

# Common tools for variables naming and URI generation

april 16th and april 20th

# Session #1 :

# Presentation of the interface and models

16 april 2021

# Plan

- Presentation of concepts
  - FAIR data, identification and URI
  - Information system
  - URI model
  - Variable Model
- Presentation of Interface
  - Variables
  - URI
- Questions and answers

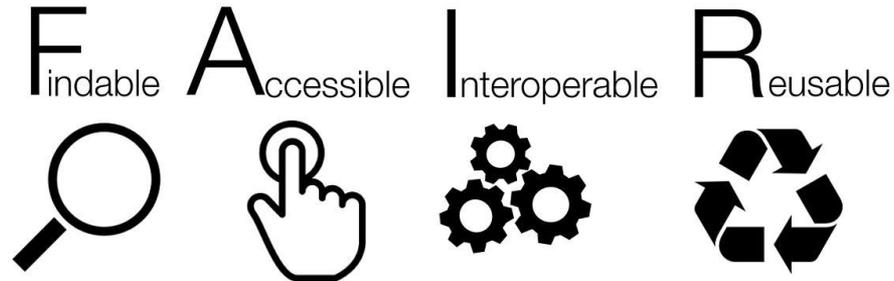
# FAIR data

**Findable** : URI, metadata and indexed portal

**Accessible** : open and standardized protocol, authentication

**Interoperable** : shared standardized format and ontologies

**Reusable** : provenance and domain relevant metadata



# FAIR data

Findable : URI, metadata and indexed portal

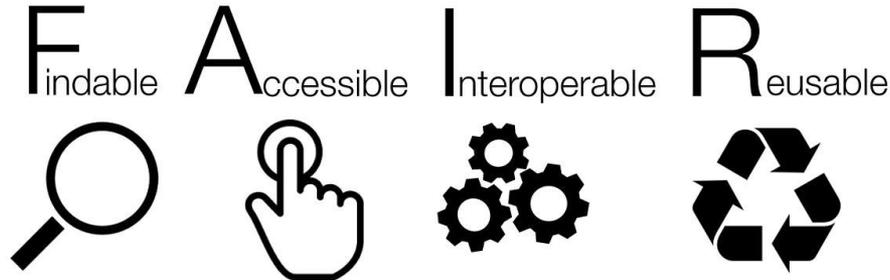


Accessible : open and standardized protocol, authentication

Interoperable : shared standardized format and ontologies



Reusable : provenance and domain relevant metadata



# Make FAIR data, structure your data

How ? Based on 2 key elements:

- **Identification and Naming convention**
  - Objects: plants, plots, experiments, sensors, events, etc
  - Persistent, unambiguous, resolvable, globally unique
  
- **Semantic and metadata**
  - Controlled vocabulary
  - Formalized relationships between entities
  - Data annotation and enrichment (search engine friendly)

# Information System

- An information system is a software that is used for data management. A lot of the tedious work is automated: e.g. generating URI.
- Thanks to identifiers, interactions between multiple elements are possible and reasoning is possible.
- Interaction with data is possible through web services allowing automation and creation of workflows.

# URI

- Every entry is unique, we don't want duplicates!
- URI accepts semantic parts, so use it wisely.
- Position of the plant is a common pattern

BUT position can change, avoid it if change is possible.

## Properties :

un-ambiguous, persistent, resolvable and stable.

*[https://your-institution.nat/installation/case\\_dependant](https://your-institution.nat/installation/case_dependant)*

Keep in mind to avoid metadata in the identifier

# Different resources, different URI schema

In small numbers like experiments, projects or documents

- You can give it a name including some year related content to be unique.  
BUT never rename it `field-expe:20hp03-sunrise-dyp12-preflo`

In big numbers you want to use some incremental schema

- Plants, plots, sensors, ... `field:set/devices/vocabularycapacitancesensor-hcsol0005`

In very large amount, different options

- Images and sensor output
- File + date `field-ev:26510077-81b1-4c2d-9514-97a6874a7fbe`
- Cryptographic numbers

# Presentation of the variable model

It is not ideal to create an ontology with all the phenotyping variables

Use a common pattern to create a variable name that anyone could easily understand

**Entity - Quality - Method - Unit**

Addition of the Germplasm (**Species**) is planned.

# Presentation of the variable model

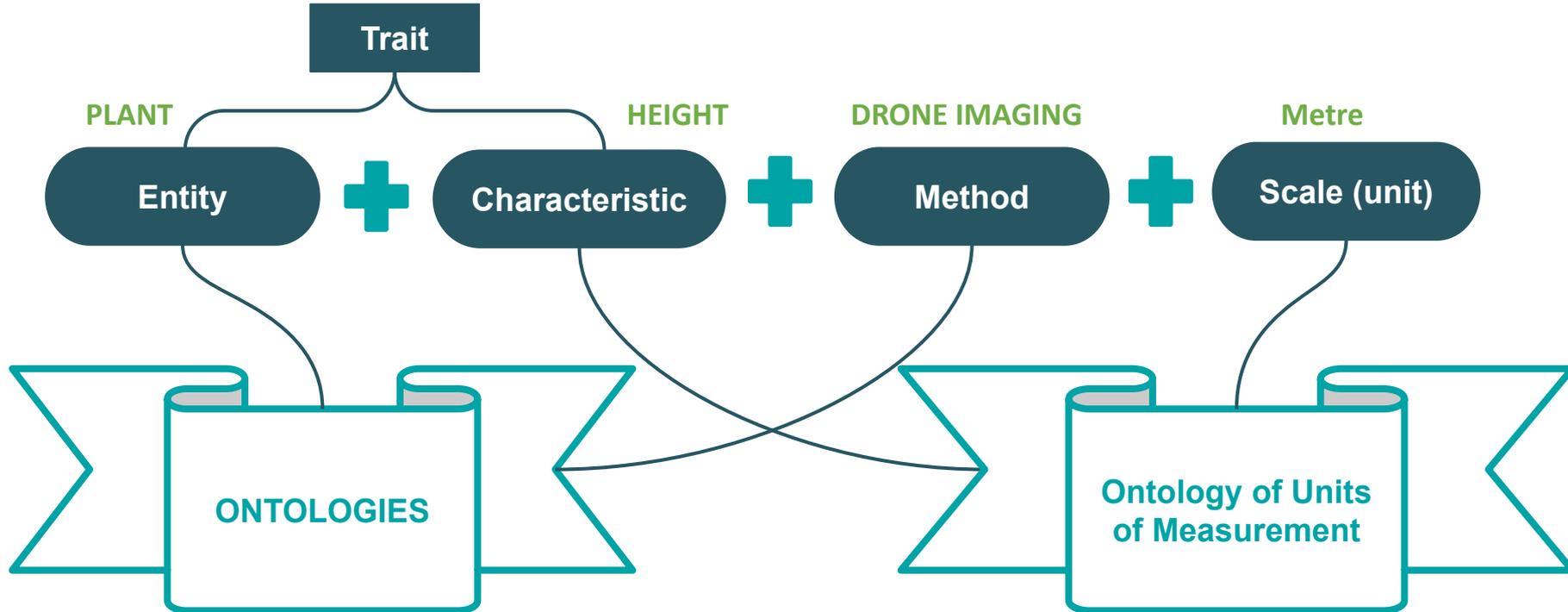
**Entity** is the target of the variable, what we are measuring on.

**Characteristic** is an element from Ontology of Units of Measurement, physical quantity, or quality.

**Method** is the the different ways we have to measure variable. Giving details of protocols and material can be useful.

**Unit** is an element from the Ontology of Units of Measurement

# Presentation of the variable model



# Variable declaration - Presentation of the interface

Common tool to declare and share variables

<http://138.102.159.36:8082/app/>

- id : guest@opensilex.org
- password : guest

The screenshot shows the EPPN Emphasis web interface. The top navigation bar is green with the EPPN logo and the text 'EPPN Emphasis'. A left sidebar contains a menu with items: 'Shared resources', 'Variables' (highlighted), 'Germplasm', 'Users management', and 'Scientific applications'. The main content area is titled 'Variables' and includes a sub-header 'Manage and configure variables, entities, characteristics'. Below this are four tabs: 'Variables', 'Entity', 'Characteristic', and 'Method'. The 'Variables' tab is active, showing a search bar with the text 'Search variables, plant height, plant, humidity, image processing,' and a message 'Showing 0 to 20 of 45 entries'.

The screenshot shows the 'Add variable' form. The title is 'Add variable'. There is a checked checkbox for 'URI' and a text field containing 'autogenerated URI'. Below this are two rows of selection fields. The first row has 'Entity' with a dropdown menu showing 'Grain' and a blue '+' button, and 'Characteristic' with a dropdown menu showing 'Humidity'. A blue error message 'Trait already existing in an ontology' is displayed below the first row. The second row has 'Method' with a dropdown menu showing 'Imaging' and a blue '+' button, and 'Unit/Level' with a dropdown menu showing 'Percentage'. Below these are two more text input fields: 'Name' containing 'grain\_humidity\_imaging\_percentage' and 'Alternative name' containing 'grain\_humidity'. At the bottom, there is a 'Data type' dropdown menu showing 'Decimal Number'.

# How to design variables in advance

File example:

<https://docs.google.com/spreadsheets/d/1s959UrLx0S4V39uLW201USnLeIK4CTQit72-A1qf9Jo/edit#gid=651449043>

A	B	C	D	E	F
	Unit name	Type	symbol	Alternative symbol	uri
area density unit	▼ kilogram per hectare		kg ha <sup>-1</sup>	kg/ha	<a href="http://purl.obolibrary.org/obo/VO_0000283">http://purl.obolibrary.org/obo/VO_0000283</a>
area density unit	▼ ton per hectare		t ha <sup>-1</sup>	t/ha	<a href="http://purl.obolibrary.org/obo/VO_0000323">http://purl.obolibrary.org/obo/VO_0000323</a>
area density unit	▼ microgram per square centimet	UnitDivision	µg cm <sup>-2</sup>	µg/cm <sup>2</sup>	
area density unit	▼ microgram per square meter	UnitDivision	µg m <sup>-2</sup>	µg/m <sup>2</sup>	
area density unit	▼ square meter per square meter	UnitDivision	m <sup>2</sup> m <sup>-2</sup>	m <sup>2</sup> /m <sup>2</sup>	
area unit	▼ square meter		m <sup>2</sup>		<a href="http://purl.obolibrary.org/obo/VO_0000080">http://purl.obolibrary.org/obo/VO_0000080</a>
area unit	▼ square centimeter		cm <sup>2</sup>		<a href="http://purl.obolibrary.org/obo/VO_0000081">http://purl.obolibrary.org/obo/VO_0000081</a>
area unit	▼ square millimeter		mm <sup>2</sup>		<a href="http://purl.obolibrary.org/obo/VO_0000082">http://purl.obolibrary.org/obo/VO_0000082</a>
area unit	▼ hectare		h		<a href="http://purl.obolibrary.org/obo/VO_0010010">http://purl.obolibrary.org/obo/VO_0010010</a>
base unit	▼ meter		m		<a href="http://purl.obolibrary.org/obo/VO_0000008">http://purl.obolibrary.org/obo/VO_0000008</a>
base unit	▼ kelvin		°K		<a href="http://purl.obolibrary.org/obo/VO_0000012">http://purl.obolibrary.org/obo/VO_0000012</a>

# Design variables

You can create your own URI, following the recommendations. This way you will find your own variables easily later on.

Variable uri	Variable ID	Variable name (long name in PHIS)	Variable abbreviation (name in PHIS)
<a href="http://phenome-emphasis/variable/4P:000001">http://phenome-emphasis/variable/4P:000001</a>	4P:0000001	Canopy_Height_Photogrammetry_Meter	C_H_P_m
<a href="http://phenome-emphasis/variable/4P:000002">http://phenome-emphasis/variable/4P:000002</a>	4P:0000002	Canopy_HeightStd_Photogrammetry_Meter	C_HS_P_m
<a href="http://phenome-emphasis/variable/4P:000003">http://phenome-emphasis/variable/4P:000003</a>	4P:0000003	Canopy_HeightFlag_Photogrammetry_Unitless	C_HF_P_uless
<a href="http://phenome-emphasis/variable/4P:000004">http://phenome-emphasis/variable/4P:000004</a>	4P:0000004	Soil_Height_Photogrammetry_Meter	S_H_P_m
<a href="http://phenome-emphasis/variable/4P:000005">http://phenome-emphasis/variable/4P:000005</a>	4P:0000005	Canopy_SR850nm675nm_BandCombination_Unitless	C_SR850nm675nm_BC_uless
<a href="http://phenome-emphasis/variable/4P:000006">http://phenome-emphasis/variable/4P:000006</a>	4P:0000006	Canopy_CI850nm570nm_BandCombination_Unitless	C_CI850nm675nm_BC_uless
<a href="http://phenome-emphasis/variable/4P:000007">http://phenome-emphasis/variable/4P:000007</a>	4P:0000007	Canopy_CI850nm730nm_BandCombination_Unitless	C_CI850nm570nm_BC_uless
<a href="http://phenome-emphasis/variable/4P:000008">http://phenome-emphasis/variable/4P:000008</a>	4P:0000008	Canopy_CI850nm710nm_BandCombination_Unitless	C_CI850nm710nm_BC_uless
<a href="http://phenome-emphasis/variable/4P:000009">http://phenome-emphasis/variable/4P:000009</a>	4P:0000009	Canopy_MCARI570nm730nm850nm_BandCombination_Unitless	C_MCARI570nm730nm850nm_BC_uless
<a href="http://phenome-emphasis/variable/4P:000010">http://phenome-emphasis/variable/4P:000010</a>	4P:0000010	Canopy_MND450nm530nm850nm_BandCombination_Unitless	C_MND450nm530nm850nm_BC_uless
<a href="http://phenome-emphasis/variable/4P:000011">http://phenome-emphasis/variable/4P:000011</a>	4P:0000011	Canopy_MND450nm570nm850nm_BandCombination_Unitless	C_MND450nm570nm850nm_BC_uless
<a href="http://phenome-emphasis/variable/4P:000012">http://phenome-emphasis/variable/4P:000012</a>	4P:0000012	Canopy_MND450nm675nm850nm_BandCombination_Unitless	C_MND450nm675nm850nm_BC_uless
<a href="http://phenome-emphasis/variable/4P:000013">http://phenome-emphasis/variable/4P:000013</a>	4P:0000013	Canopy_MND450nm730nm850nm_BandCombination_Unitless	C_MND450nm730nm850nm_BC_uless
<a href="http://phenome-emphasis/variable/4P:000014">http://phenome-emphasis/variable/4P:000014</a>	4P:0000014	Canopy_MND850nm730nm450nm_BandCombination_Unitless	C_MND850nm730nm450nm_BC_uless

# How to design variables ?

- Recommendations to create URI :

<https://hal.archives-ouvertes.fr/hal-02390920/document>

- Use components from different Ontologies :

Units and characteristics : <https://github.com/HajoRijgersberg/OM#om>

- Browse different concepts and ontologies :

<http://agroportal.lirmm.fr/>

# The problem with 'Traits' and Methods

'Trait' (Entity-Characteristic)	Method		
	Method 1	Method 2	Method 3
Trait 1			
Trait 2			
Trait 3			

Do Traits and Methods have a meaning independent of each other, or can they only be interpreted together, within the context of a variable?

# Example. Modelling option 1

Trait		Method	Unit
Entity	Characteristic		
Canopy	Height	ImageAnalysisAriBased	cm
Canopy	Height	ImageAnalysisExcessGreenBased	cm

Do both ways of measuring 'Height' give the same result?

If yes, the Method is just a refinement of the way the Trait was measured. As a consequence:

- Two Traits can be compared without needing the Method
- The difference between Method and Provenance is fuzzy. Why not leave out the method?

If not, the Method is an essential semantic refinement of the Trait. As a consequence:

- You cannot compare two Traits without the Method
- Method will be tightly connected to a Trait

## Example. Modelling option 2

Trait		Method	Unit
Entity	Characteristic		
Canopy	HeightAriBased	ImageAnalysis	cm
Canopy	HeightExcessGreenBased	ImageAnalysis	cm

Consequences:

- The concept of Method is more or less empty, because the definition of the Trait tells all

# Germplasm

To be completed later with the addition of a 5th element to variables.

Selected Germplasm 0 Actions ▾

Showing 0 to 17 of 17 entries

<input type="checkbox"/>	Name	Type	Species
<input type="checkbox"/>	Lambada	Variety	Fragaria
<input type="checkbox"/>	Mount Everest	Variety	Fragaria
<input type="checkbox"/>	starlette	Variety	Fragaria
<input type="checkbox"/>	Deluxe	Variety	Fragaria

 **Sweet Charlie**  
Germplasm

← Details Annotations Documents

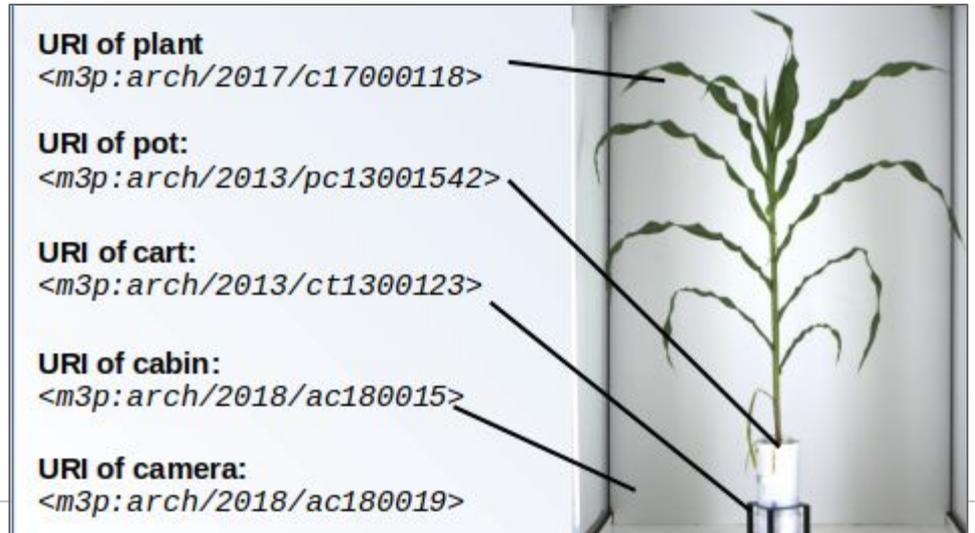
 Description  

**URI** Test-germplasm:variety.sweet-charlie  
**Type** Variety  
**Name** Sweet Charlie  
**Species:** Fragaria  
[http://aims.fao.org/aos/agrovoc/c\\_3076](http://aims.fao.org/aos/agrovoc/c_3076)

# URI/QR Code generation

URI : Uniform Resource Identifier

- Standardized, Un-ambiguous, Actionable
- Generated by tools under responsibility of scientific coordinator
- Use URI for every objects



# URI/QR Code generation - Presentation of the interface

Common tool to generate URI and QR Code

<http://138.102.159.36:8082/app/>

The screenshot shows a web interface for managing applications. On the left is a sidebar with 'Other applications' selected. The main area is titled 'Other applications' and contains buttons for '+ Add application' and 'Reload application server'. Below these is a status bar indicating 'Application server status : Online'. A search bar contains the text 'Detection of outlier, Data map representation, etc.'. At the bottom, a table displays application entries.

Name	Description
URI and QR codes generator	Generate URI and QR codes

The form is divided into three numbered steps:

- 1 Import your file**  
A file with one row for each resource you want to generate a URI for.  
Parcourir... Aucun fichier sélectionné.  
Details: ▾
- 2 Host Name**  
test  
Installation name  
your installation  
Object Type  
plant ▾
- 3 Data to put in the URI**  
Project related  
aProject  
Year  
2020

**Generate URI**

You can find the file in your Download folder (default settings)

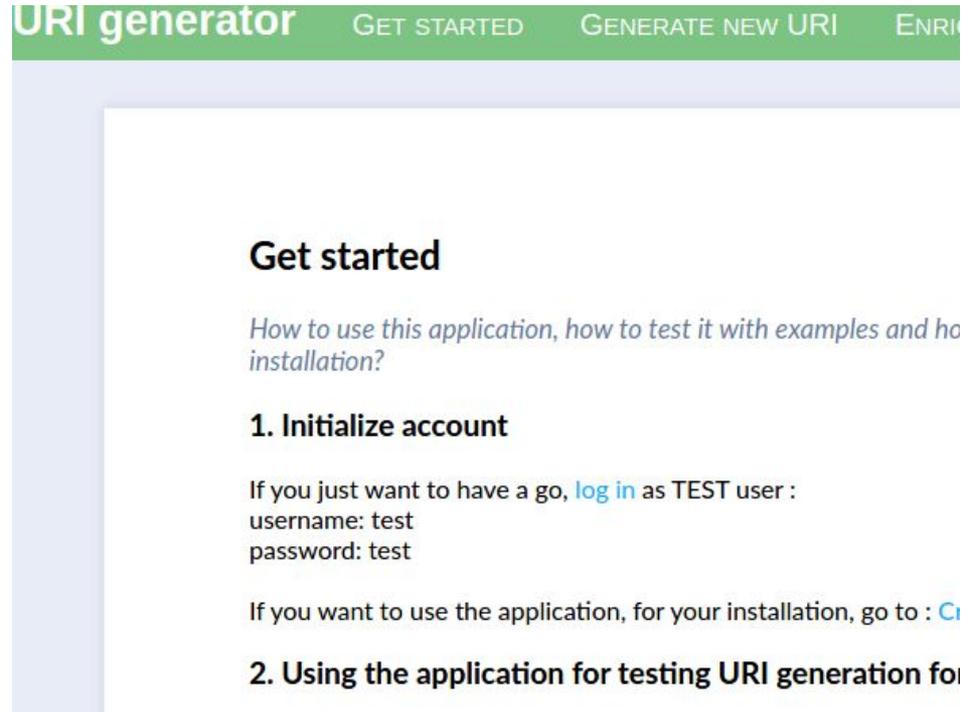
# Stand alone app

This app was designed to be stand alone.

The code is accessible [here](#)

Documentation within the app. Including data.

Can be modified (in python) if you need an adjustment



The screenshot shows the top navigation bar of the 'URI generator' application. The bar is green with white text for the title and navigation links. Below the bar, the main content area is white with a light blue border on the left and top. The text 'Get started' is prominently displayed, followed by a sub-heading and a list of instructions for using the application.

**URI generator** GET STARTED GENERATE NEW URI ENRI

## Get started

*How to use this application, how to test it with examples and how to install it?*

- 1. Initialize account**

If you just want to have a go, [log in](#) as TEST user :  
username: test  
password: test

If you want to use the application, for your installation, go to : [C...](#)
- 2. Using the application for testing URI generation for**

# Preparation of the next session *20th april 2021*

## Next Session's Agenda :

- Reminders of previous Session
- Practice on your own cases / Discussion on the best practices
- Short presentation of the implementation of these tools in the information system PHIS

## “Homework” :

- Play with the URI / QR code generator
- Think about variables you'll need to declare
- Bring your files