Systèmes d'information et analyses de données

Journée PHENOME-EMPHASIS 13 avril 2023,

INRAE-Paris

Pascal Neveu, Cyril Pommier,

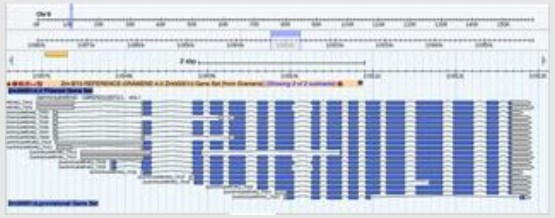
Isabelle Alic

ANR11-INBS-0012

Genome / Environment / Phenome

PHENOME EMPHASIS FRANCE

Genetics & Genomics



Phenomics

Plant Breeding
Meta analyses

Climate Change Studies
Genotype by Environment
Phenology

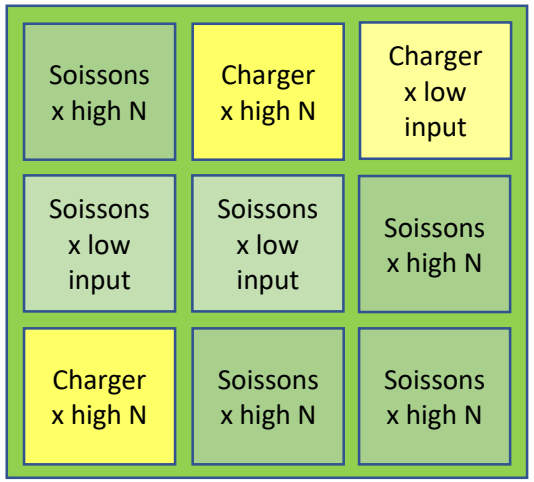
- ***Reproducibility***
- ***Tracibility***

- ***Expensive, lot of resources needed***
- ***Complex design***
- ***Cannot be reproduced***
- ***Huge and very complex datasets***

Environment

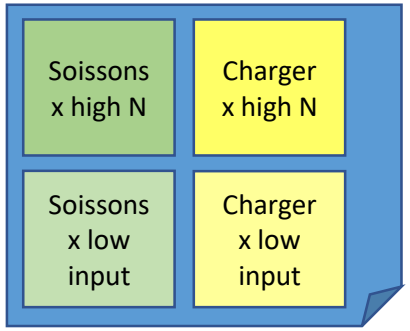
Plant Phenomics simplified data life cycle

« Raw » data, pheno/env measures, variables



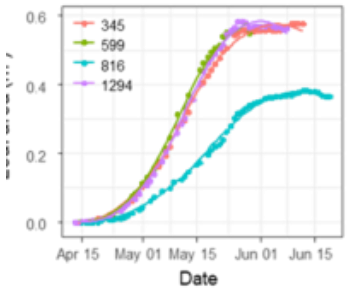
Derivation, Reduction

« computed » data, reduced, indicators



Genotype	Treatment	N input	Date	Rep	Fusariose
Soissons	low input	15,32253129	15/11/2011	1	5
Soissons	low input	15,31430556	16/11/2011	2	7

Genotype	Treatment	Fusariose
Soissons	low input	6



661300270	Ardon	2004	45.645632645603683	12/01/2004	284.3
661300270	Ardon	2005			
661300444	Ardon	2004	38.96112577281653	12/01/2004	228.8
661300444	Ardon	2005			
661300312	Cavallermaggiore	2004	52.4	01/01/2004	249.9
661300312	Cavallermaggiore	2005			
661300371	Cavallermaggiore	2004	45.74	01/01/2004	230.2
661300371	Cavallermaggiore	2005			
661300487	Cavallermaggiore	2004	72.52	01/01/2004	309.8
661300487	Cavallermaggiore	2005			
661300585	Cavallermaggiore	2004	71.739999999999995	01/01/2004	305.7
661300585	Cavallermaggiore	2005			
661300468	Headley	2004	45.27	01/01/2004	
661300468	Headley	2005			
661300469	Headley	2004	70.930000000000007	01/01/2004	
661300469	Headley	2005			
661300533	Headley	2004	57.67	01/01/2004	258.8

Plant Phenomics simplified data life cycle

Raw data long term conservation

Data acquisition

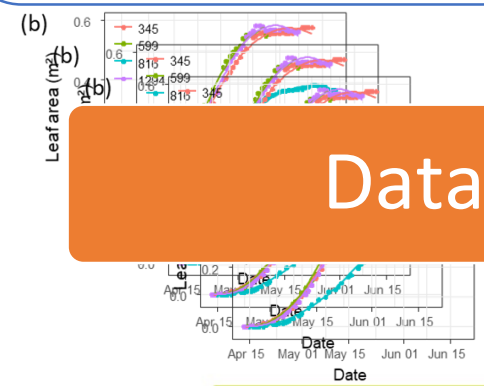
- **VARIABLES**
- Plant/microplot level
- Traceability
- Raw measures
- Data Cleaning
- Platform IS (Emphasis IS, PHIS, …)
- Analysis Reproducibility
- Provenance

Data computation

- **INDICATORS**
- Statistical integration
- Genotype level (mostly)
- New computation for each scientific question
- One raw dataset → many computed datasets

Data publication

- One Data Publication by datasets.
- **Platform IS**
  - Phenomic, plant level
- **FAIR Data Repositories**
  - Reduced



Data

Genotype	traitement	Fusariose
Soisson	low input	5
Soisson		
Charger	low input	1
Charger	high N	2

Knowledge

Variety charger

intensiv cultural practice

Phenomic data management and description → Data standards

Semantic

- Description of the data
- Crop ontology
- www.cropontology.org
- Phenotyping/environment
- variable = Trait + Method + Unit/Scale
- *Biologist* driven

Structure

- Formatting, Modelling and Organizing Data
- Minimal Information About Plant Phenotyping Experiment
- www.miappe.org
- Dataset : greenhouse, field, multiannual, multilocal
- iables and plant material
- t & *Computer scientist* driven

Persistent Unique Identifiers

URI, sensor ID, accessions ID, Trait ID, DOI,…

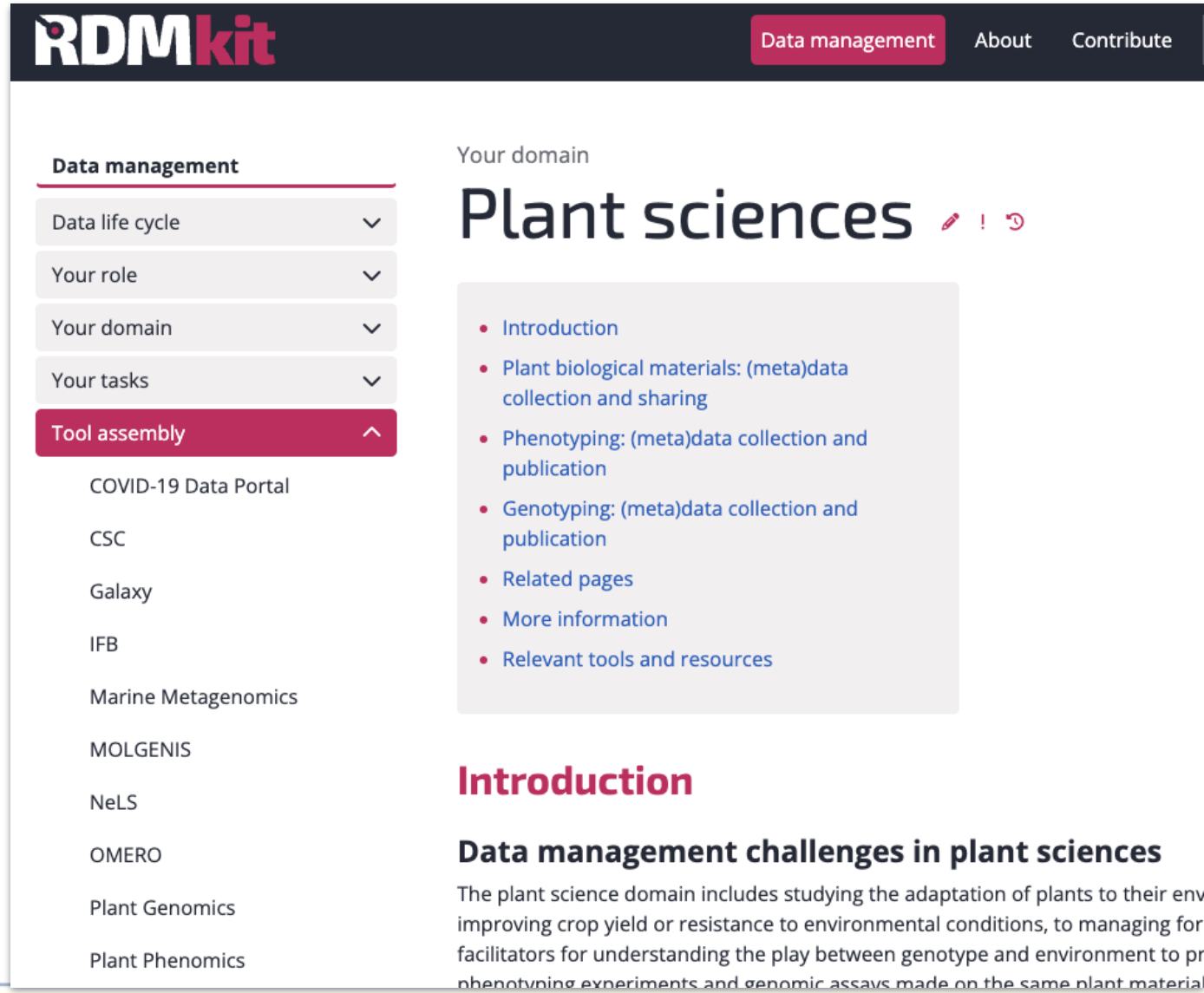
Technical

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

Plant Phenomic data standards and tools

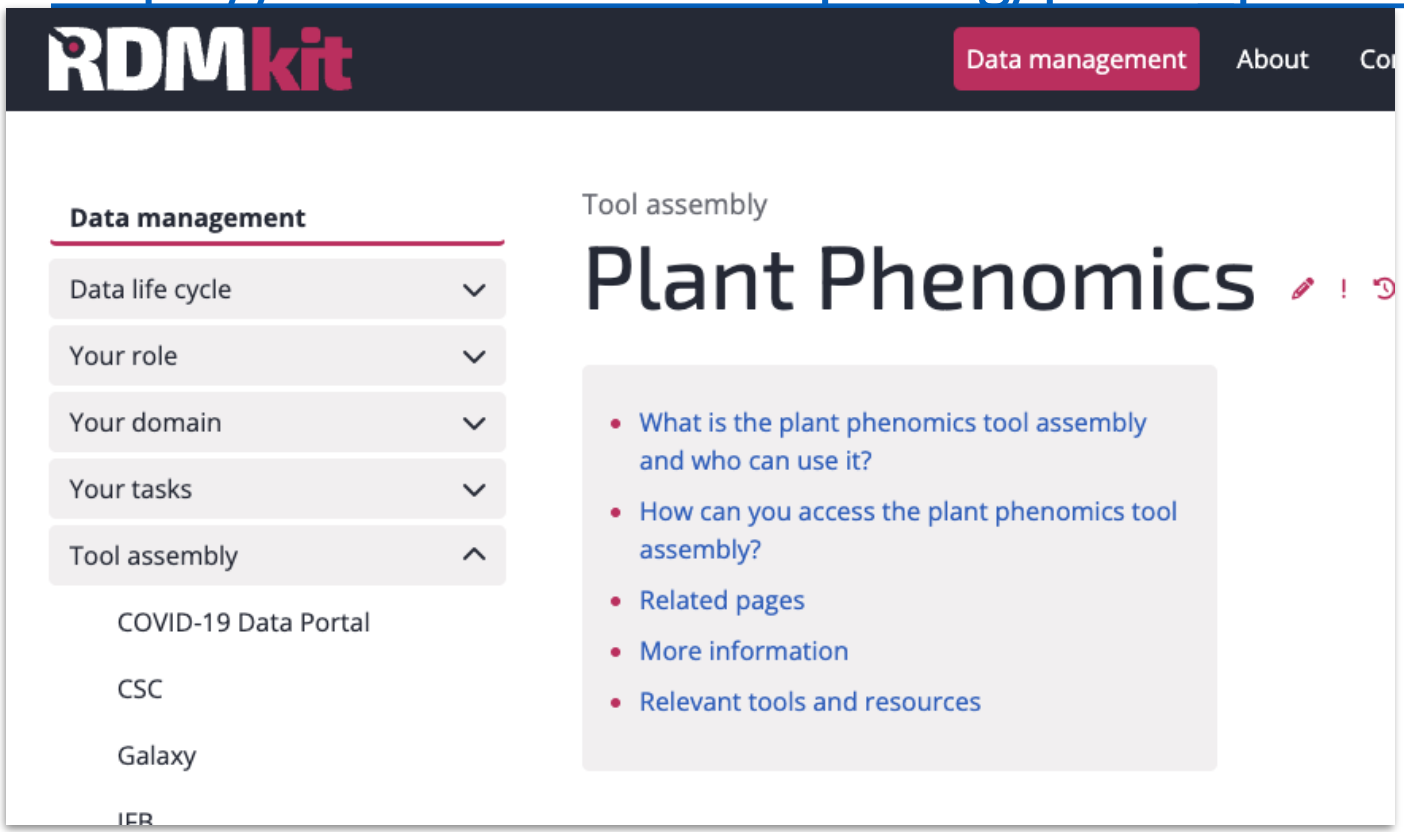
- Implementations
  - File templates
  - Recherche Data Gouv
  - FAIDARE Data portal
  - PHIS/Sixtine Information System

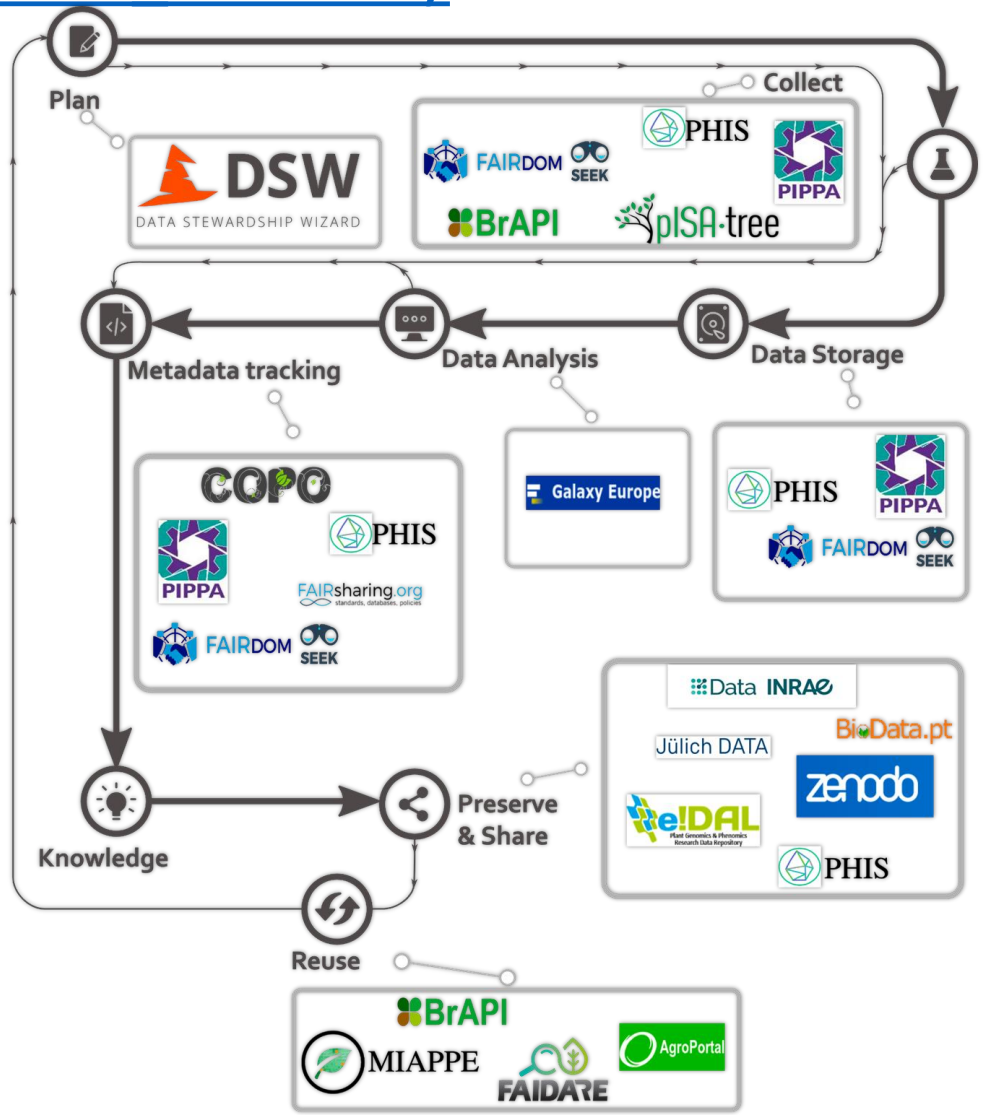
- Research Data Management Kit
  - List recommended standards, tools practices …
  - https://rdmkit.elixir-europe.org/



Research Data Management Kit for Plant Phenomics

https://rdmkit.elixir-europe.org/plant_phenomics_assembly

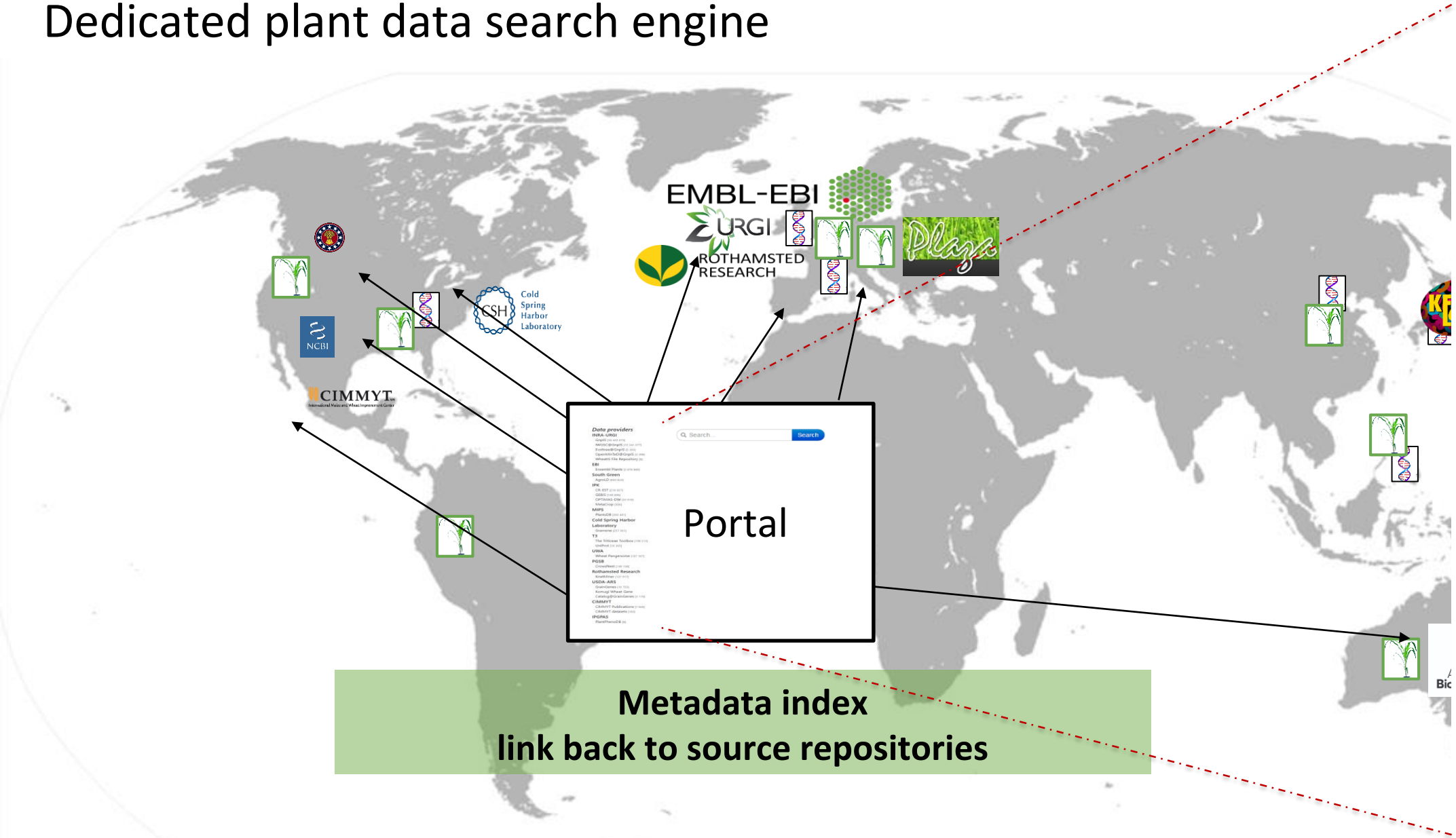




FAIDARE Plant data portal for global heterogenous data federation

PHENOME EMPHASIS FRANCE

Dedicated plant data search engine



Data providers

INRA-URGI
- GnpIS [36 420 075]
- IWGSC@GnpIS [19 241 077]
- Evoltree@GnpIS [5 363]
- OpenMinTeD@GnpIS [3 398]
- WheatIS File Repository [6]

EBI
- Ensembl Plants [2 878 860]

South Green
- AgroLD [450 623]

IPK
- CR-EST [218 927]
- GEBIS [148 696]
- OPTIMAS-DW [33 919]
- MetaCrop [586]

MIPS
- PlantsDB [263 401]

Cold Spring Harbor Laboratory
- Gramene [257 561]

T3
- The Triticeae Toolbox [199 213]
- UniProt [16 365]

UWA
- Wheat Pangenome [187 167]

PGSB
- CrowsNest [140 138]

Rothamsted Research
- KnetMiner [137 917]

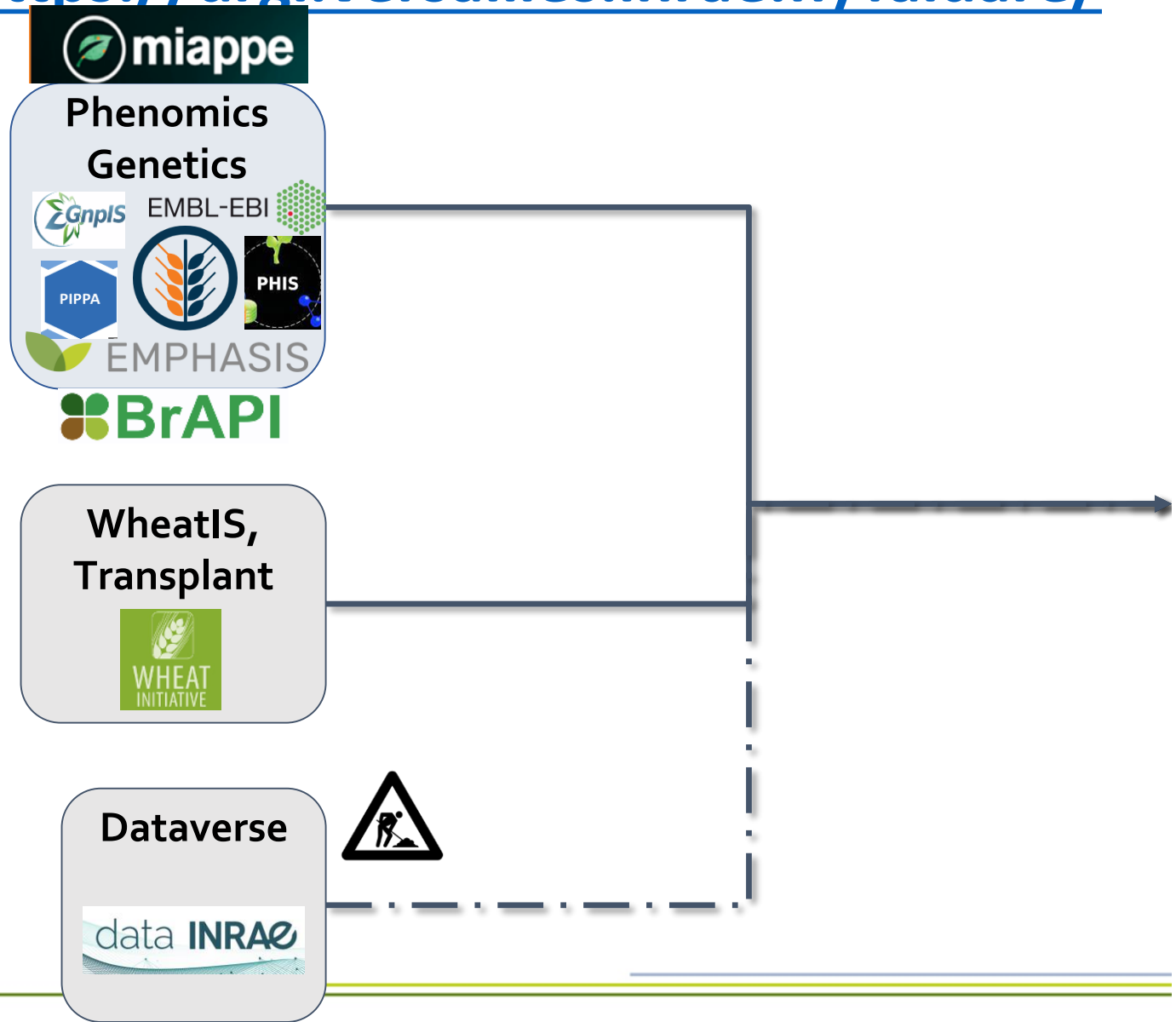
USDA-ARS
- GrainGenes [16 755]
- Komugi Wheat Gene
- Catalog@GrainGenes [3 119]

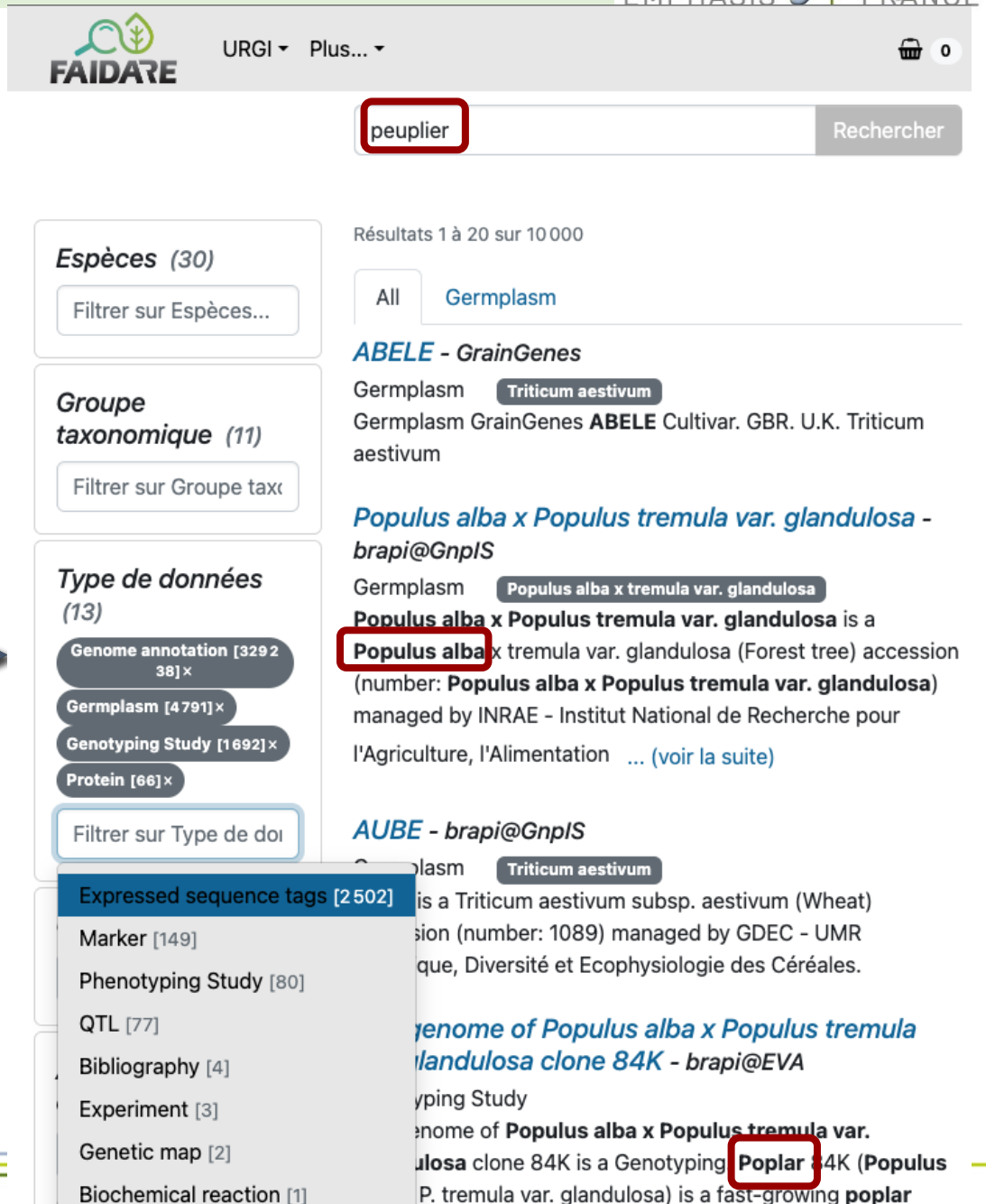
CIMMYT
- CIMMYT Publications [1 605]
- CIMMYT datasets [183]

IPGPAS
- PlantPhenoDB [8]

FAIDARE Plant data portal for global heterogenous data federation

https://urgi.versailles.inrae.fr/faidare/





File template & Recherche Data Gouv

https://github.com/MIAPPE/MIAPPE/tree/master/MIAPPE_Checklist-Data-Model-v1.1/MIAPPE_template

	A	B
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). *This section is mandatory.*
2	**Study***	A study (or experiment) comprises a series of assays (or measurements) of one or more types, undertaken to answer a particular biological question. *This section is mandatory.*
3	**Person***	A human involved in the investigation or specifically any of its studies. *This section is mandatory.*
4	**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.
5	**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. *This section is mandatory.*
	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).

General Definitions | Investigation | Study | Person | Data file | +

File template & Recherche Data Gouv

PHENOME

miappe

10-Info-ObservedVariable.tab
Données tabulaires - 21.8 Ko
Publié 3 nov. 2022
106 téléchargements
17 Variables, 91 Observations UNF
List of phenotypic and environment
(https://www.miappe.org/)
Observed Variable

11-Info-Study.tab
Données tabulaires - 6.4 Ko
Publié 3 nov. 2022
87 téléchargements
11 Variables, 19 Observations UNF
List of studies, including locations, used in the dataset, following the MIAPPE data standard
(https://www.miappe.org/)
Study

12a-Info-EIPO-Environmental_Traits.csv
Comma Separated Values - 2.3 Ko
Publié 3 nov. 2022
85 téléchargements

RÉPUBLIQUE FRANÇAISE recherche.data.gouv.fr

Recherche ▾ À propos Guide d'utilisation Support Français ▾ S'inscrire Se connecte

URGI

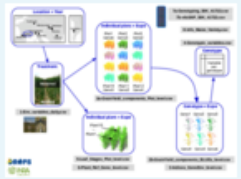
URGI Plant and Fungi Dataverse URGI - Unité de Recherche Génomique Info

(www.urgi.versailles.inrae.fr)

Recherche Data Gouv > Data INRAE > Omics Dataverse > URGI Plant and Fungi Dataverse >

A multi-site experiment in a network of European fields for assessing the maize yield response to environmental scenarios

Version 4.0



Millet, Emilie J.; Pommier, Cyril; Buy, Mélanie; Nagel, Axel; Kruijer, Willem; Welz-Bolduan, Therese; Lopez, Jeremy; Richard, Cécile; Racz, Ferenc; Tanzi, Franco; Spitkot, Tamas; Canè, Maria-Angela; Negro, Sandra S.; Coupel-Ledru, Aude; Nicolas, Stéphane D.; Palaffre, Carine; Bauland, Cyril; Praud, Sébastien; Ranc, Nicolas; Presterl, Thomas; Bedo, Zoltan; Tuberosa, Roberto; Usadel, Björn; Charcosset, Alain; van Eeuwijk, Fred A.; Draye, Xavier; Tardieu, François; Welcker, Claude, 2019, "A multi-site experiment in a network of European fields for assessing the maize yield response to environmental scenarios", https://doi.org/10.15454/IASSTN, Recherche Data Gouv, V4, UNF:6:zS2/ccOQxFrKlUt+1S0Cvg== [fileUNF]

Citer le jeu de données ▾

Pour en apprendre davantage sur le sujet, consulter le document Data Citation Standards [en].

Modalités d'accès au jeu de données ▾

Contact | Partager

Statistiques d'utilisation sur les jeux de données

12 624 consultations

5 857 téléchargements

3 citations

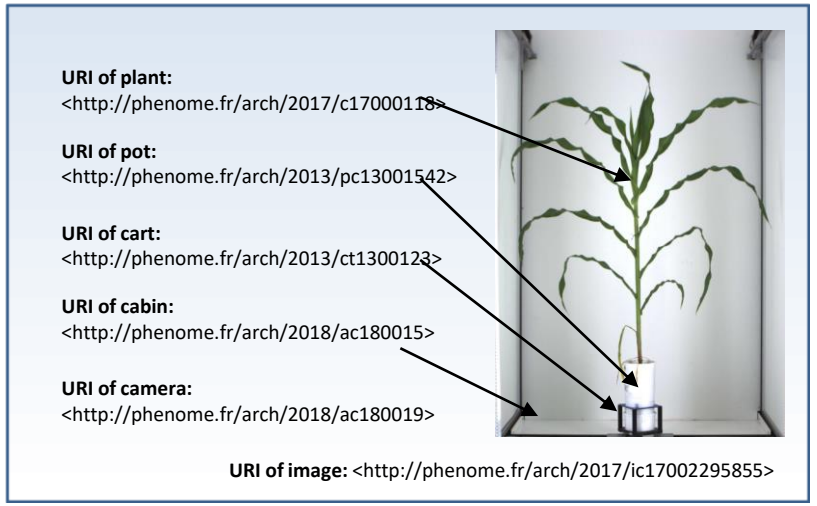
Full Phenomic RDM capabilities

- Dedicated experimental Information System
- PHIS / SIXTINE

Make FAIR data, structure your data

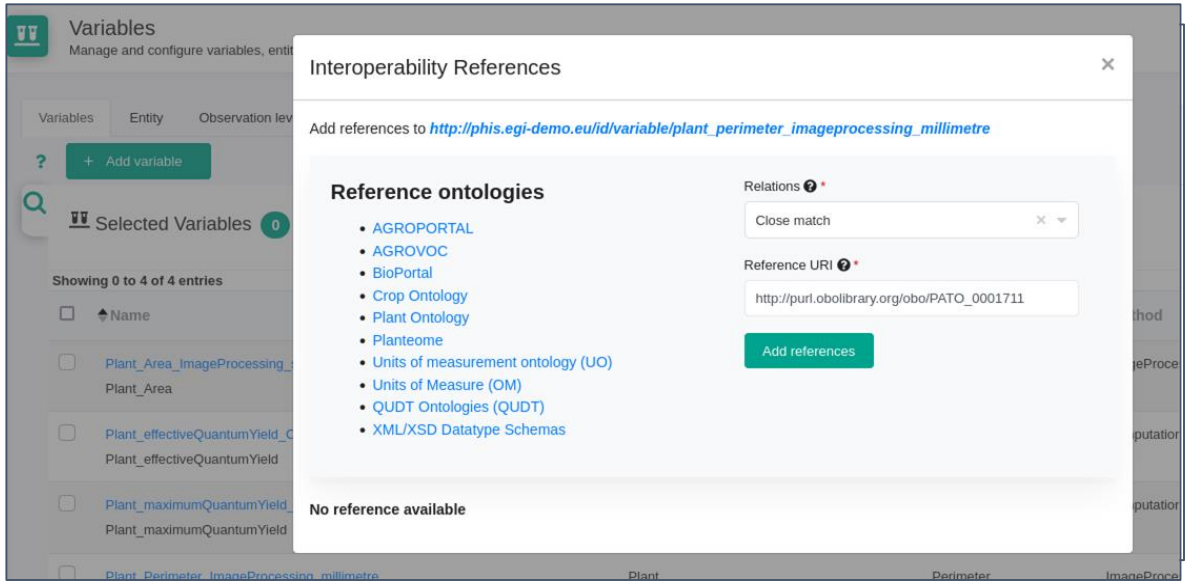
Identification

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



Semantics

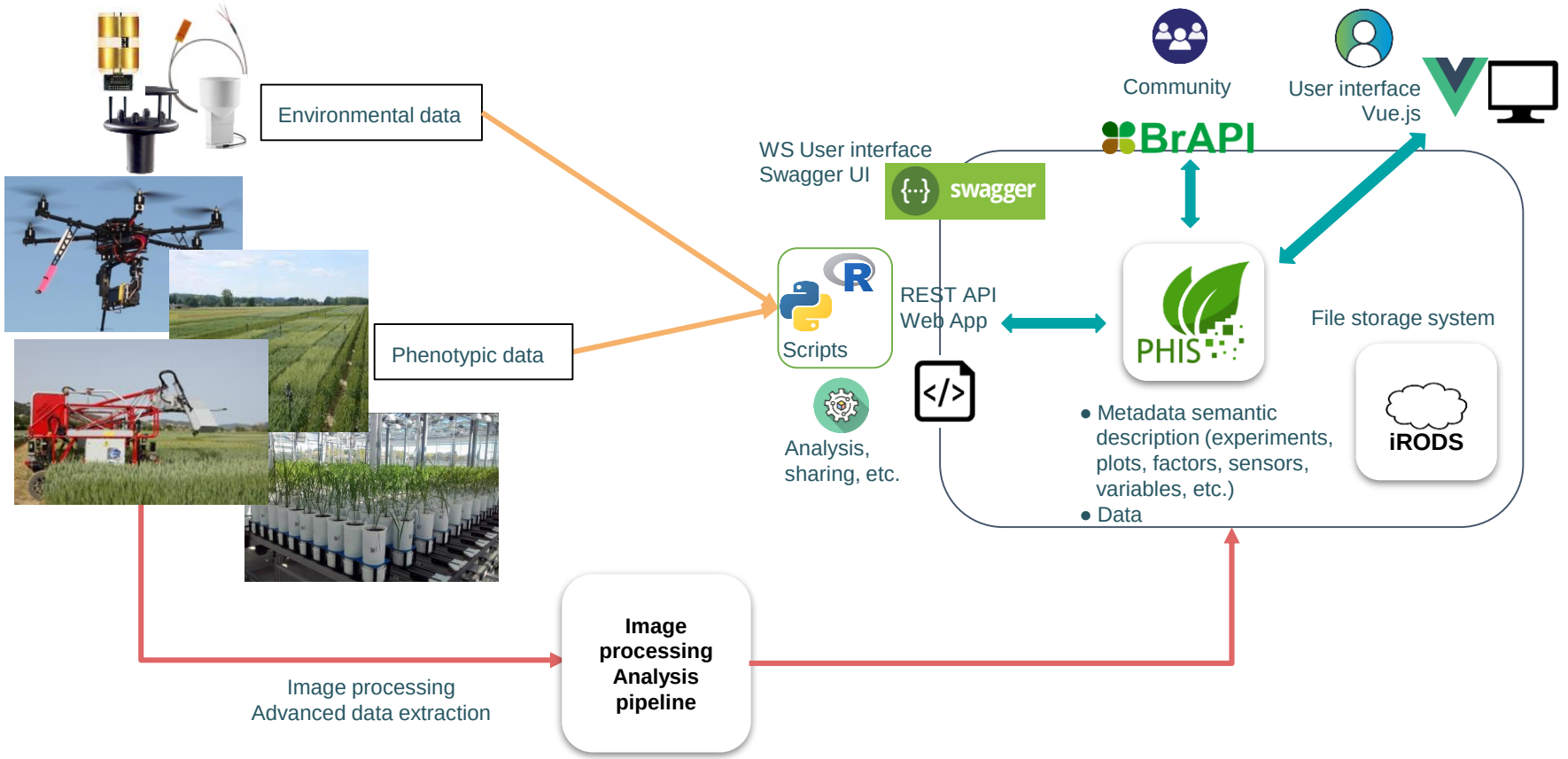
- Naming Conventions
- Controlled vocabulary
- Formalized relationships between entities
- Data annotation and enrichment



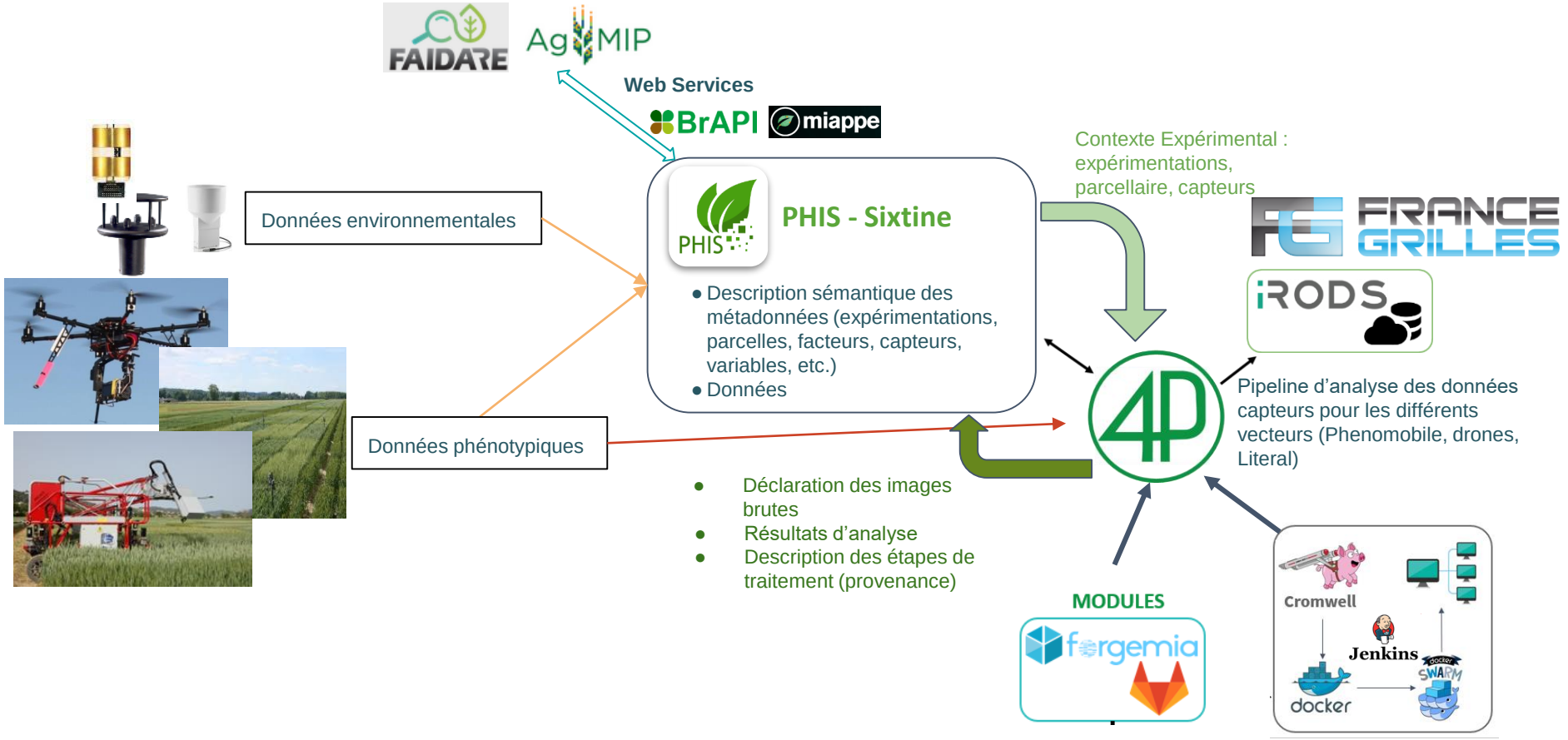
How?

PHIS, an **ontology driven** Information System

PHIS: an Ontology driven Information System for Plant Phenomics



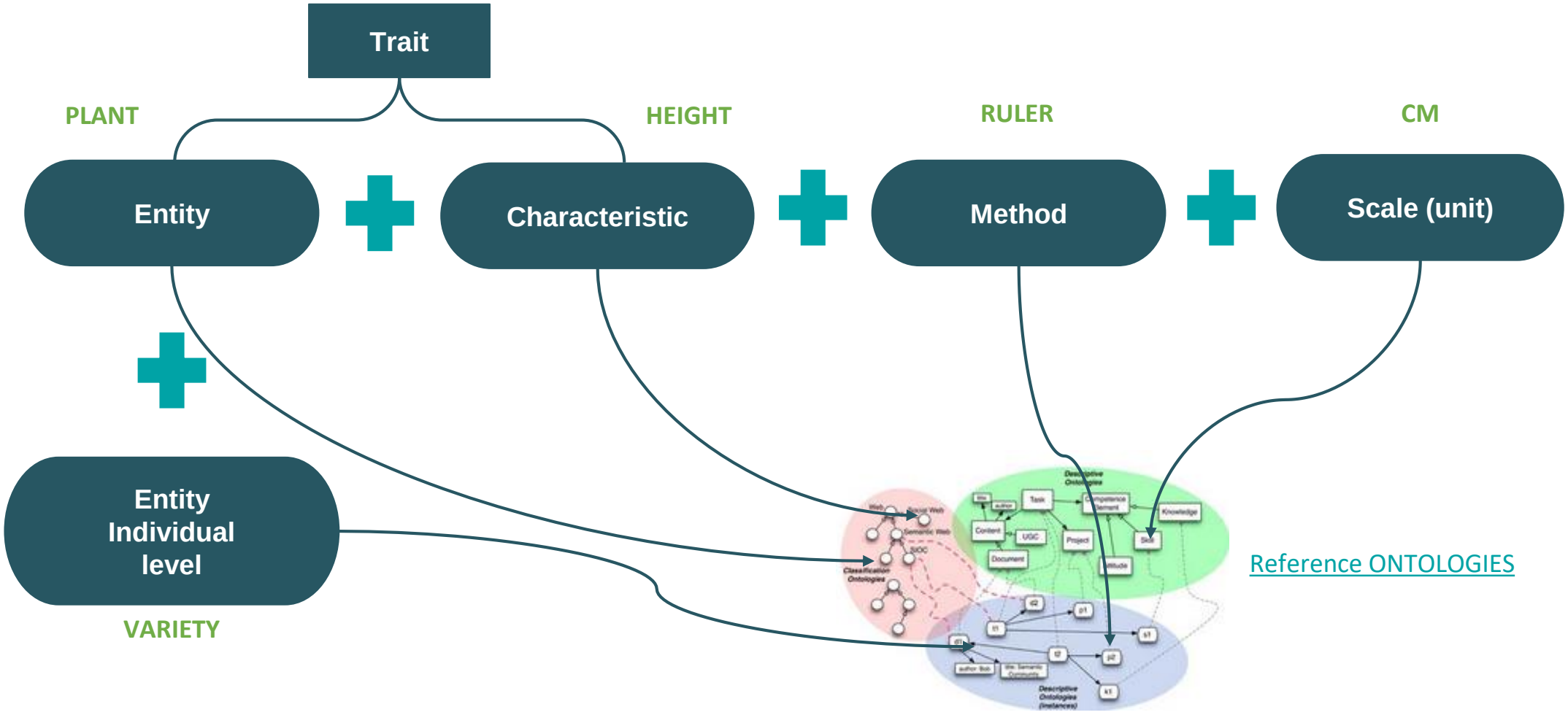
PHIS-4P: example of a connection to a data processing platform



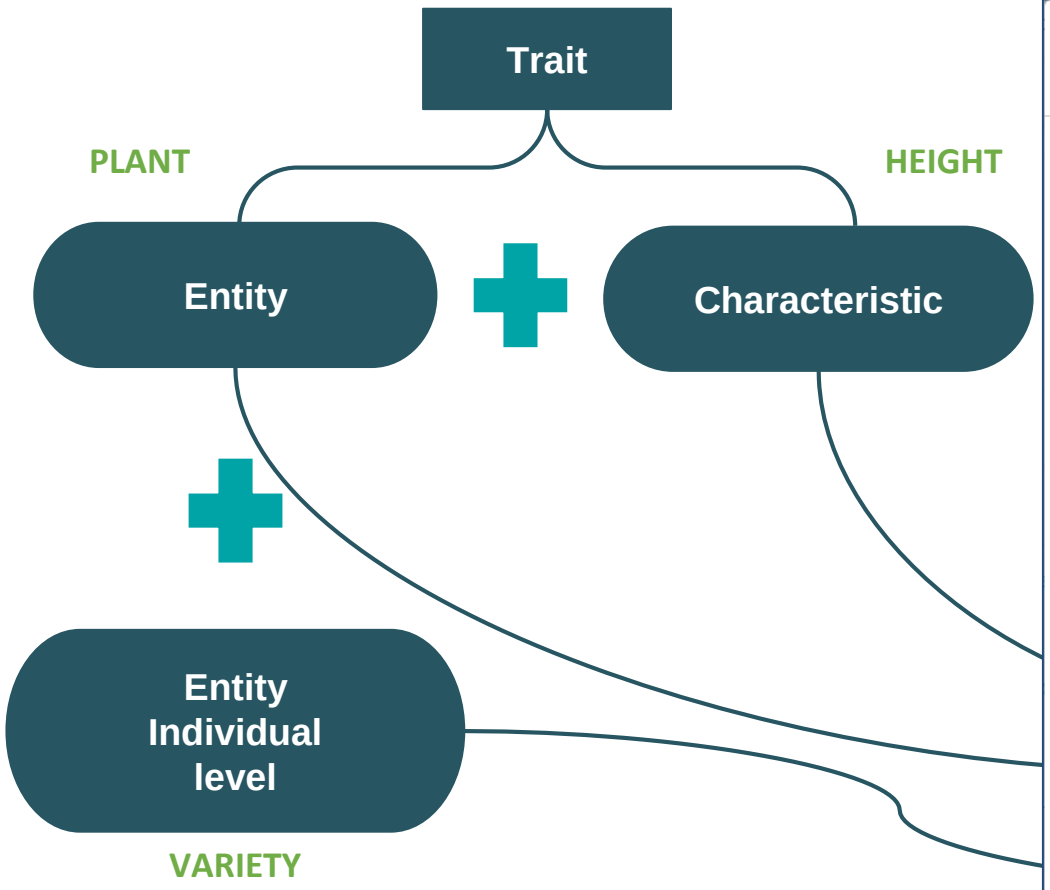
Resource Sharing

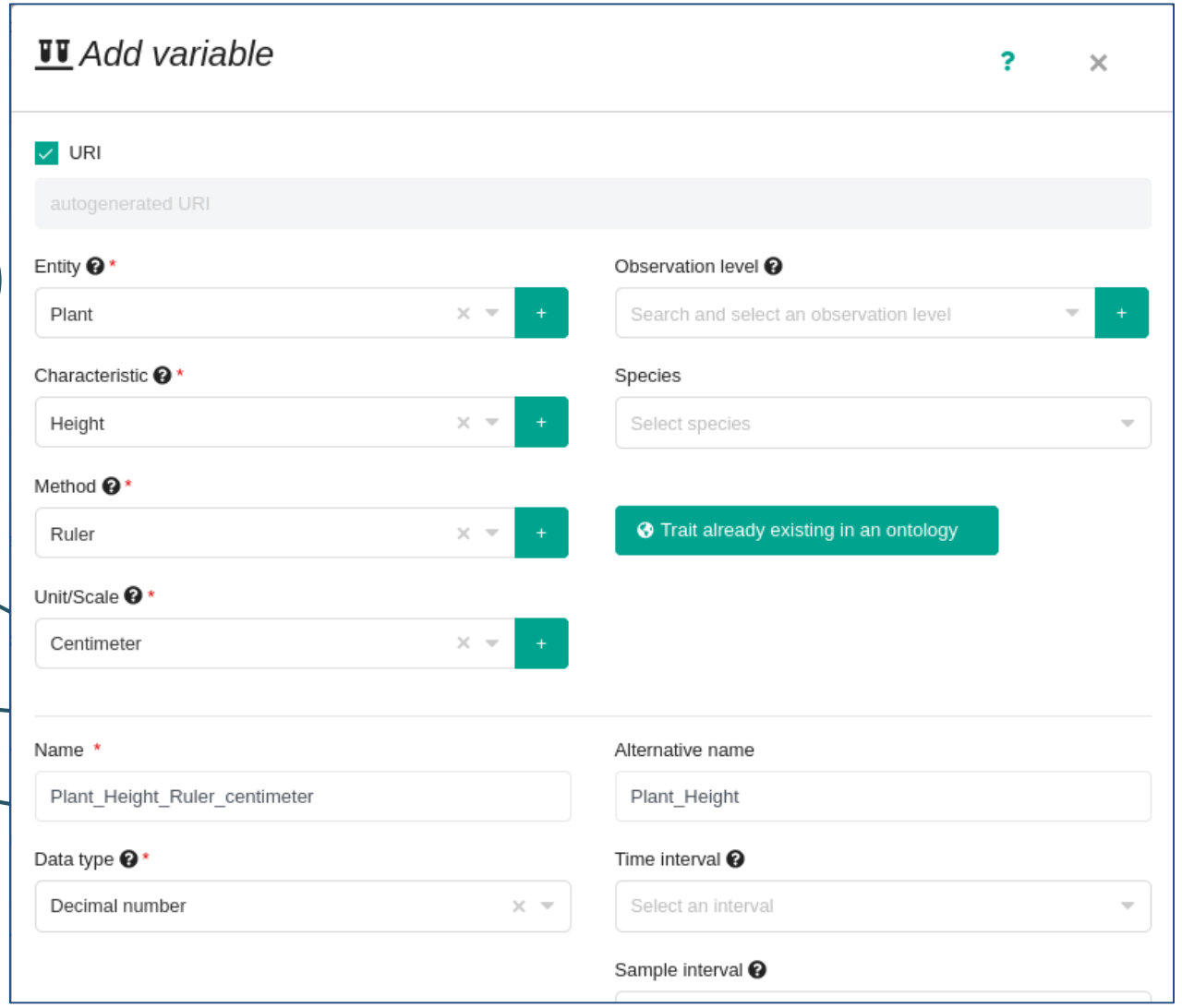
- **Naming conventions**
- Link to **domain ontologies** (Crop Ontology, Agrovoc, etc.) and **standard ontologies** (Dublin Core, Ontology of Annotation, etc.)
- Specific instances for creating, storing, and exporting **sharable resources**
- **Data traceability** protocols
- **Standard Compatibility** (BrAPI, MIAPPE)

Resource Sharing - Variables model

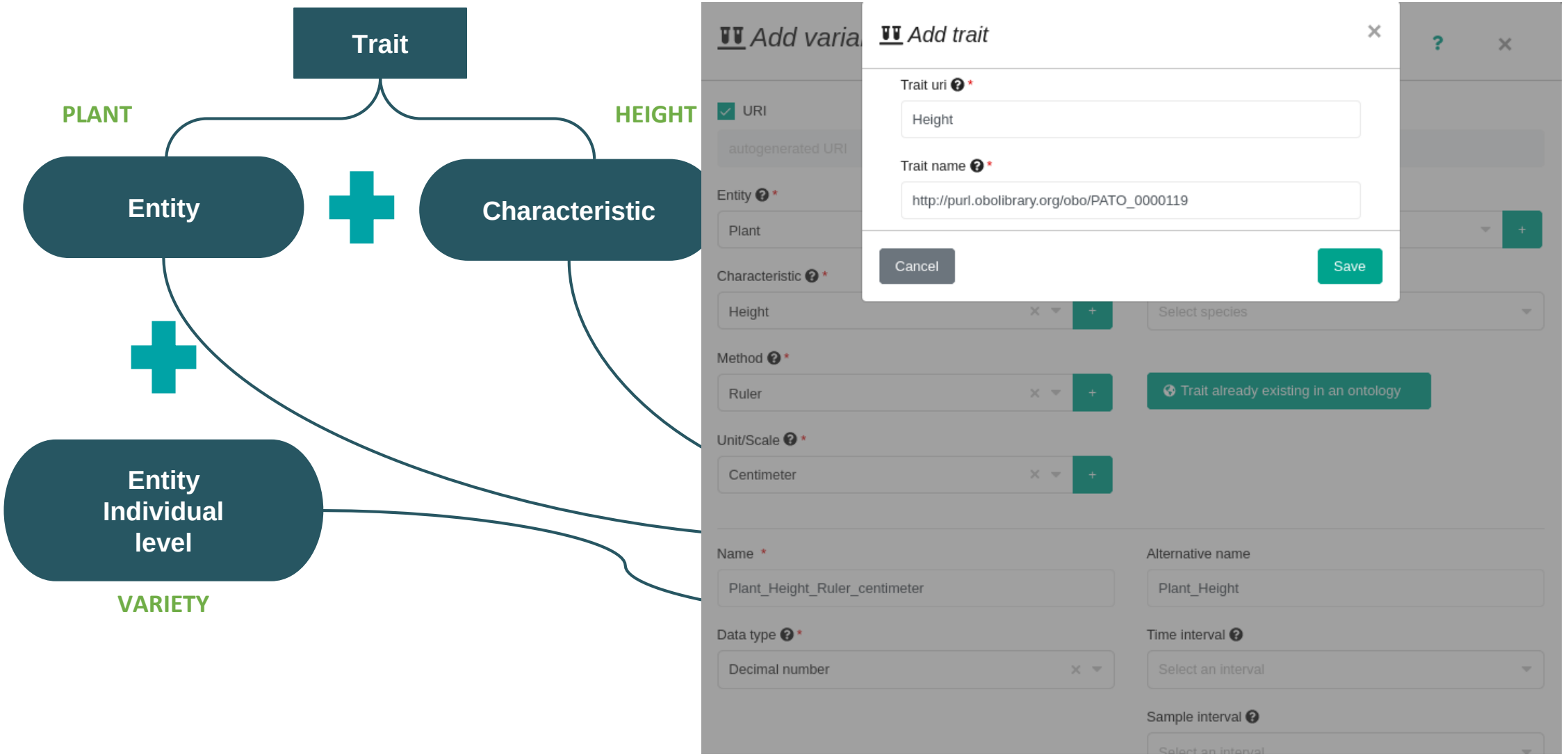


Resource Sharing - Variables model



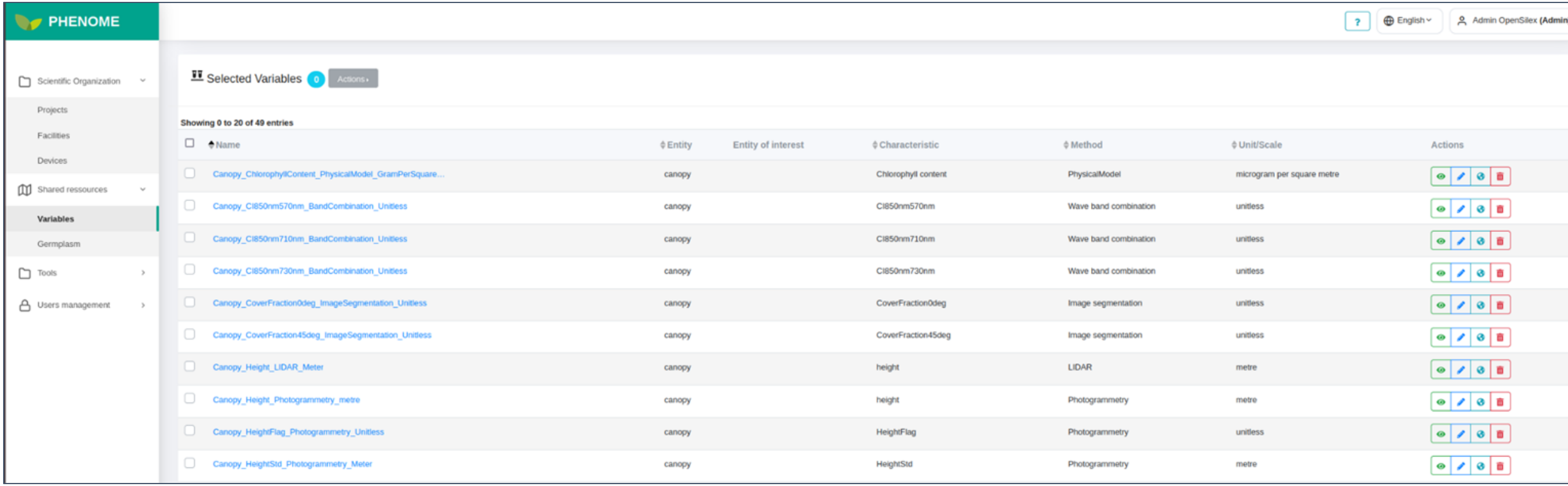


Resource Sharing - Variables model



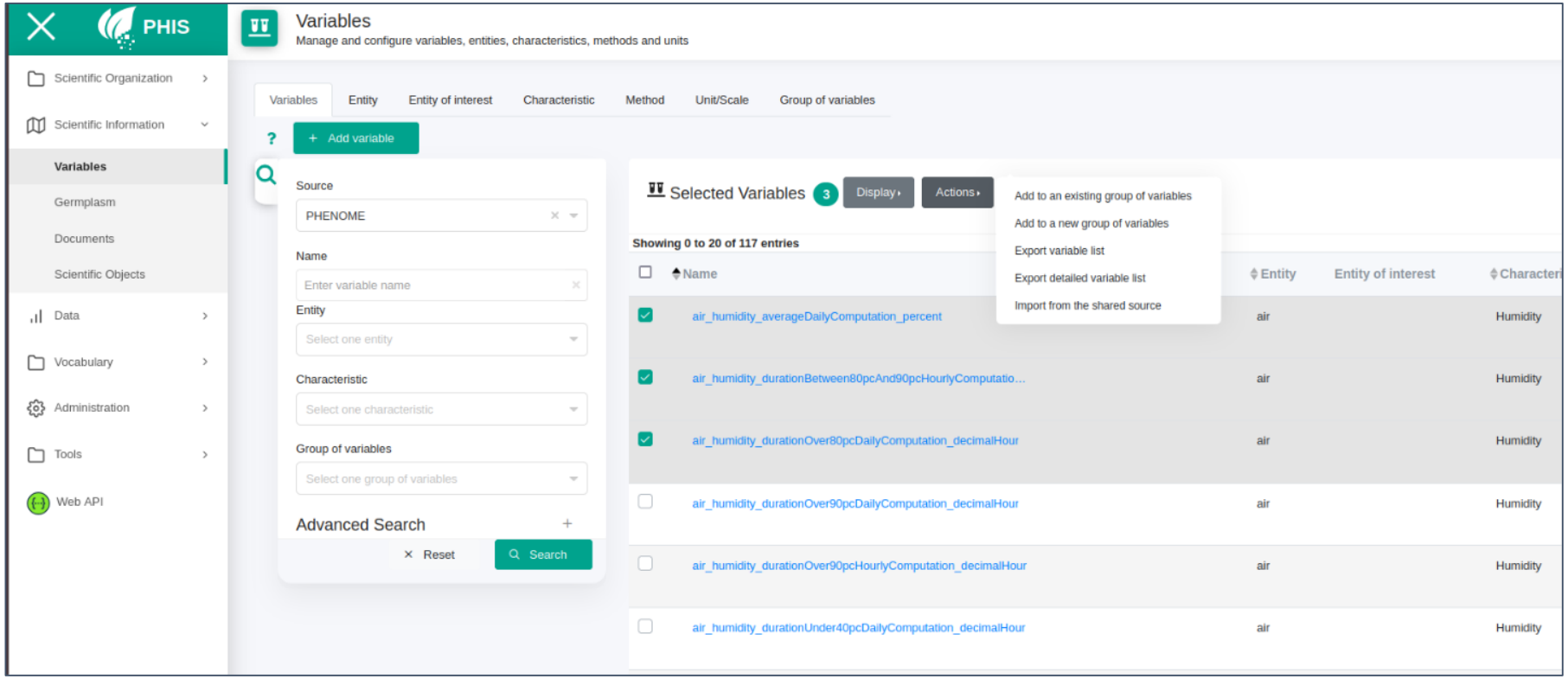
Resource Sharing - Shared resources instances

- Variables



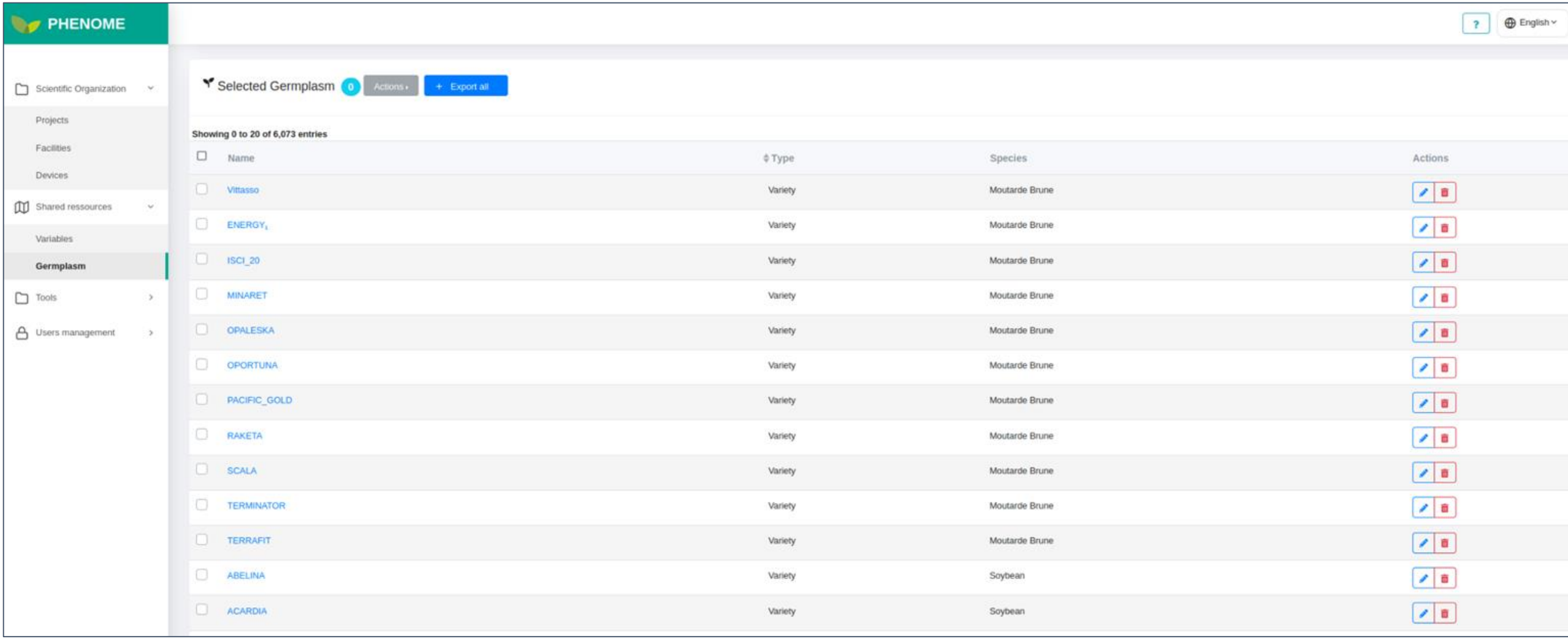
Resource Sharing - Shared resources instances

- Variables



Resource sharing - Shared resources instances

- Germplasm



Resource Sharing - Data traceability - Provenance

What is Provenance ?

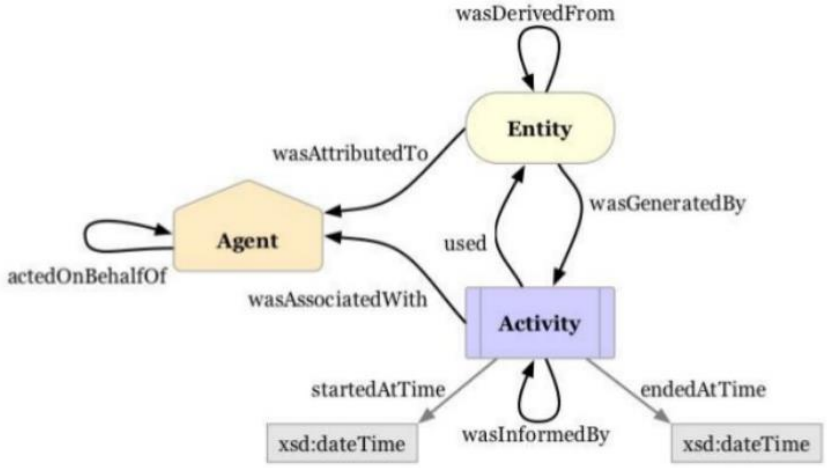
That is metadata to know

- Who played a role in data generation at different scale?
- Who is the main responsible?
- How data was transformed?
- Which tools were used?
- When and how data were produced?
- etc.

Resource Sharing - Data traceability - Provenance

- **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

Representation model for provenance (PROV-O)



Resource Sharing - Data traceability - Provenance

Exemple : PhenoArch

Greenhouse experiment

Plant

camera ?
settings ?
angle ?

RGB Image

Provenance
Image Acquisition

treatment ?
settings ?

Binary Image

Provenance
Binarisation

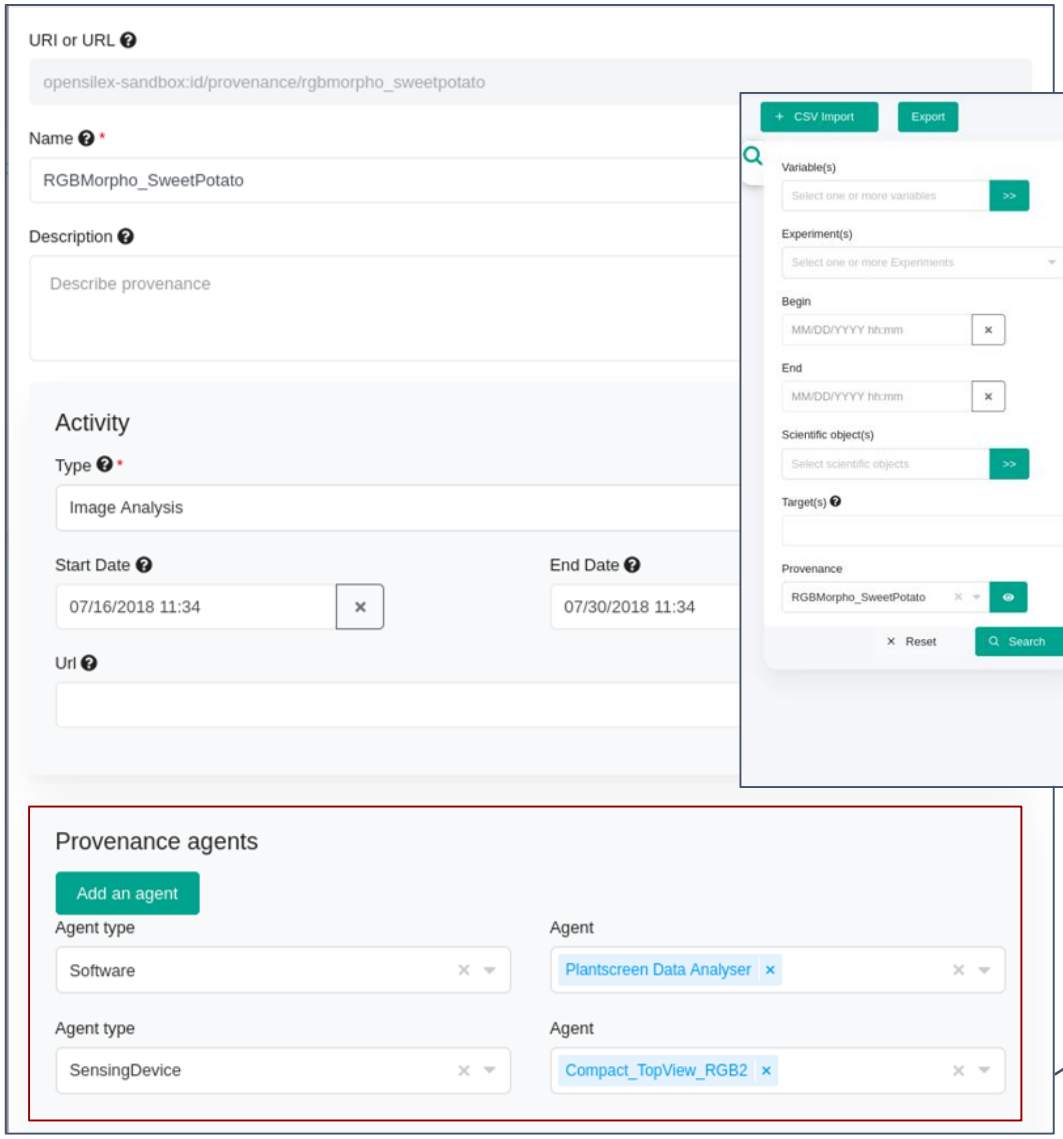
traitement ?
settings ?

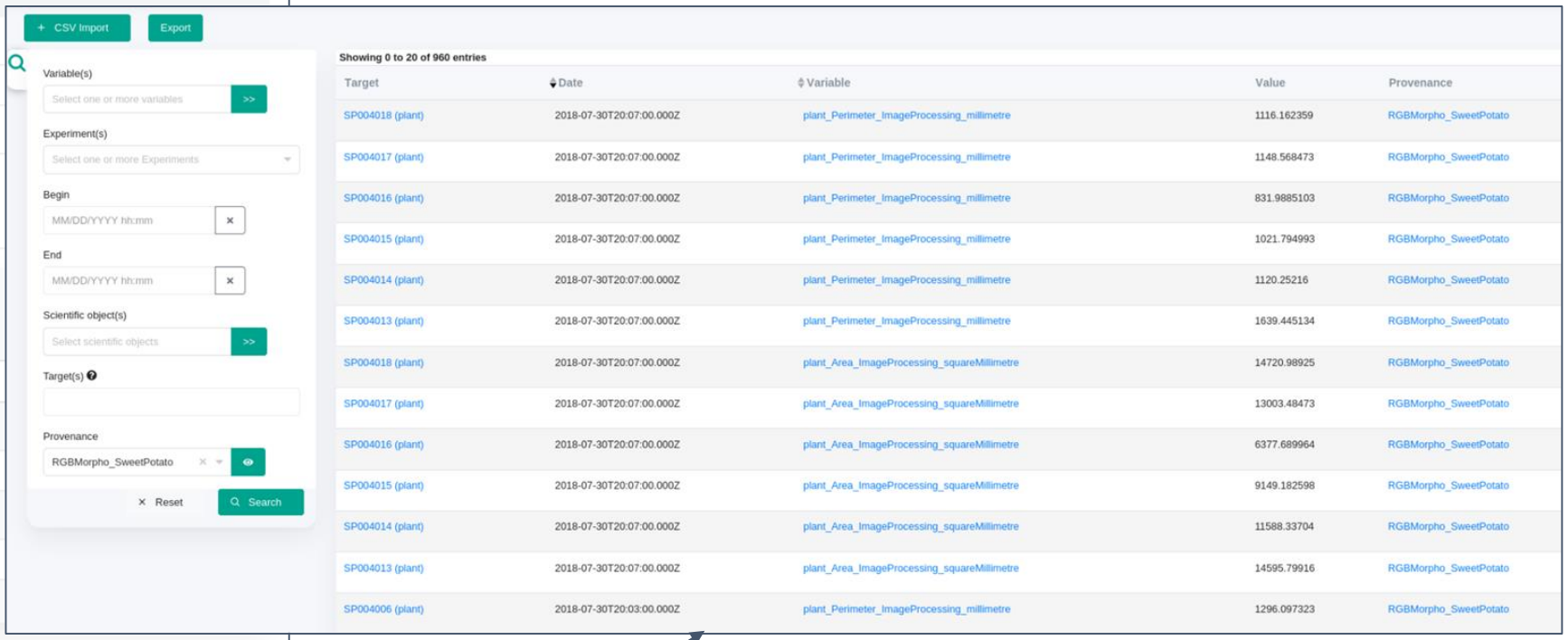
hauteur de plante = 342 pixels

Elaborated Data

Provenance Data
(traits extraction)

Resource Sharing - Data traceability - Provenance



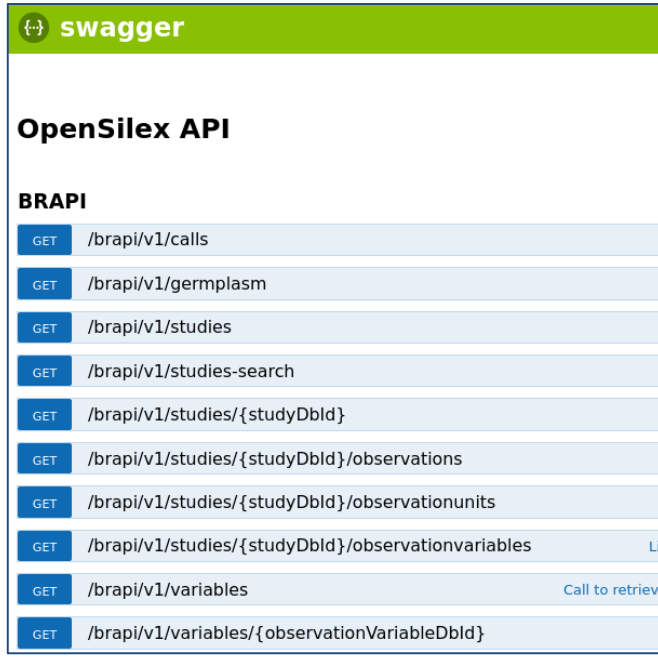


Data are linked to a provenance

Resource Sharing - Standards Compatibility

- **BrAPI** - https://brapi.org/
  - Standardized RESTful web service API specification for communicating plant breeding data
  - **PHIS:** BrAPI WebServices implemented (v1.3)

=> Makes **indexing of PHIS data by FAIDARE** possible!



- **MIAPPE -** https://www.miappe.org/
  - Minimum Information About Plant Phenotyping Experiments
  - Open, community driven, data standard designed to harmonize data from plant phenotyping experiments
  - Mapping with other standards (ISA-Tools , BrAPI)
  - **PHIS is MIAPPE-compliant**

User Friendly Interfaces

Declaration of the experimental context

SweetPotatoViruses_2018

Experiment

Details | Factors 1 | Scientific objects | Data 42K+ | Visualization | Map | Annotations | Documents 1

Description

Name SweetPotatoViruses_2018

State Finished | Public

Period 2018-07-01 - 2018-07-30 (29 days)

URI opensilex-sandbox:id/experiment/sweetpotatoviruses_2018

Objective

Phenotyping viral infection in sweetpotato using a high-throughput chlorophyll fluorescence and thermal imaging platform

Description

Virus diseases caused by co-infection with Sweet potato feathery mottle virus (SPFMV) and Sweetpotato chlorotic stunt virus (SPCSV) are a severe problem in the production of sweetpotato (Ipomoea batatas L.). Traditional molecular virus detection methods include nucleic acid-based and serological tests. In this study, we aimed to validate the use of a non-destructive imaging-based plant phenotype platform to study plant-virus synergism in sweetpotato by comparing four virus treatments with two healthy controls.

Context

Projects NextGen Phytosanitation

Organizations National Plant Phenotyping Infrastructure , University of Helsinki

Facilities PlantScreen™ Compact System

Species sweetpotato

Factors Viral Disease

Groups NaPPI Team

Contacts

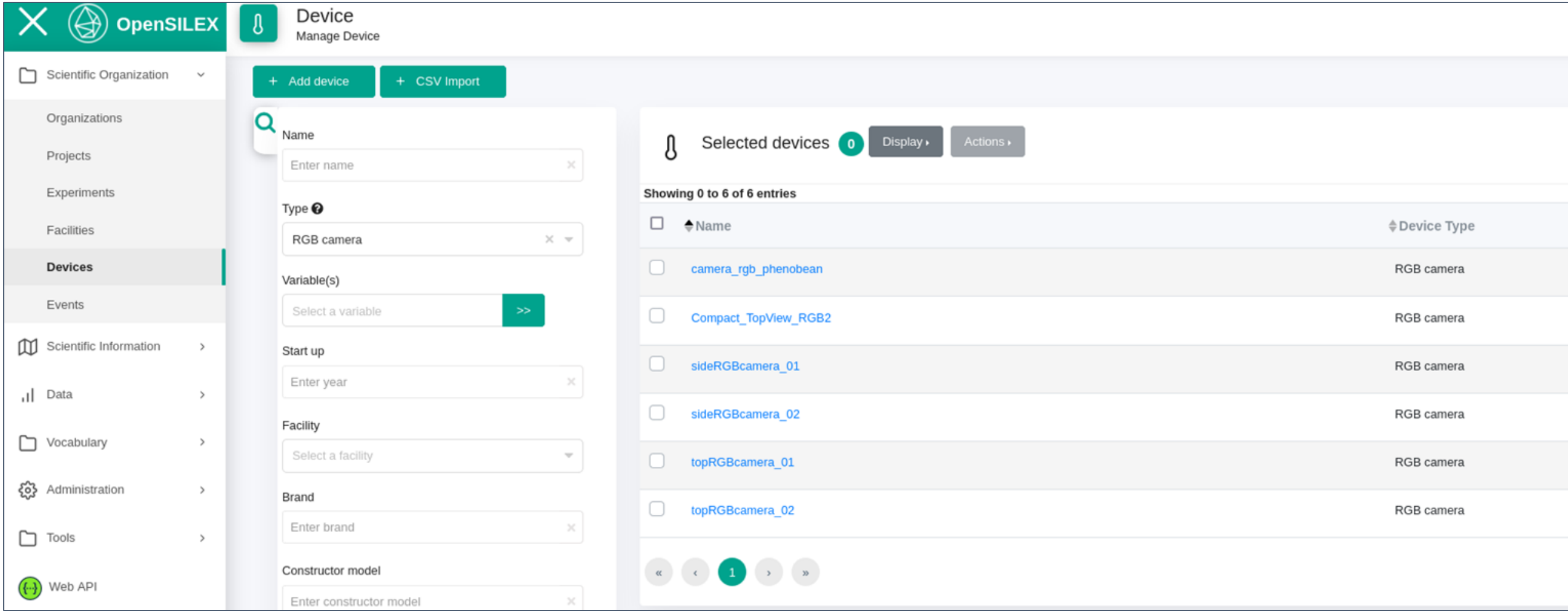
Scientific supervisors Jari Valkonen , Linping Wang , Sylvain Poque

Technical supervisors Sylvain Poque

Declared by admin admin

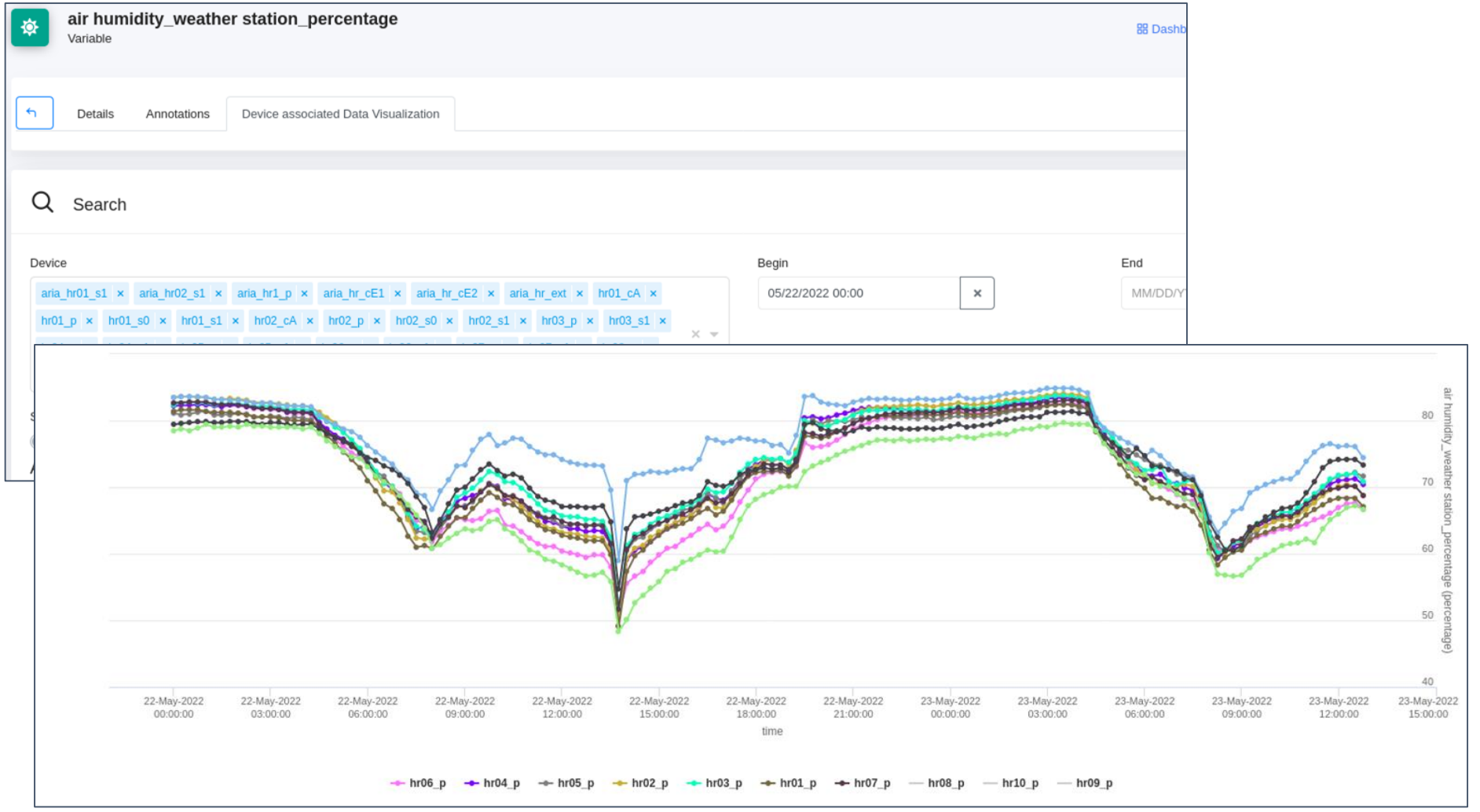
User Friendly Interfaces

Declaration of the Devices



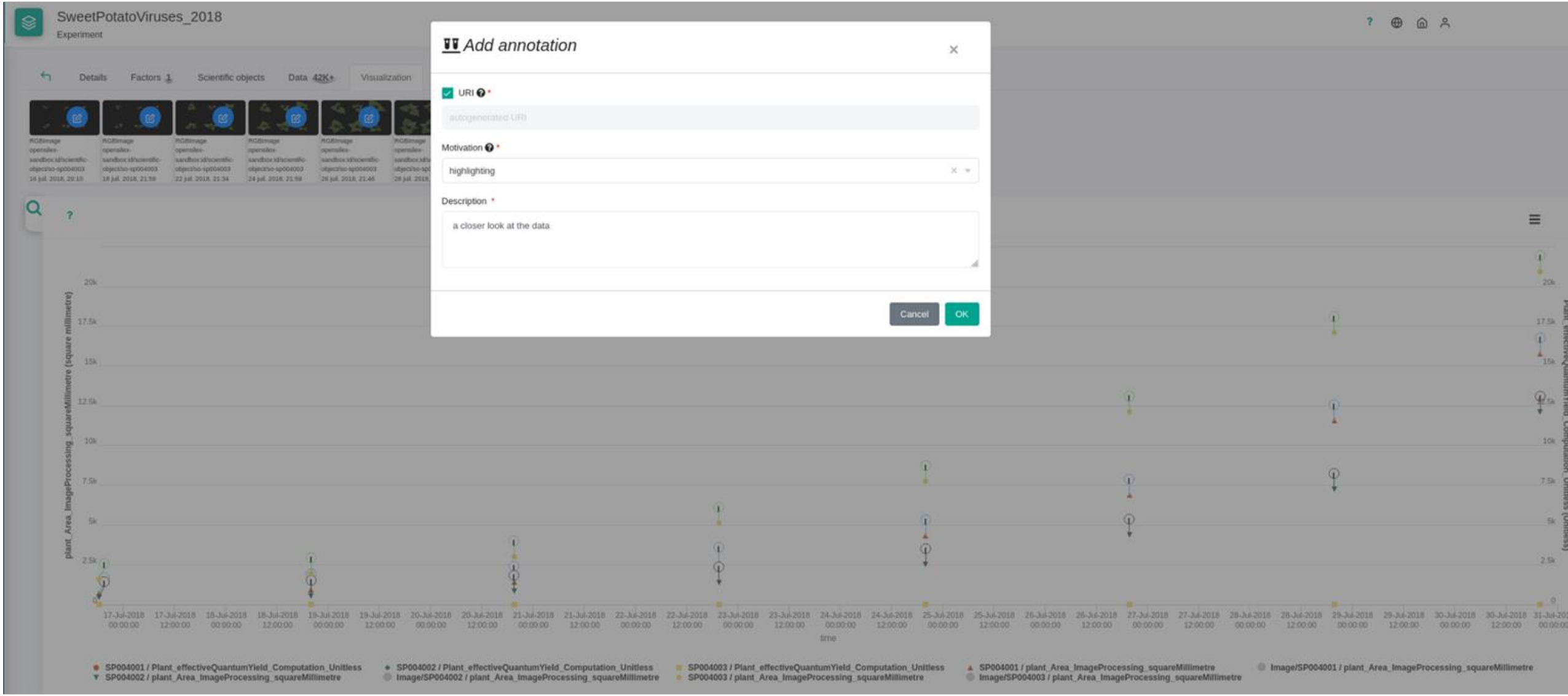
User Friendly Interfaces

Data Visualization - Environmental data



User Friendly Interfaces

Data Visualization to raw image provenance



PHIS Users Community

Phenome Platforms

- M3P - PhenoArch
- Pheno3c
- Phenotoul
- Diaphen
- New users:
  - Phenotic
  - 4PMI

Emphasis community

- UCL
- University of Helsinki
- University of Copenhagen
- WUR NPEC

Conclusions

- A **new generation of information systems** (e.g. PHIS) is needed
    - Giving value to complex data requires **structuring according to FAIR principles**
    - A better formalization of concepts (using ontologies) and data is required for **interdisciplinary research**
      Advanced data management makes **data available for AI and data analytics**
- **PHIS**: used in several phenotyping platforms
- Work in conjunction with **standards** (BrAPI - MIAPPE) and offer shared resources (variables/genetic)
- Successfully linked to an image analyses platform using FAIR principles for its input and output data
- **Ongoing developments**
    - New features / enhanced ergonomy
    - Adaptation to the needs / evolution of research
    - Support
    - Link with research teams
- **OpenSILEX**: opening to **other communities**
- Support and training available!

PHIS - OpenSILEX useful links

- OpenSILEX
  - OpenSILEX website: http://opensilex.org/
  - OpenSILEX demo: http://opensilex.org/sandbox/app/
  - How to contribute to OpenSILEX?
    - Github repository: https://github.com/OpenSILEX/
    - Developer documentation: https://opensilex.github.io/docs-community-dev/
    - OpenSILEX Docker: https://github.com/OpenSILEX/opensilex-docker-compose
    - User documentation: https://opensilex.github.io/phis-docs-community/

- PHIS
  - PHIS website: http://phis.inrae.fr/
  - Research paper: https://nph.onlinelibrary.wiley.com/doi/full/10.1111/nph.15385
  - Variables declaration tutorial: https://www.youtube.com/watch?v=Pvz9o-b_Mok