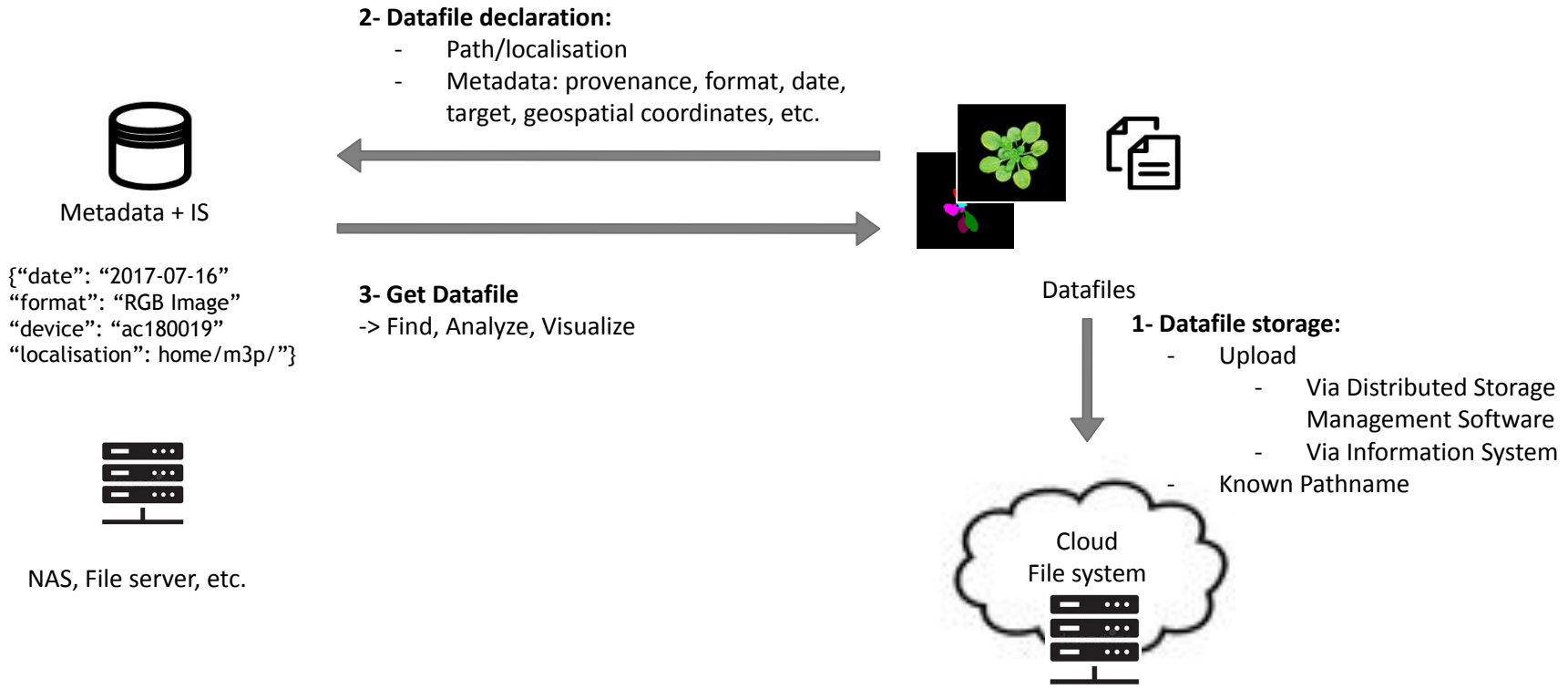
Datafiles storage & link to Information System

Different storage solutions



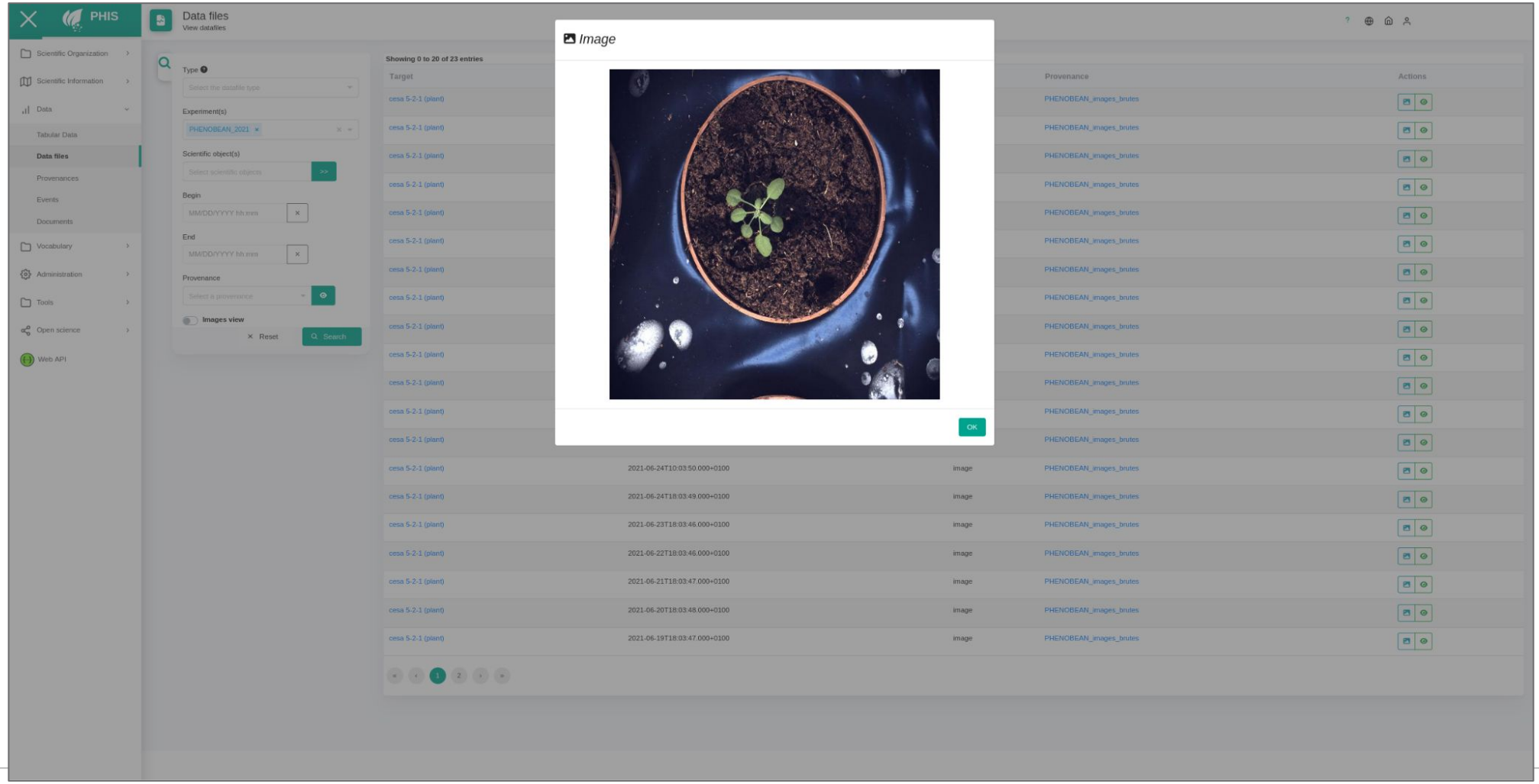
Datafiles storage & link to Information System

Implementation Example



Datafiles storage & link to Information System

Implementation Example: view images directly from the web interface



Implementation Example: PHIS, an Information System for Plant Phenomics

Environmental data

Phenotypic data

Image processing
Advanced data extraction

Image processing Analysis pipeline

RÉPUBLIQUE FRANÇAISE recherche.data.gouv.fr

Community

User

User interface
Vue.js

WS User interface
Swagger UI

swagger

Scripts

REST API
Web App

PHIS

File storage system

iRODS

Analysis, sharing, etc.

- Metadata semantic description (experiments, plots, factors, sensors, variables, etc.)
- Data

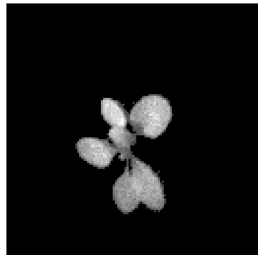
Data & Datafiles insertion

Image processing - Analysis Pipeline

Plant

camera ?
settings ?
angle ?

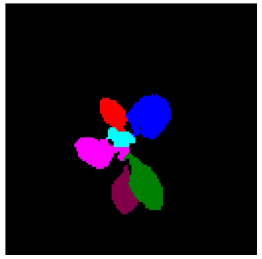
Provenance
Image Acquisition



Image

treatment ?
model ?
settings ?

Provenance
Segmentation



Segmented Image

treatment ?
model ?
settings ?

Provenance Data
Pipeline Segmentation
(traits extraction)

Plant
Area =
6249 px

Elaborated Data
(e.g. Area)

Use Case - Sylvain Poque

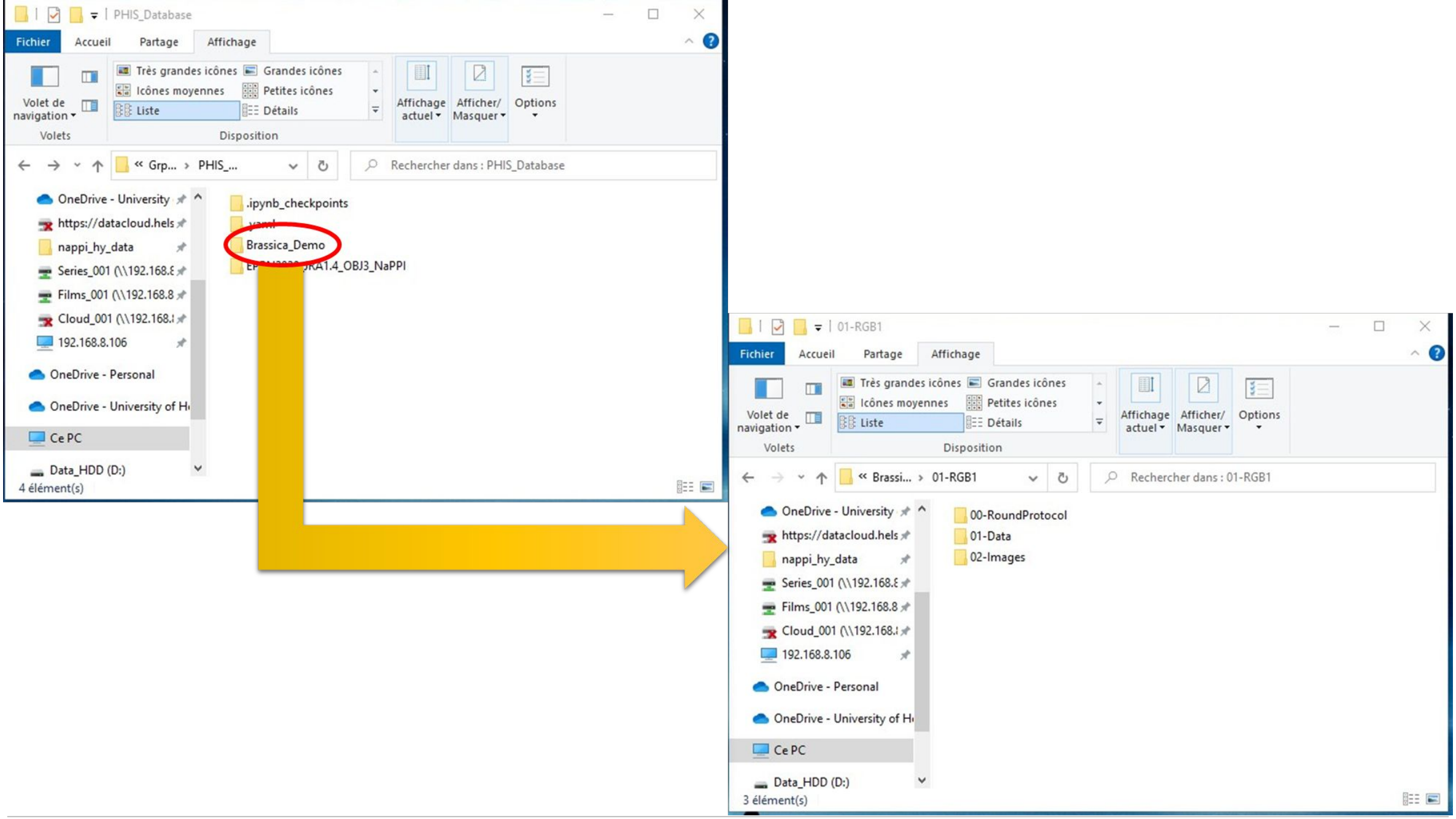
Sending Data to PHIS Information System

UNIVERSITY OF HELSINKI

UNIVERSITY OF HELSINKI

PHIS

```
<Measure>
  <AdaptTime>00:00:00</AdaptTime>
  <Prescription id="1" name="Recipe New:1">
    <RGB1 height="Default">
      <Offset>0</Offset>
      <Frame angle="0" />
      <Frame angle="120" />
      <Frame angle="240" />
    </RGB1>
    <RGB2 height="Automatic">
      <Offset>0</Offset>
      <Frame angle="0" />
    </RGB2>
    <Analyse>
      <MaskErosionLevel>1</MaskErosionLevel>
      <RGB1>
        <PlantMask>
          <Formula>0-((2*R)+(3*B)-(5*G))</Formula>
          <Threshold>0.1098039</Threshold>
          <MedianFilterSize>4</MedianFilterSize>
          <MinSize>100</MinSize>
          <MinHoleSize>10</MinHoleSize>
          <CropObjectsOnBorders>False</CropObjectsOnBorders>
          <UseReflectionReduction>False</UseReflectionReduction>
          <SkipBadExposedPoints>True</SkipBadExposedPoints>
```


PHIS_Database
Fichier
Accueil
Partage
Affichage
Volet de navigation
Très grandes icônes
Grandes icônes
Icônes moyennes
Petites icônes
Liste
Détails
Affichage actuel
Afficher/ Masquer
Options
Volets
Disposition
« Grp... › PHIS_...
Rechercher dans : PHIS_Database
OneDrive - University
https://datacloud.hels
nappi_hy_data
Series_001 (\\192.168.8
Films_001 (\\192.168.8
Cloud_001 (\\192.168.
192.168.8.106
OneDrive - Personal
OneDrive - University of H
Ce PC
Data_HDD (D:)
4 élément(s)
.ipynb_checkpoints
Brassica_Demo
01-RGB1
« Brassi... › 01-RGB1
Rechercher dans : 01-RGB1
00-RoundProtocol
01-Data
02-Images
3 élément(s)

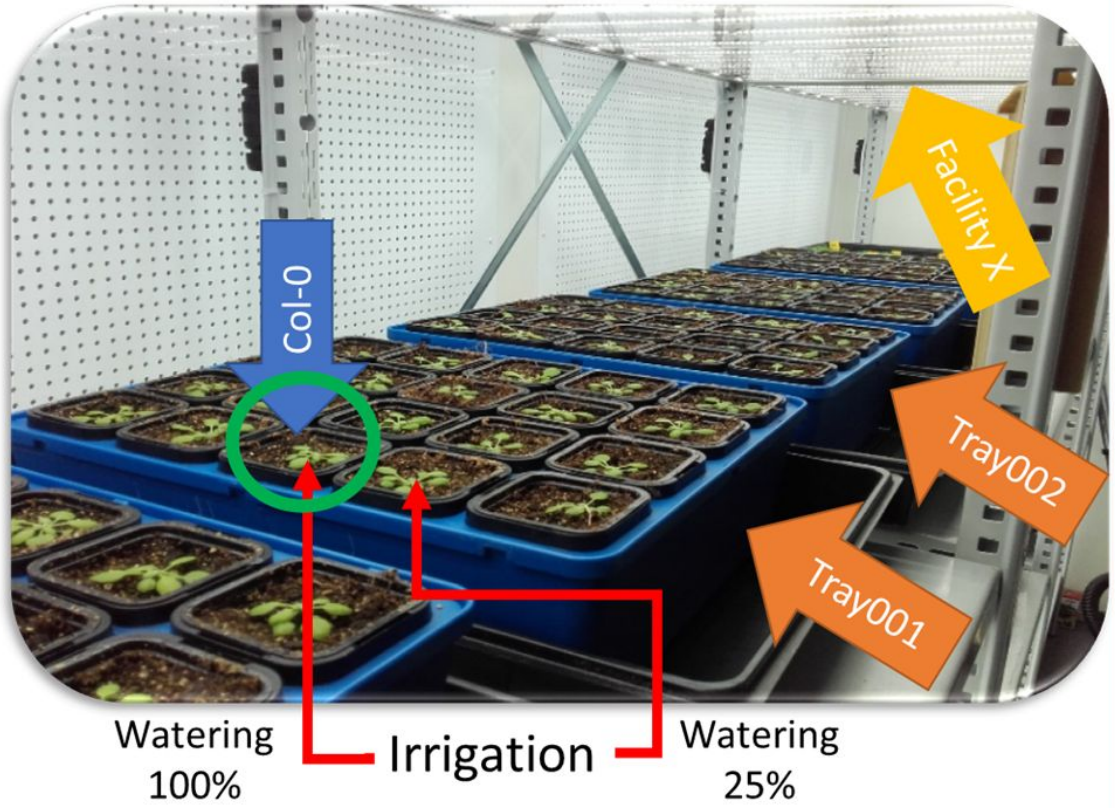
PHIS Identification & Semantics

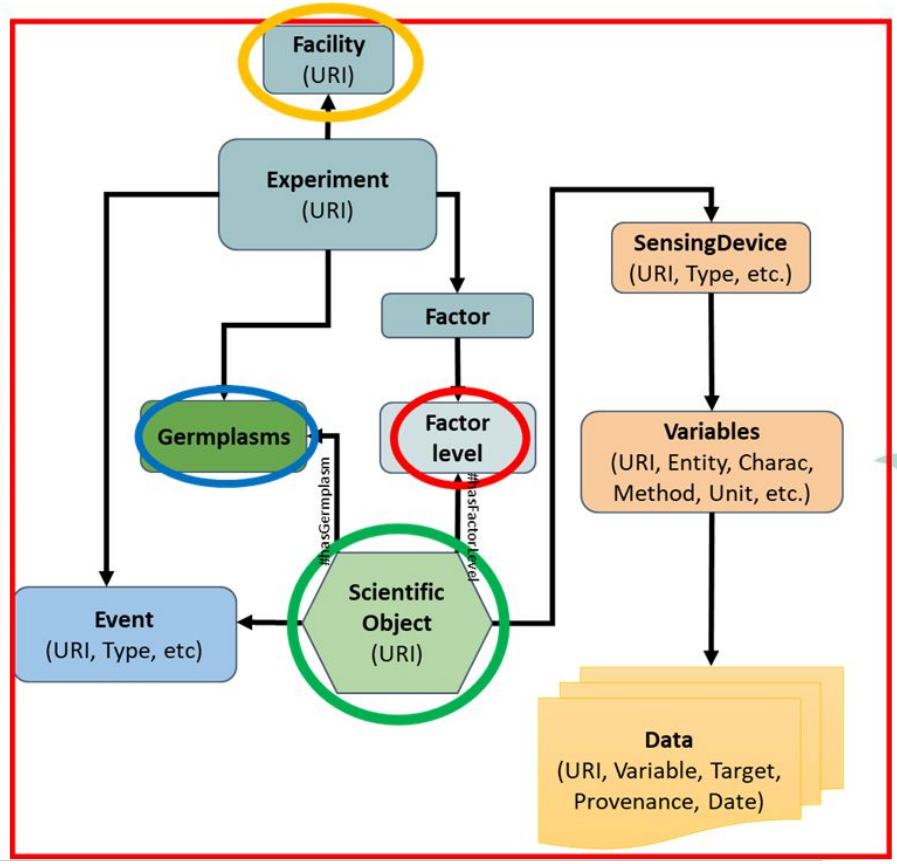
Identification

- Everything can be identified: plants, experiments, sensors, events, etc.
- Persistent, unambiguous, resolvable

Semantics

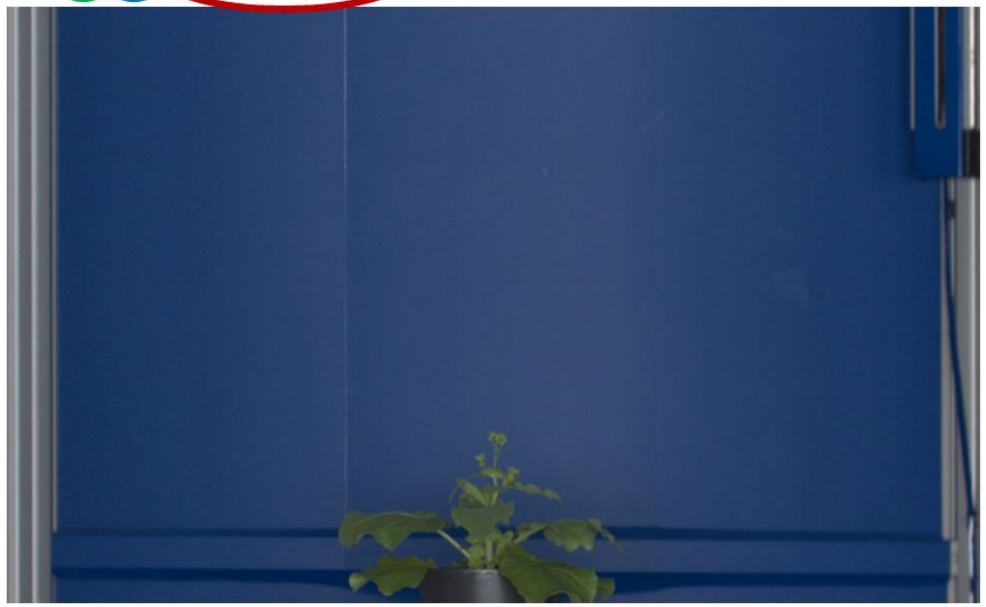
- Naming Conventions
- Controlled vocabulary
- Formalized relationships between entities
- Data annotation and enrichment





Example of NaPPI: Data structure outputs

69 25 180806.Bra.37-RGB1-000-FishEyeCorrected.png



69 42 180806.Bra.37-RGB1-000-FishEyeCorrected.png

Photon Systems Instruments

Measuring Date	Experiment ID	Round Order	Tray ID	Plant ID	Geno	Germplasm	PID	Position	Angle	AREA_MM	PERIMETER_MM
2018-08-31	69	14	180806.Bra.37	BRA04_037	SemiDwarf	BC4F2_4003_SemiDwarf	RGB1	A1	0	4990.566734	745.6099917
2018-09-02	69	19	180806.Bra.37	BRA04_037	SemiDwarf	BC4F2_4003_SemiDwarf	RGB1	A1	0	8288.080108	1142.58486
2018-09-06	**69**	**25**	**180806.Bra.37**	**BRA04_037**	**SemiDwarf**	**BC4F2_4003_SemiDwarf**	**RGB1**	**A1**	0	**26405.13174**	**2902.298909**
2018-09-08	69		180806.Bra.37	BRA04_037	SemiDwarf	BC4F2_4003_SemiDwarf	RGB1	A1	0	40337.01069	4089.33801
2018-09-10	69	34	180806.Bra.37	BRA04_037	SemiDwarf	BC4F2_4003_SemiDwarf	RGB1	A1	0	53688.43063	6107.038481
2018-09-12	69	37	180806.Bra.37	BRA04_037	SemiDwarf	BC4F2_4003_SemiDwarf	RGB1	A1	0	64638.33809	7413.620659
2018-09-13	69	39	180806.Bra.37	BRA04_037	SemiDwarf	BC4F2_4003_SemiDwarf	RGB1	A1	0	70297.66953	6896.141705
2018-09-14	**69**	**42**	**180806.Bra.37**	**BRA04_037**	**SemiDwarf**	**BC4F2_4003_SemiDwarf**	**RGB1**	**A1**	0	**76262.47543**	**7984.092591**
2018-09-17	69		180806.Bra.37	BRA04_037	SemiDwarf	BC4F2_4003_SemiDwarf	RGB1	A1	0	97027.26002	9549.016183
2018-09-18	69	51	180806.Bra.37	BRA04_037	SemiDwarf	BC4F2_4003_SemiDwarf	RGB1	A1	0	102117.0628	10932.23031

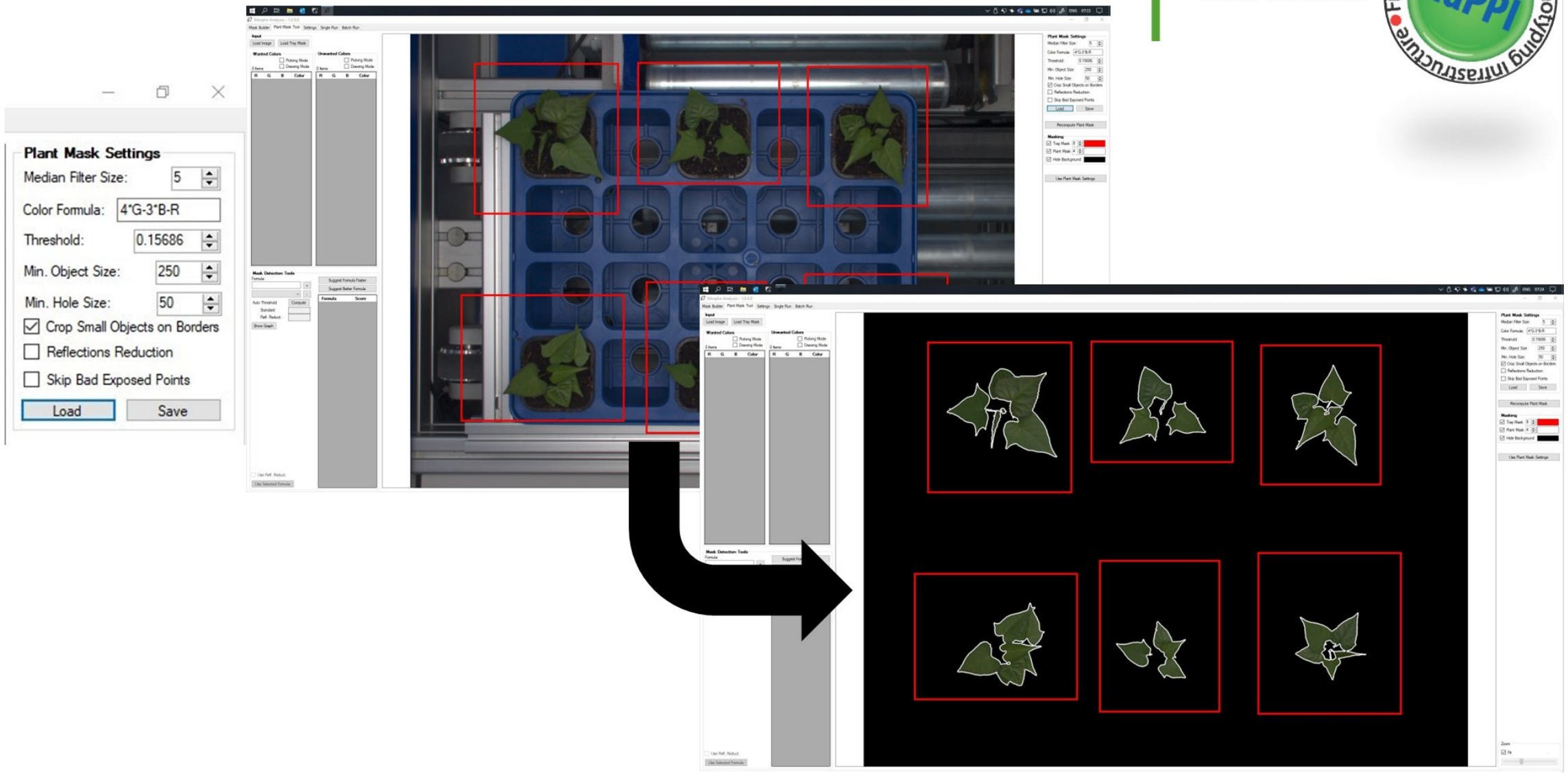
Scientific Object

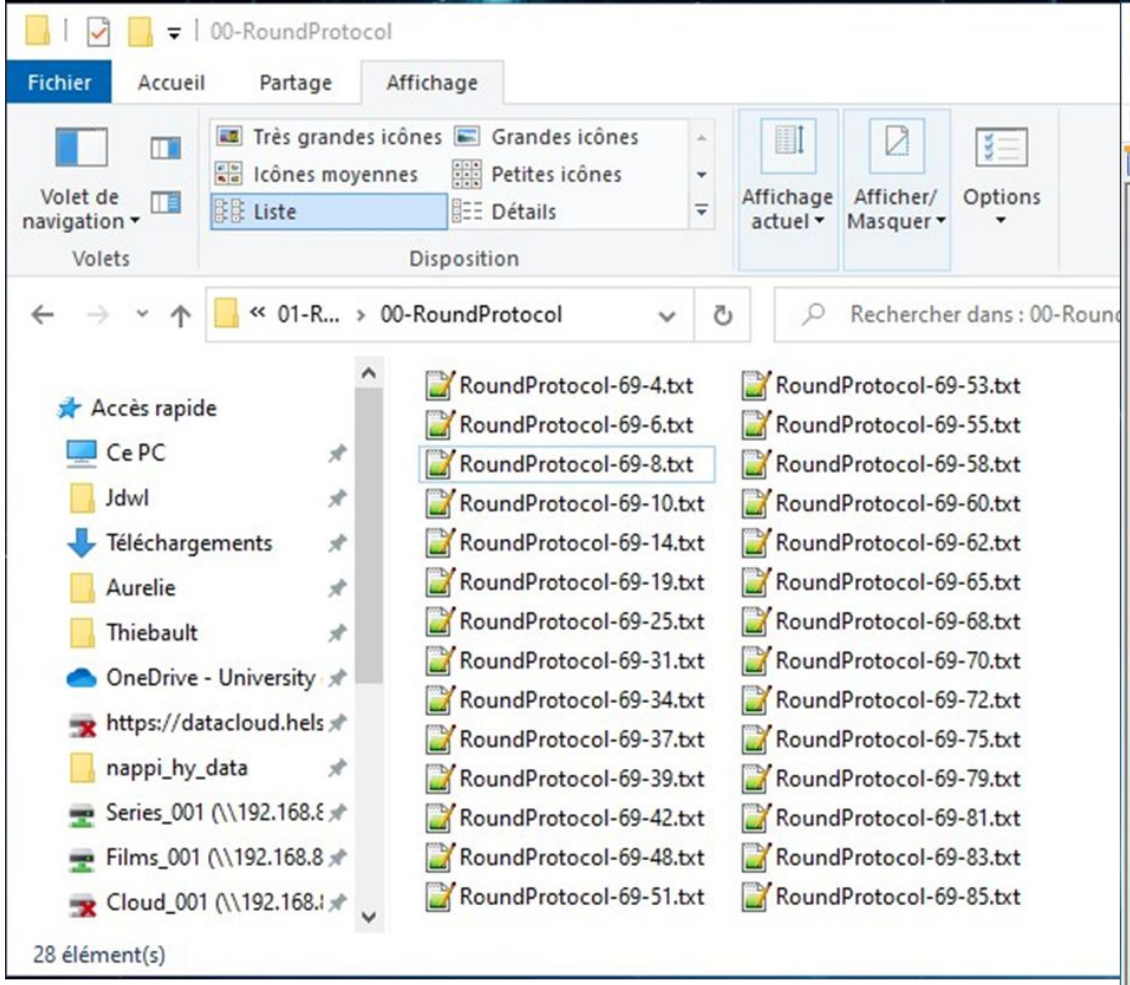
Germplasm

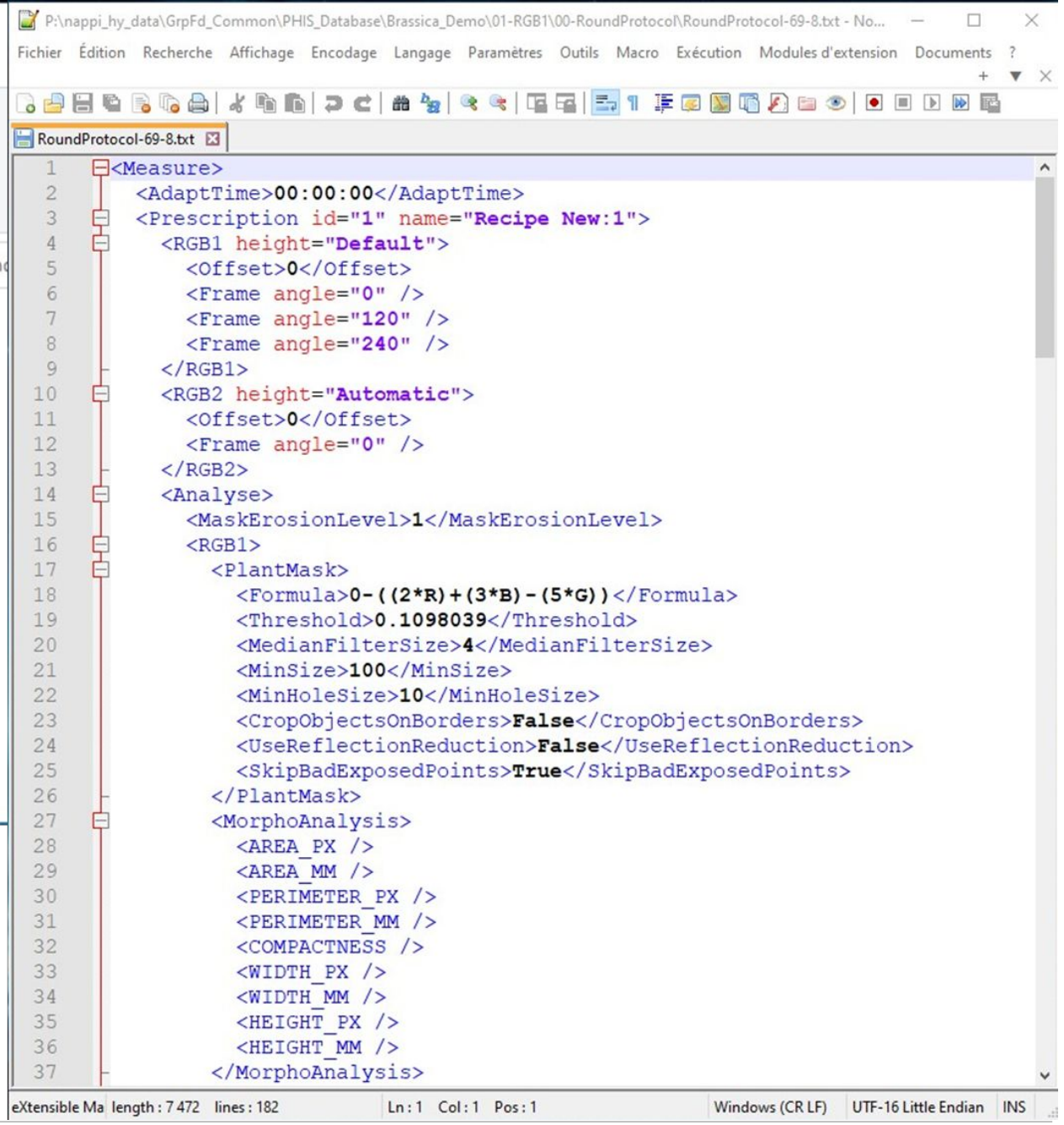
Extracting info from images with PSI software

Photon Systems Instruments

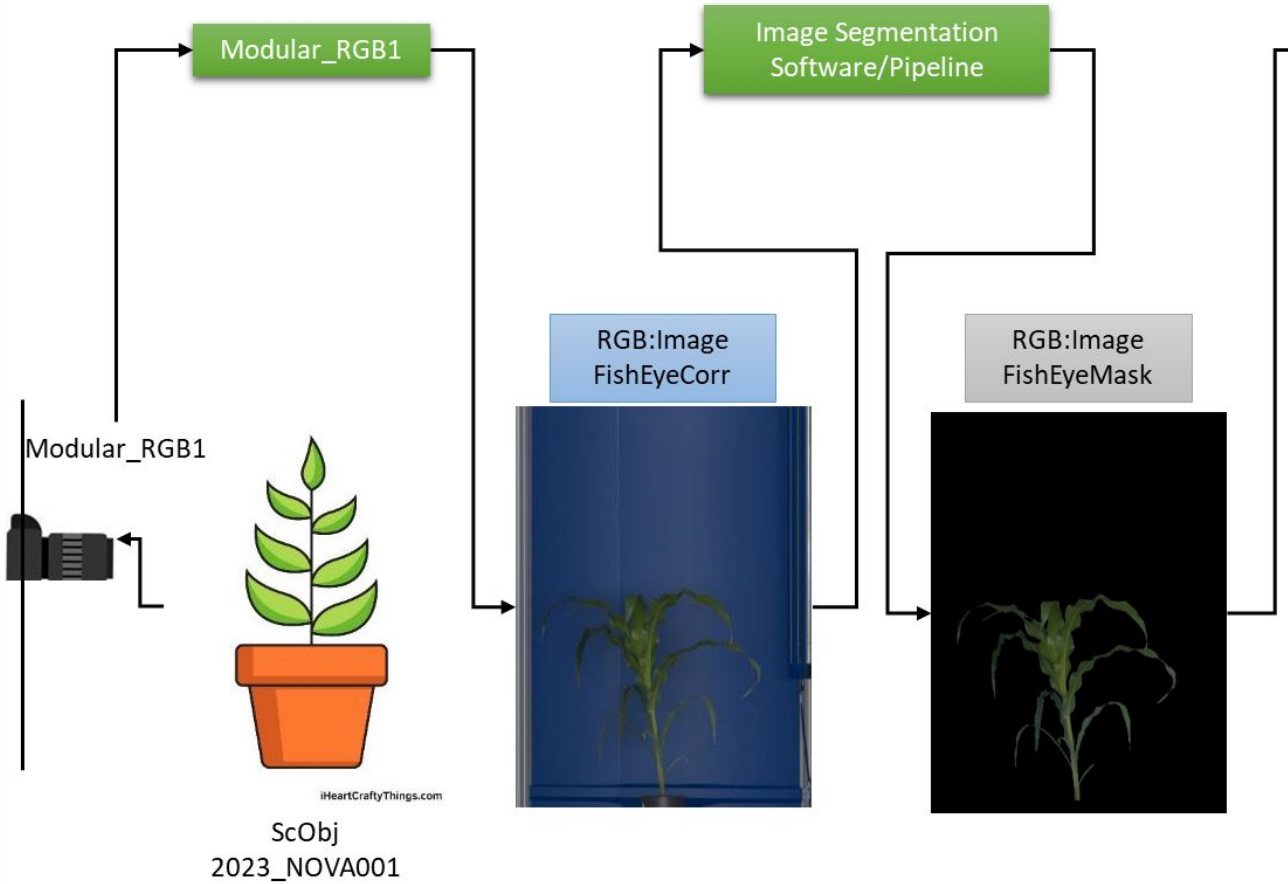
Finnish National Plant Phenotyping Infrastructure NaPPI




00-RoundProtocol
Fichier
Accueil
Partage
Affichage
Très grandes icônes
Grandes icônes
Icônes moyennes
Petites icônes
Liste
Détails
Volet de navigation
Volets
Disposition
Affichage actuel
Afficher/ Masquer
Options
« 01-R... › 00-RoundProtocol
Rechercher dans : 00-Round
Accès rapide
Ce PC
Jdwl
Téléchargements
Aurelie
Thiebault
OneDrive - University
https://datacloud.hels
nappi_hy_data
Series_001 (\\192.168.8
Films_001 (\\192.168.8
Cloud_001 (\\192.168.
RoundProtocol-69-4.txt
RoundProtocol-69-6.txt
RoundProtocol-69-8.txt
RoundProtocol-69-10.txt
RoundProtocol-69-14.txt
RoundProtocol-69-19.txt
RoundProtocol-69-25.txt
RoundProtocol-69-31.txt
RoundProtocol-69-34.txt
RoundProtocol-69-37.txt
RoundProtocol-69-39.txt
RoundProtocol-69-42.txt
RoundProtocol-69-48.txt
RoundProtocol-69-51.txt
RoundProtocol-69-53.txt
RoundProtocol-69-55.txt
RoundProtocol-69-58.txt
RoundProtocol-69-60.txt
RoundProtocol-69-62.txt
RoundProtocol-69-65.txt
RoundProtocol-69-68.txt
RoundProtocol-69-70.txt
RoundProtocol-69-72.txt
RoundProtocol-69-75.txt
RoundProtocol-69-79.txt
RoundProtocol-69-81.txt
RoundProtocol-69-83.txt
RoundProtocol-69-85.txt
28 élément(s)


P:\nappi_hy_data\GrpFd_Common\PHIS_Database\Brassica_Demo\01-RGB1\00-RoundProtocol\RoundProtocol-69-8.txt - No...
Fichier Édition Recherche Affichage Encodage Langage Paramètres Outils Macro Exécution Modules d'extension Documents ?
RoundProtocol-69-8.txt
1 <Measure>
2 <AdaptTime>00:00:00</AdaptTime>
3 <Prescription id="1" name="Recipe New:1">
4 <RGB1 height="Default">
5 <Offset>0</Offset>
6 <Frame angle="0" />
7 <Frame angle="120" />
8 <Frame angle="240" />
9 </RGB1>
10 <RGB2 height="Automatic">
11 <Offset>0</Offset>
12 <Frame angle="0" />
13 </RGB2>
14 <Analyse>
15 <MaskErosionLevel>1</MaskErosionLevel>
16 <RGB1>
17 <PlantMask>
18 <Formula>0-((2*R)+(3*B)-(5*G))</Formula>
19 <Threshold>0.1098039</Threshold>
20 <MedianFilterSize>4</MedianFilterSize>
21 <MinSize>100</MinSize>
22 <MinHoleSize>10</MinHoleSize>
23 <CropObjectsOnBorders>False</CropObjectsOnBorders>
24 <UseReflectionReduction>False</UseReflectionReduction>
25 <SkipBadExposedPoints>True</SkipBadExposedPoints>
26 </PlantMask>
27 <MorphoAnalysis>
28 <AREA_PX />
29 <AREA_MM />
30 <PERIMETER_PX />
31 <PERIMETER_MM />
32 <COMPACTNESS />
33 <WIDTH_PX />
34 <WIDTH_MM />
35 <HEIGHT_PX />
36 <HEIGHT_MM />
37 </MorphoAnalysis>
eXtensible Ma length : 7 472 lines : 182 Ln : 1 Col : 1 Pos : 1 Windows (CR LF) UTF-16 Little Endian INS

PHIS, Provenance



Tray ID	Angle	ShootSystem_ProjectedArea	ShootSystem_Height
2020_Maize001	0	4135.986114	1541.387944
2020_Maize001	120	3955.347733	1469.520866
2020_Maize001	240	4295.914362	1544.069552

Conclusion

Take home message

- **Variables:** Take time to define your variables will allow you to understand them, reuse them, share them

  => Portals exist to help you discovering Ontologies and Concepts! (Agroportal, BioPortal, Ontobee, etc.)
- **Datafiles:** study the storage solution best suited to **your needs and uses**

  Remember you can have separate solutions for your information system!
- **Provenance:** Make your data really **R**eusable!
- **Data Transfer Pipeline:** for improved automatization!

Thank you for your attention!

https://www.phenome-emphasis.fr/

OpenSILEX Team - http://opensilex.org/

http://www.phis.inrae.fr/

Special thanks to: Silvana Moscatelli, François Tardieu, Romain Chapuis, Anne Tireau, Renaud Colin …

EMPHASIS

emphasis@fz-juelich.de

emphasis.plant-phenotyping.eu

EMPHASIS_EU

EMPHASIS.EU

EMPHASIS on Plant Phenomics

ESFRI

EMPHASIS is an ESFRI-listed project.

EMPHASIS-PREP is funded by the
European Union (Grant Agreement: 739514).

EUROPEAN INFRASTRUCTURE
FOR PLANT PHENOTYPING