



HAL
open science

Semantics and plant phenotyping data structuration for data analytics

Pierre-Etienne Alary, Pascal Neveu, Anne Tireau, Nadine Hilgert, Llorenç Cabrera Bosquet, Francois Tardieu

► **To cite this version:**

Pierre-Etienne Alary, Pascal Neveu, Anne Tireau, Nadine Hilgert, Llorenç Cabrera Bosquet, et al.. Semantics and plant phenotyping data structuration for data analytics. 7. International Food Legumes Research Conference (IFLRC VII), May 2018, Marrakech, Morocco. hal-02791501

HAL Id: hal-02791501

<https://hal.inrae.fr/hal-02791501>

Submitted on 5 Jun 2020

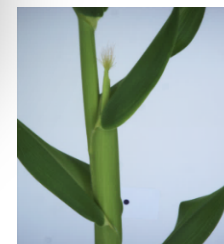
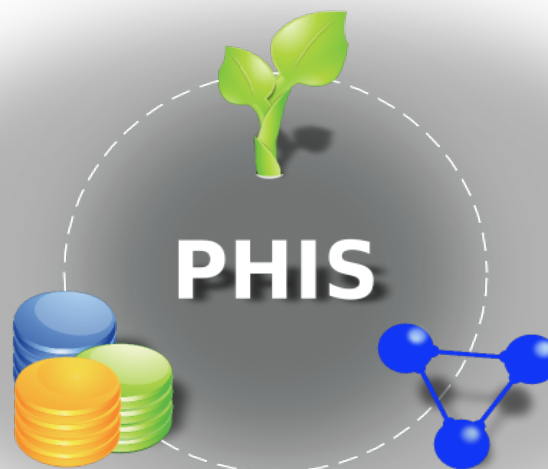
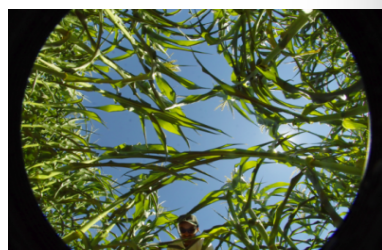
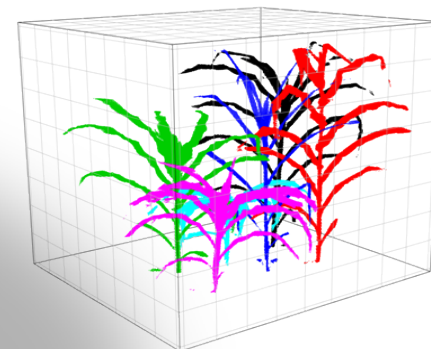
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

Semantics and Plant Phenotyping Data Structuration for Data Analytics



Phenotyping Hybrid Information System

***Pierre-Etienne Alary, Llorenç Cabrera-Bosquet,
Anne Tireau, Morgane Vidal, ... , Pascal Neveu***

MISTEA-LEPSE, INRA-Montpellier, France

Phenotyping infrastructure: multi-scale approach

Aim : Characterize large collections of genotypes at high-throughput enabling quantitative genetics studies (GWAS – GS)



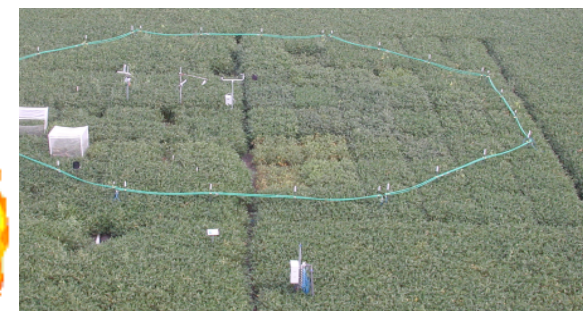
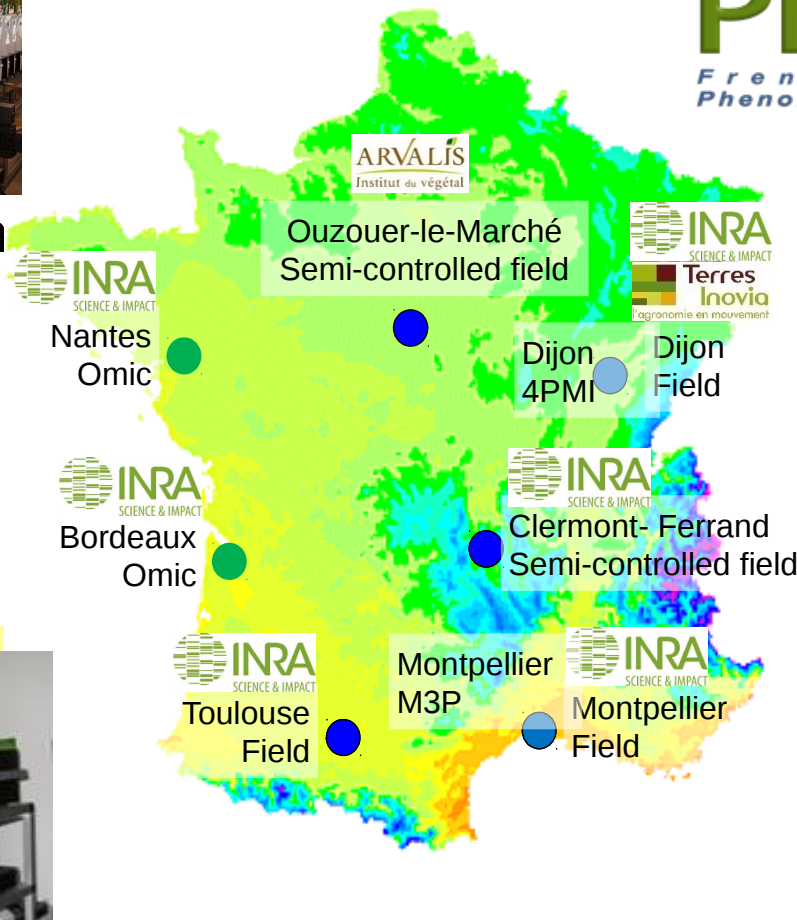
Controlled condition

- Traits not measurable in the field
- Individual plants
- Model parameters



High throughput omics

- Composition and quality traits



Field

- Canopies
- Validation in agronomic conditions
- Yield

Use case: PhenoArch generates daily...

... Structured data

- > 44000 raw and segmented images
- ca. 200000 phenotypic observations
- >10000s points of environmental data

... Non structured data

- Visual scores
- Lab book notes
- Events
- Measurements in companion platforms or compartments

But also and very important...

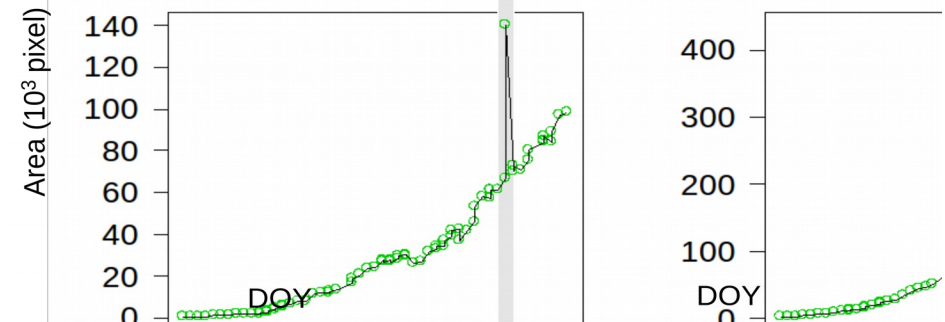
- *Plant accidents*
- *Human mistakes*
- *platform failures*
- *↑ number of non exploitable / misleading data*



Side images



Top images



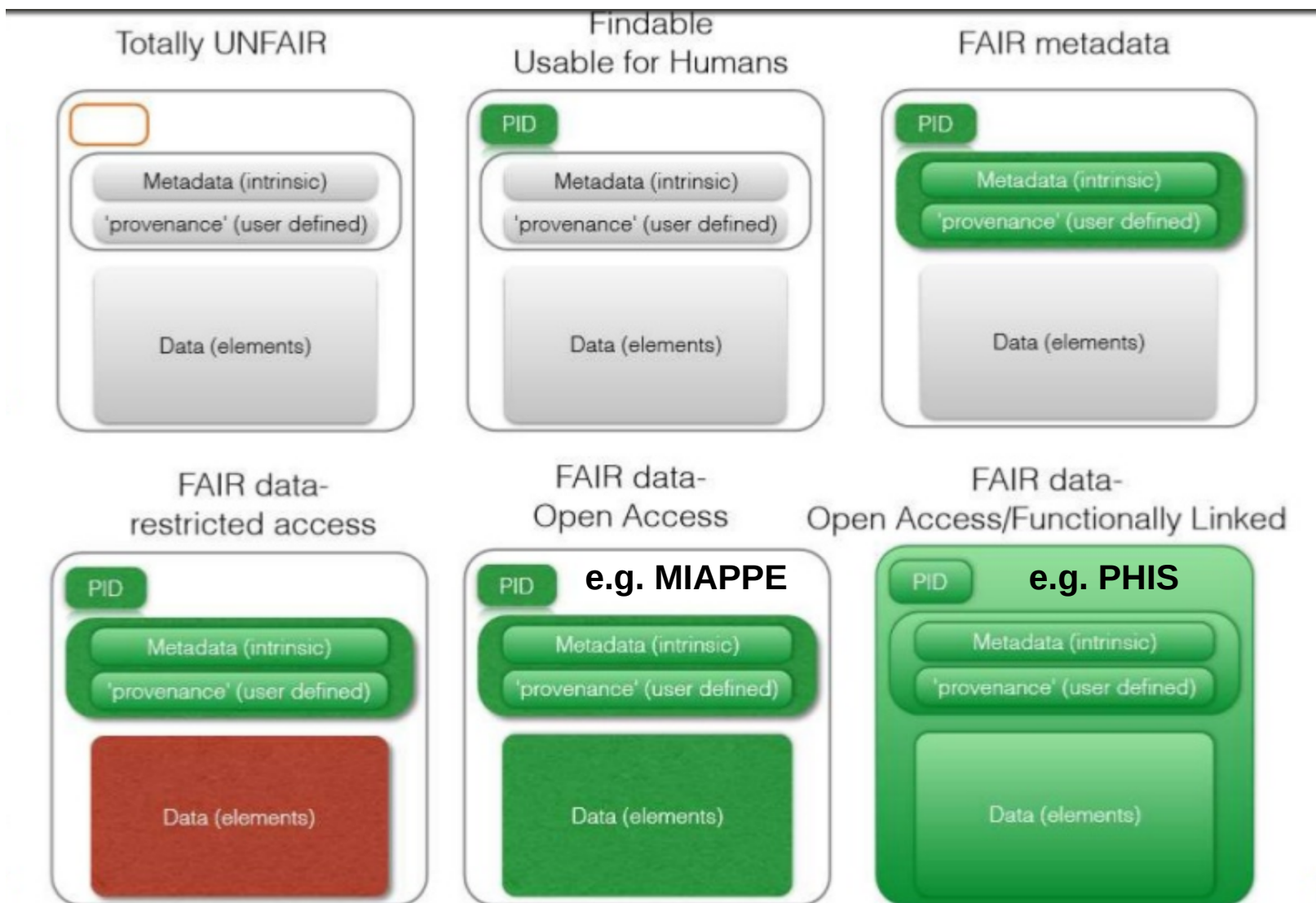
We need an Information system capable to

- **Store, organise and manage:**
 - Highly heterogeneous data (e.g. images, growth curves)
 - Multi-spatial and temporal scale data (leaf to canopy level)
 - Multi-source data (field, platform)
 - Environmental data
- **Enrich datasets with knowledge and metadata (enable reuse of data and meta-analyses)**
- **Interoperate and integrate data into external resources (e.g. modelling platforms or external databases)**
- **Provide FAIR data (functionally linked data)**

Findable, Accessible, Interoperable, Reusable

FAIR DATA

- *Findable, Accessible, Interoperable, Reusable*



PHIS approaches


- **Hybrid system based on different storage databases** (PostgreSQL, MongoDB, iRODS, RDF4J)
- **Object identification and organisation** (plants, sensors, plant organs, plots, etc.) based in:
 - Persistent Identifiers (e.g.URI, IRI, doi)
 - Two specific **application Ontologies (OWL)** formalise Objects and Events
- **Enrichment of data with knowledge and metadata via annotation**
- **User-friendly and interactive web interface**
- **Flexible Web Services for the integration of a diverse and multi-source data**

URIs (Uniform Resource Identifiers)

⇒ standardized, unique, unambiguous identification of objects



Prefix m3p: `<http://phenome-fppn.fr/m3p>` (a)



URI of plant:
`<m3p:arch/2017/c17000118>`

URI of pot:
`<m3p:arch/2013/pc13001542>`


URI of cart:
`<m3p:arch/2013/ct1300123>`

URI of cabin:
`<m3p:arch/2018/ac180015>`

URI of camera:
`<m3p:arch/2018/ac180019>`

URI of image:
`<m3p:arch/2017/ic17002295855>`

Prefix diaphen: `<http://phenome-fppn.fr/diaphen>` (b)



URI of plot:
`<diaphen:2017/o1700029>`

URI of plant:
`<diaphen:2017/17000147>`

URI of leaf:
`<diaphen:2017/l17000590>`

URI of camera:
`<diaphen:2018/ac180002>`

URI of image:
`<diaphen:2017/ic14001480237>`

How do we link all these objects?

OEPO (Ontology for Experimental Phenotypic Objects)


=> assigns types to **objects** involved in phenotyping experiments (e.g. sensors, plants, experimental context) and defines a specialization hierarchy between them according to the specificities of the installations and experiments.

OEEv (Ontology of Experimental Events (OEEv))

=> characterises **events** that occur during an experiment, (e.g. dates of sowing, application of a given treatment, harvesting), experimental management or any category of technical problem.

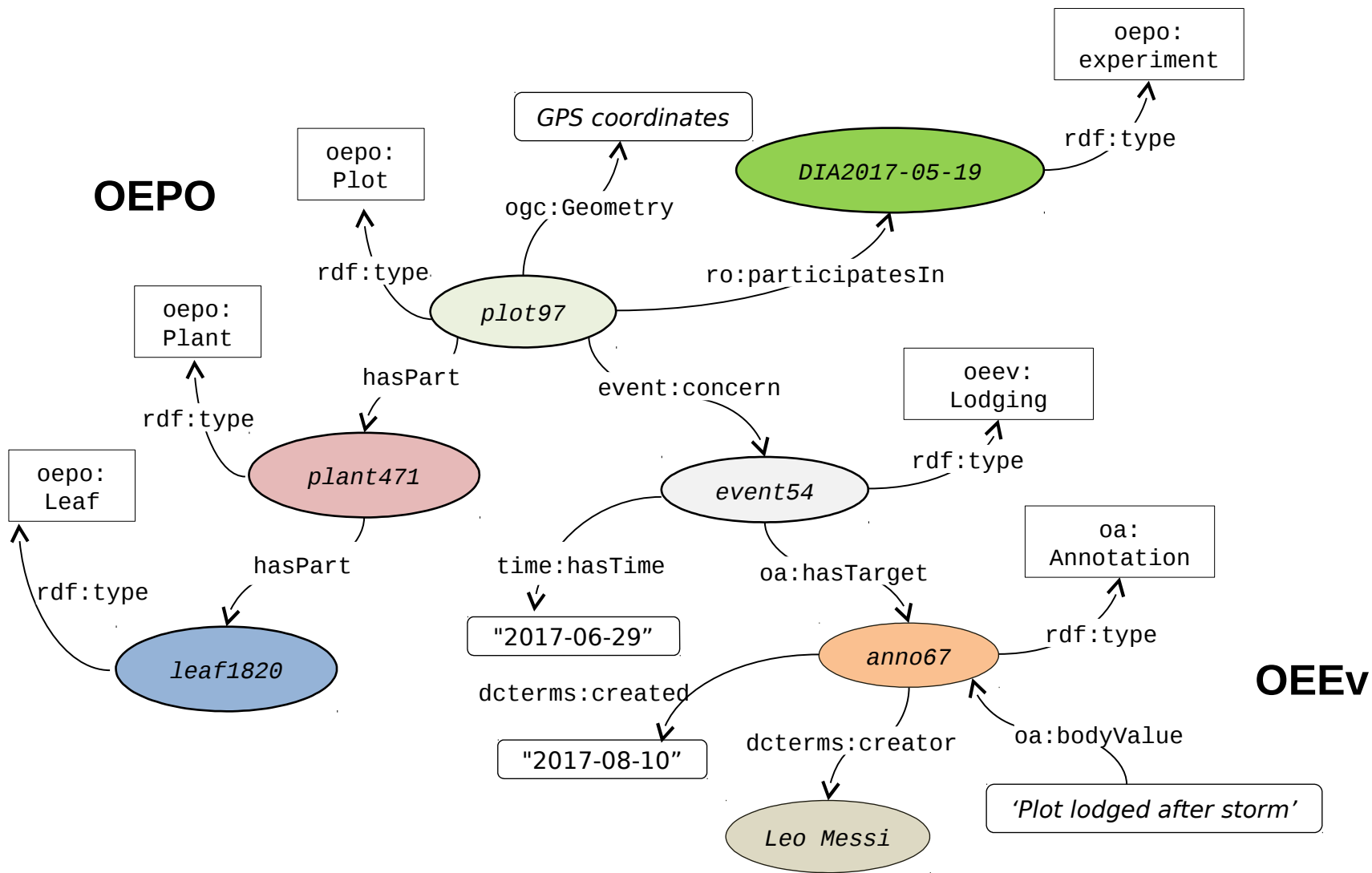
Whereas OEPO and OEEv refer to objects and events that can be specific to either field or controlled conditions, a number of the terms and definitions follow recommendations adopted by **EMPHASIS** and reuses **reference ontologies** including:

- Crop Ontology
- Plant Ontology
- FAO/Bioversity Multi Crop Passport Descriptors
- Semantic Sensor Network Ontology
- PATO
- PPEO (MIAPPE)
- other ontological resources such as the AGROVOC
- ...

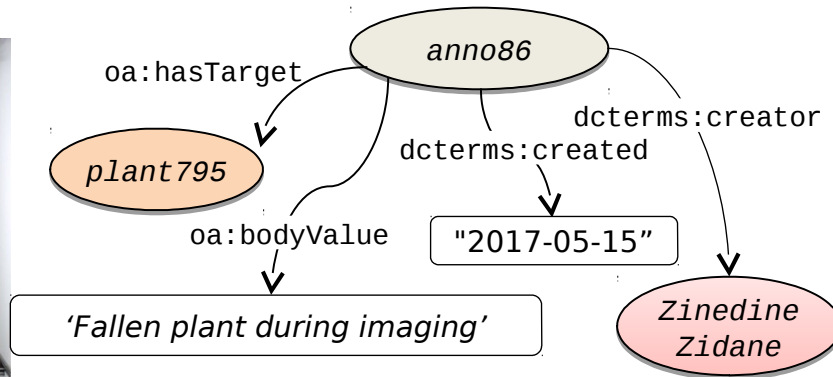
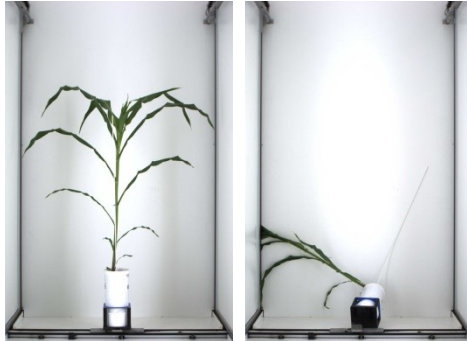
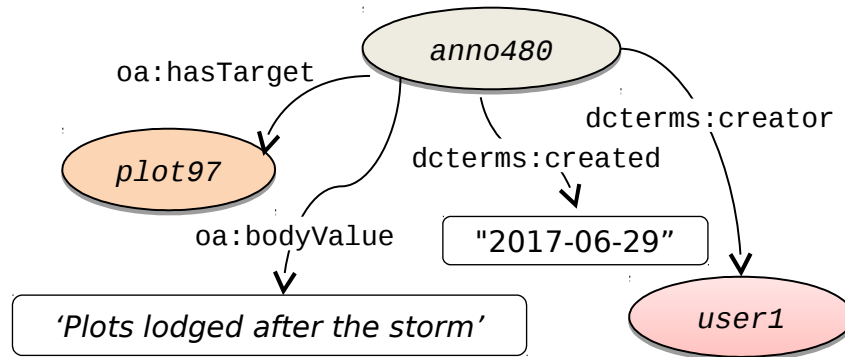


Reference ontologies

OEPO and OEEv Application Ontologies



Event Annotation



Events follow the [Web Annotation Data Model](#) that allows assigning motivation and purpose attributes to annotations (e.g. `oa:describing`, `oa:replying`, etc.). [Dublin Core](#) properties such as `dcterms:created` or `dcterms:creator` are also used.

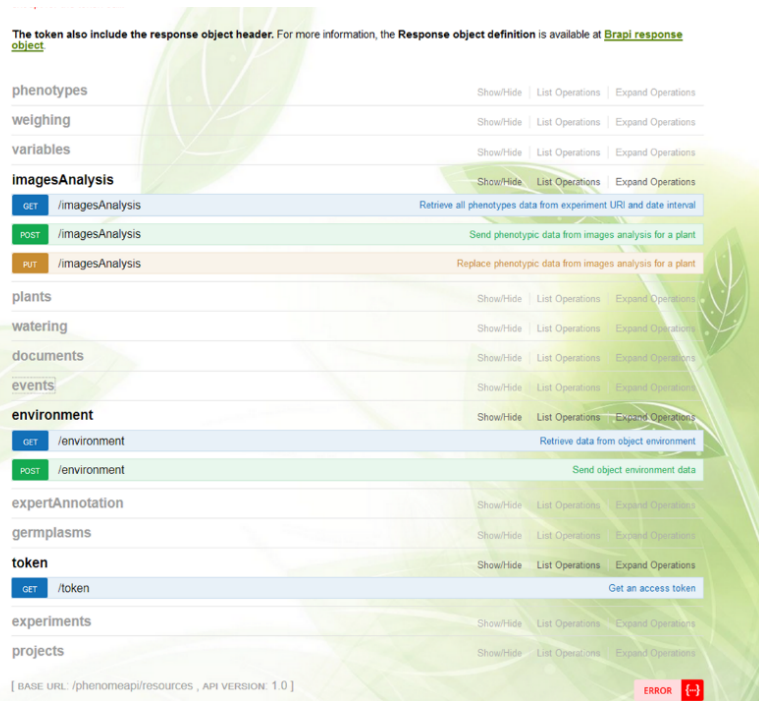
Event Annotation

On the web interface, clicking on a data point automatically displays the images associated to this point together with the associated annotations

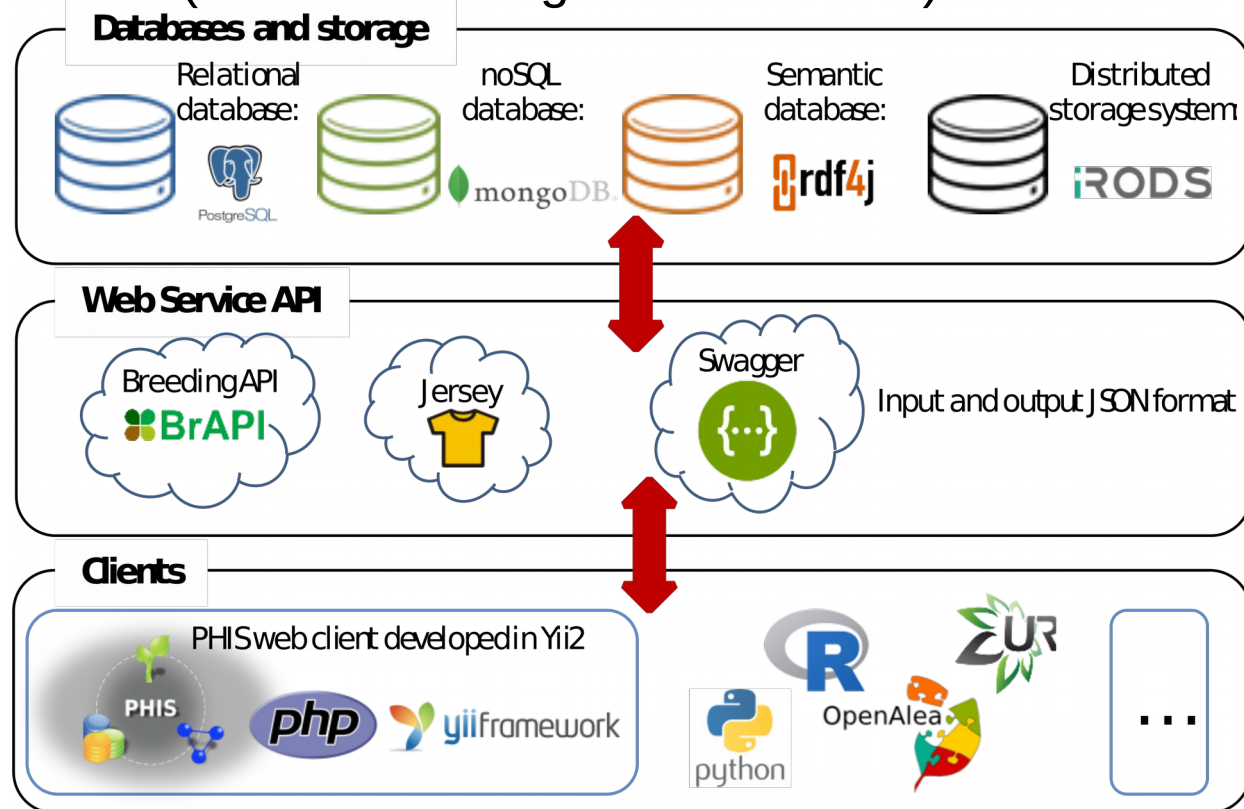
The screenshot displays the Phenotyping Hybrid Information System M3P interface. At the top, there are navigation menus for 'Experimental Organization', 'Data', 'Tools', and a user profile for 'Pierre-Etienne Alary'. Below the navigation, there are four small images of plants in pots. A tooltip above the graph shows the date 'Friday, Apr 28, 04:27:01' and a data point: '0017/DZ_PG_20/ZM4344/WD/Veg_1/01_17/ARCH2017-03-30~Convex Hull Plant Area~side60~lepse: 243,378.00'. The main graph plots 'convex_hull_area' on the y-axis (ranging from -250k to 2,000k) against 'time' on the x-axis (ranging from 17. Apr to 8. May). Multiple colored lines represent different plant samples, showing an overall upward trend. A legend at the bottom identifies the samples: 0010/DZ_PG_19/ZM4367/WD/Veg_1/01_10/ARCH2017-03-30~Convex Hull Plant Area~side60~lepse (Images), 0017/DZ_PG_20/ZM4344/WD/Veg_1/01_17/ARCH2017-03-30~Convex Hull Plant Area~side60~lepse (Images), 0063/DZ_PG_41/ZM4378/WW/Rep_1/02_03/ARCH2017-03-30~Convex Hull Plant Area~side60~lepse (Images), and 0091/DZ_PG_18/ZM4373/WW/Rep_1/02_31/ARCH2017-03-30~Convex Hull Plant Area~side60~lepse (Images). On the right, a 'PHIS - Mozilla Firefox' browser window is open, displaying an 'Event or Expert Annotation' form. The form fields include: Author: pierre-etienne.alary@supagro.fr; IP: 10.146.2.250; Confidential: (oui); Target (choose one): plant 0017/DZ_PG_20/ZM4344/WD/Veg_1/01_17/ARCH2017-03-30~Convex Hull Plant Area~side60~lepse, nearby image side60 2017-05-12T07:56:37+02:00; Datetime Event: 2017-05-12T07:45:07+02:00; Category: a dropdown menu; Subject: an empty text field; Content: an empty text area; and a 'save' button.

Web Service API

- **RESTFul** (Representational state transfer)
- Services available using URIs.
- Developed in Java with **Jersey** implementation of JAX-RS (Java API for RESTful Web Services) standard
- Implements the **Breeding API** : new version (v1.2) to be implemented
- Swagger interactive documentation (data-interchange format JSON)



API documentation (Swagger)



PHIS architecture

Environmental sensors

Each object is related to different objects, properties and data

Phenotyping Hybrid Information System Experimental Organization Data Tools Llorenç Cabrera-Bosquet

Home / Device / View / tair06_p

tair06_p

[Return to the list](#)

Device Alias	tair06_p
URI	http://www.phenome-fppn.fr/m3p/arch/2011/sa110018
Device Type	TemperatureSensor
Related Concept	http://purl.oclc.org/NET/ssnx/meteo/aws#ElectricalResistanceThermometer
Brand	CAMPBELL SCIENTIFIC
Model	HMP45C
Position (X,Y)	Greenhouse PhenoArch
Position (meter)	<i>(not set)</i>
Variable	air temperature:weather station:degree celsius
In Service Date	2011-05-01
Date of Purchase	2011-01-01
Date of last Calibration	2017-03-22
Documents	Hmp45c.pdf

Upload Document

Select files... [Browse ...](#)

[Add Document](#)

Data of tair06_p

air temperature:weather station:degree celsius

Sunday, Sep 17, 10:45
tair06_p: 27.9

15.Sep 12:00 16.Sep 12:00 17.Sep 12:00 18.Sep 12:00 19.Sep 12:00 20.Sep 12:00 21.Sep 12:00

Periode of time
 Week Month Range

2017-09-15 2017-09-22

Data visualisation

- Display of images, and time courses of phenotypic and environmental variables are automatically adapted to the particular experiments and variables
- Raw images, segmented images and metadata can be displayed in both field and greenhouse experiments

Phenotyping Hybrid Information System M3P

Experimental Organization Data Tools Lioreny Cabrera-Bosquet

Graphic Visualization

Search Criteria ↕

Experiments (ex. ARCH2016-04-15)
ARCH2014-02-10 ×

Genotypes (ex. Lo1270_H)
Select genotype

Plants list (ex. 0001/Lo1270_H/ZM2887/G151/WD3/01_01/ARCH2016-04-15) (1047 plant(s))
2058/7G-015/316-10/1079/D/ARCH2014-02-10 ×

Variables Selection

Imaging (ex. Convex Hull Plant Area)
imaging | Convex Hull Plant Area ×

View Type (ex. side60)
side0 ×

Watering and Weighing (ex. warerweightAfter)
watering | weightAfter ×

Weather (ex. micrometeor humidity)
Select a variable

Metadata

Include Images with 5 images to display

Include Annotations

Show Graphics

© INRA MISTEA-LEPSE 2014-2017 (PHIS v.2.6 - 04th October 2017) ; © INRA MISTEA - SILEX



Date: 2017-07-04 20:33

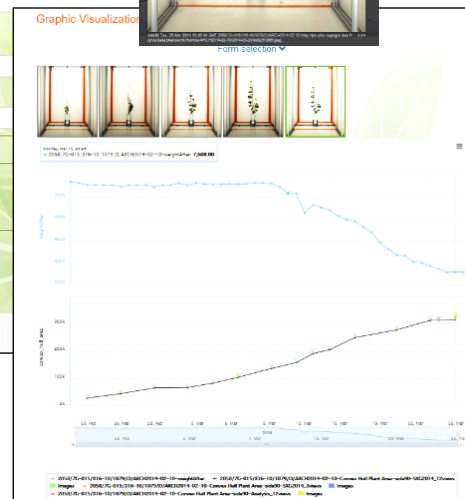
Provider: Saphen

View Type: warrersphere

Panel ID: Wb/News/phenotype-kpi-818/saphen2017-07-04/000273/27302_PG_22242259/021/04/2017-05-18

Image Visualization

© INRA MISTEA-LEPSE 2014-2017 (PHIS v.2.6 - 04th October 2017) ; © INRA MISTEA - SILEX



Variable management and interoperability

Home / Variables / Create Variable

Create Variable

Variable Label *

MyNewTrait_MyNewMethod_NA

Trait

Trait label -

Internal Label

MyNewTrait

Comment

This is a comment for y new trait, on which my new variable is focused.

Method

Method label -

Internal Label

MyNewMethod

Comment

This is a comment for my new method, used to produce the values of my new variable.

Unit

Unit label +

Ontologies References

In order to fill ontological references (URI) you can go to these ontologies :

- [AGROPORTAL](#) ⓘ
- [AGROVOC](#) ⓘ
- [PLANT ONTOLOGY](#) ⓘ
- [PLANTEOME](#) ⓘ
- [CROP ONTOLOGY](#) ⓘ
- [UNIT ONTOLOGY](#) ⓘ

Variable = Trait + Method + Unit

PHIS includes extensible scientific computing modules based on R packages for calculating elaborated variables and generating experimental summaries and reports.

Global Greenhouse Report

Name	Global Greenhouse Report
URI	http://www.phenome-fppn.fr/id/analysis/daglobal
Description	Visualization of a specified variable of an experiment. A HTML report is produced by this program.
Documents	

Experiment *

ARCH2017-03-30

Trait *

objectSumArea

View Label *

side90

Run

Data Analysis

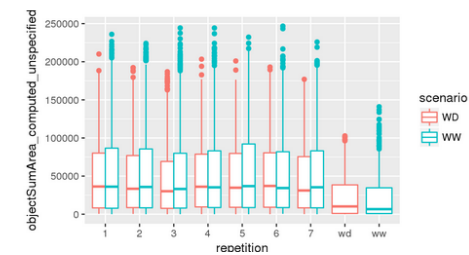
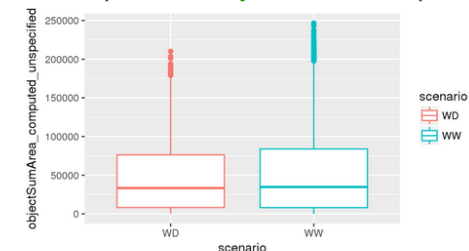
Showing 1-5 of 5 items.

#	Name ↓		URI	Description
1	Daily report	Greenhouse	http://www.phenome-fppn.fr/id/analysis/dailyreportphis	Daily description of a PhenoArch experiment (imagery, environnement and so on...) running of it. A HTML report is produced by this program.
2	Environment Report	Field	http://www.phenome-fppn.fr/id/analysis/daenvirfield	Description of the environment of a field experiment (meteo...). A HTML report program.
3	Environment Greenhouse Report		http://www.phenome-fppn.fr/id/analysis/daenvir	Description of environment of PhenoArch experiment (meteo...). A HTML report program.
4	Global Report	Greenhouse	http://www.phenome-fppn.fr/id/analysis/daglobal	Visualization of a specified variable of an experiment. A HTML report is produced by
5	Thermal Calculation Report	Time	http://www.phenome-fppn.fr/id/analysis/dathermal	For a PhenoArch experiment, a thermal time is calculated according to the user's parent's metho). A HTML report and a csv file are produced by this program.

ARCH2017-03-30 - ZA17 experiment on objectSumArea parameter and side90 view

- 60 genotypes
- 2 scenarios: WD, WW
- 9 repetitions
- 1467 pots
- Scientist supervisors: Cabrera-Bosquet, Tardieu, Turc, Welcker
- Technical supervisors: Brichet, LUCHAIRE, Suard
- Experiment performed from 2017-03-30 to 2017-06-30
- Genotypes used in this experiment: IPG004, IPG007, IPG017, IPG026, IPG029, IPG062, IPG063, IPG066, IPG073, IPG077, IPG082, IPG089, IPG101, IPG103, IPG109, IPG110, IPG111, IPG116, IPG117, IPG119, IPG120, IPG121, IPG128, IPG131, IPG136, IPG138, IPG146, IPG148, IPG152, IPG153, IPG155, IPG158, IPG159, IPG164, IPG165, IPG167, IPG169, IPG173, IPG176, IPG181, IPG188, IPG189, IPG190, IPG194, IPG195, IPG202, IPG216, IPG228, IPG233, IPG234, IPG239, IPG303, IPG304, IPG310, IPG311, IPG312, IPG313, IPG314, IPG318, IPG321

Description of objectSumArea parameter



In short PHIS:

- ✓ Allows management of huge and complex data thanks to a flexible design
- ✓ Enables and facilitates cloud computing
→ distributed computing, distributed storage
- ✓ Focuses on Data Provenance in order to produce reproducible research
- ✓ Is based on Open technologies
- ✓ Uses international identification systems (URI and DOI)
- ✓ Uses Semantics (ontologies, standardized vocabularies)
- ✓ Allows portal interoperability (towards BrAPI v1.2 compliance)



MISTEA

Pascal Neveu
Anne Tireau
Morgane Vidal
Arnaud Charleroy
Guilhem Heinrich
Nadine Hilgert
Isabelle Sanchez
Pierre-Etienne Alary



LEPSE

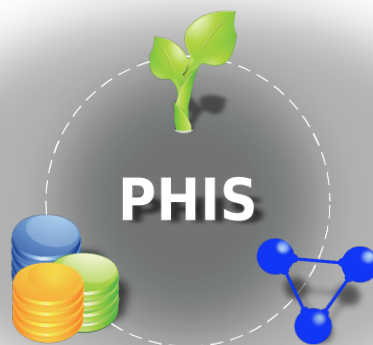
François Tardieu
Vincent Nègre
Jonathan Mineau-Cesari
Nicolas Brichet
Llorenç Cabrera-Bosquet



PHENOME partners

Patrick Moreau
Cyril Pommier
Romain Chapuis
Boris Adam
Philippe Burger
...

pierre-etienne.alary@inra.fr
pascal.neveu@inra.fr



www.plant-phenotyping.eu



EMPHASIS-PREP: Preparation of European Infrastructure for Multi-Scale Plant Phenotyping And Simulation for Food Security in a Changing Climate

