

INRAE

> OpenSILEX

**Pour des Systèmes d'Information
pilotés par des ontologies**

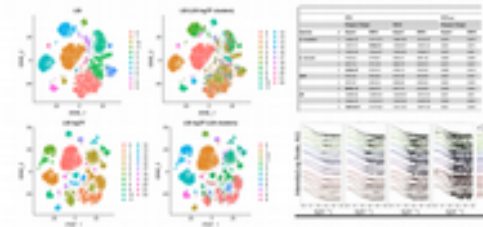
Contact : Pascal.Neveu@inrae.fr

 Département
MathNum

 **Mistea**
Mathématiques, Informatique et Statistique
pour l'Environnement et l'Agronomie



- **Experimentations in agriculture and environment domains**
 - Expensive, require a lot of resources and often very hard
 - Cannot be reproduced
 - **Very complex datasets**



★ Strong needs of **transparency** and reproducibility

- **Give value to data:** re-analyses, meta-analyses and new analyses
→ impossible without advanced data management



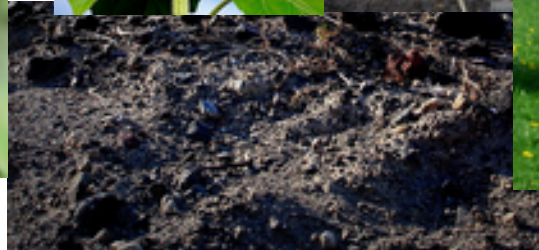
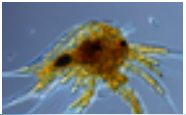


Different environments and ecosystems

Different stages

Different scales

Different interactions





Instrumentation that evolves and used at various frequencies

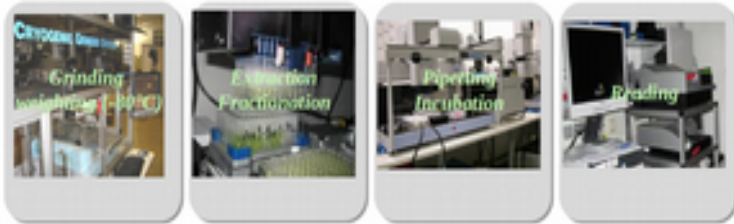
In heterogeneous facilities

By Different teams

« omics » Platforms

Various data complex types

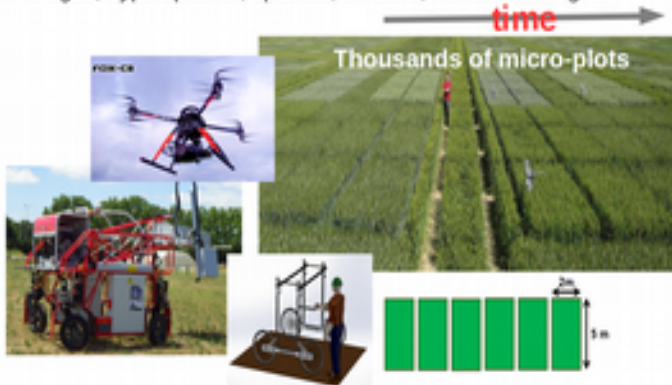
- Genomics
- Composition and the structure of biopolymers
- Quantification of metabolites and enzyme activities



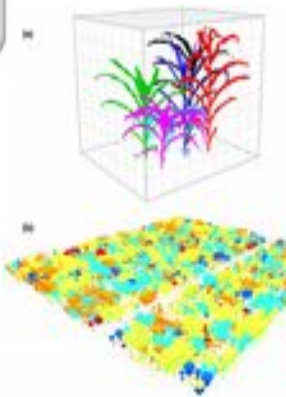
Field Platforms

Various scales and data types

- Cell, organ, plant, population
- Images, hyperspectral, spectral, sensors, human readings...

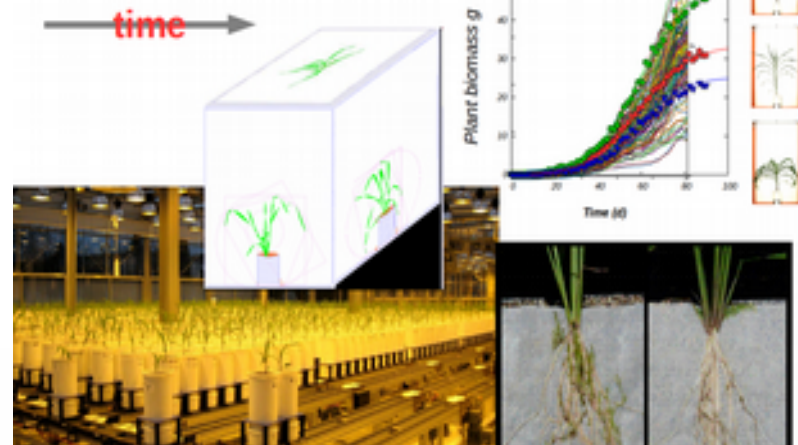


Virtual experiments



Green house Platforms

Various scales and data types



Farm Platforms

Various scales and data types from thousands of farms

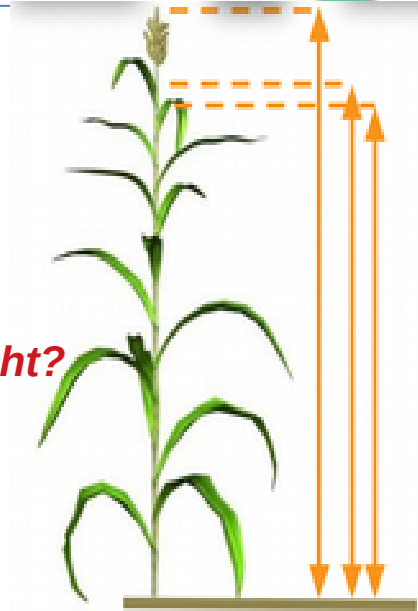
- organ, plant, population, site
- Images, sensors, human readings...





- **Manage metadata in file names**
(not standardized, very often not machine readable, poor metadata quality, etc.)
- **Variable naming**
e.g. same name for several variables, no variable ID
- **A part of data are stored on personal computer**
- **Unstable files** (machine incompatible organisations)
- **Ambiguous ID**
- **Context, faults are not described**
- **No data links**
- **No missing data representation**
- **No licence**
- **Etc.**

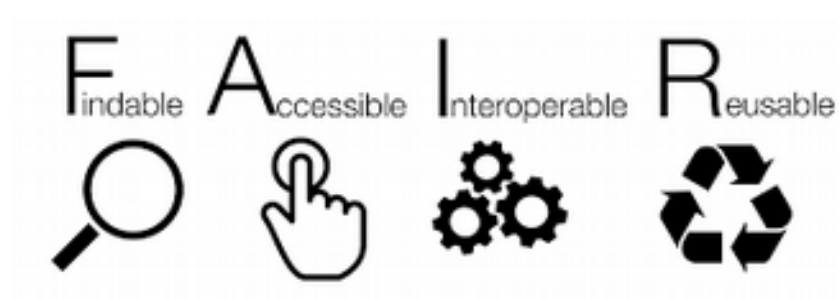
plant_height?



Plot566
in 2016

Plot566
in 2017





Findable: **persistent ID**, indexed in portals, standardized and relevant metadata
coordinated and sustainable data services

Accessible: open and standardized protocols, **license rights**, **cultural evolution**

Interoperable (technology, syntax, semantics): shared standardized formats, vocabularies and methods **for knowledge representation**

Reusable: **provenance**, relevant metadata for understanding **across disciplines**, and *robust analysis methods*



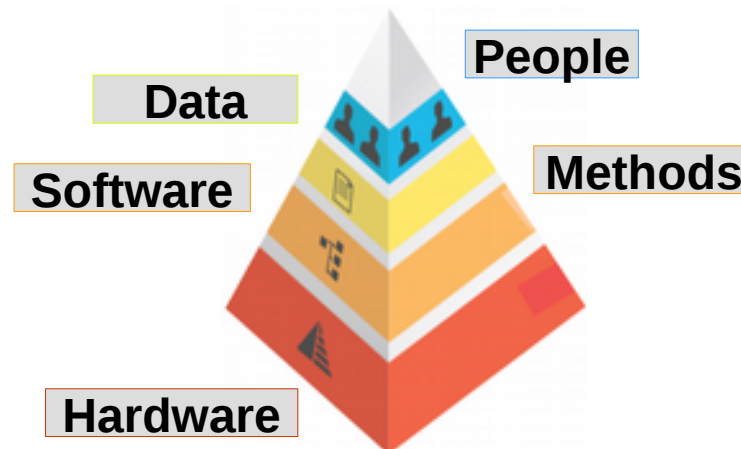
A need for a new generation of Information Systems





OpenSILEX: an open source software set

- Methods, tools, components to implement information systems for **experimental data in agriculture and environment**





Implements and enables good practices:

- Making data FAIR
- Taking into account flexibility needs
- Improving understanding and reproducibility of data processing
- Making easier DMP and Open Science

Structuring of data based on 2 key elements:

- **Identification** of entities (plants, plots, devices, experiments, events, etc.)
 - Persistent, unambiguous, resolvable
- **Semantics** (based on ontology set) provide:
 - Schemas for data
 - Controlled and standardized vocabulary
 - Representation models
 - Formalized relationships between entities
 - Data annotation and enrichment (search engine friendly)



OpenSILEX → Ontology driven Information System

Scientific objects (plant, plant organ, plot, etc.) are formalized (**OWL**)
Identified by **URI** standardized, unambiguous, shared, etc

Events (management, faults, meteo, etc) are formalized (**OWL**)
Identified by **URI**

Variables, Observations, Factors, Documents, Devices, Softwares
are formalized and associated with these Objects and Events (**OWL**)
Identified by **URI**

Organisation and linking of Objects and Events → done with a controlled
semantic (reference ontologies, vocabularies, thesaurus, taxonomies) and
application Ontologies (**RDF, OWL, SKOS**)*

Identification



URI

- **Standardized** and easy integration in Web application
- **Unambiguous**
- **Actionable** (dereferencable)

URI → generated by tools under responsibility of local coordinator

URI of plant

`<http://phenome.fr/arch/2017/c17000118>`

URI of pot:

`<http://phenome.fr/arch/2013/pc13001542>`

URI of cart:

`<http://phenome.fr/arch/2013/ct1300123>`

URI of cabin:

`<http://phenome.fr/arch/2018/ac180015>`

URI of camera:

`<http://phenome.fr/arch/2018/ac180019>`



URI of image:

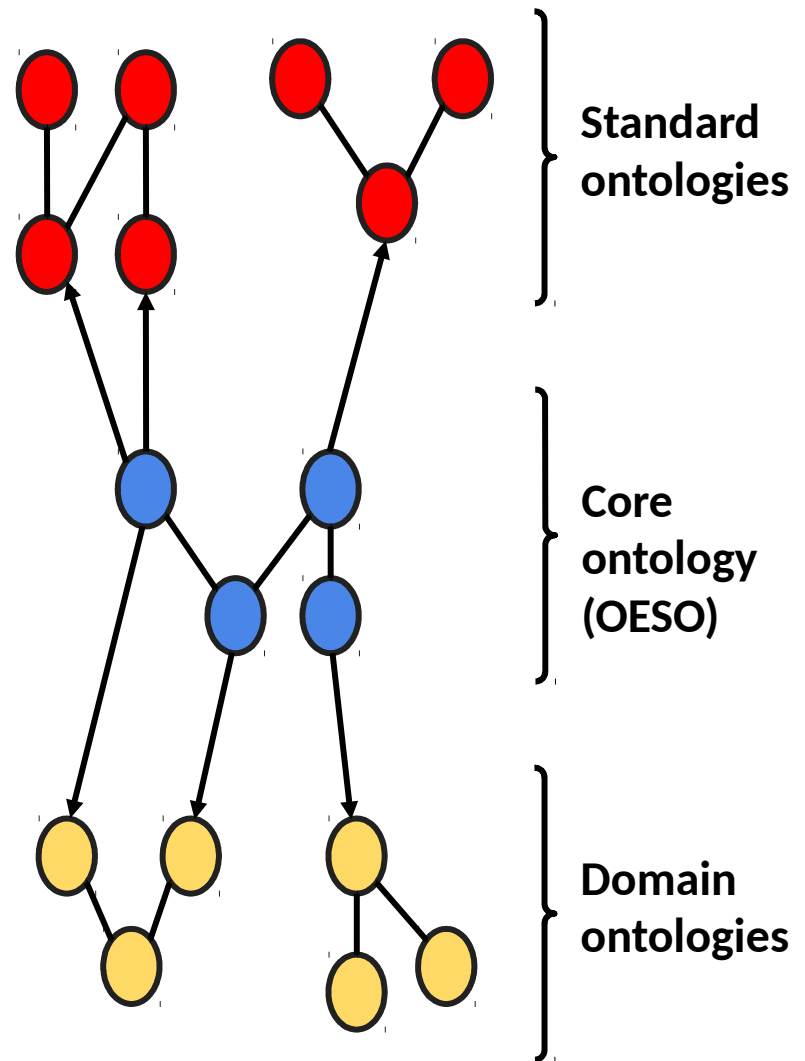
`<m3p:arch/2017/ic17002295855>`



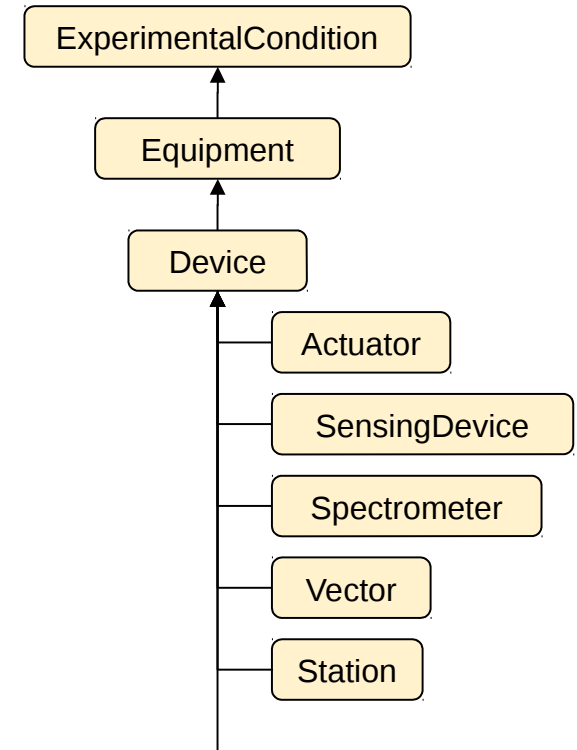
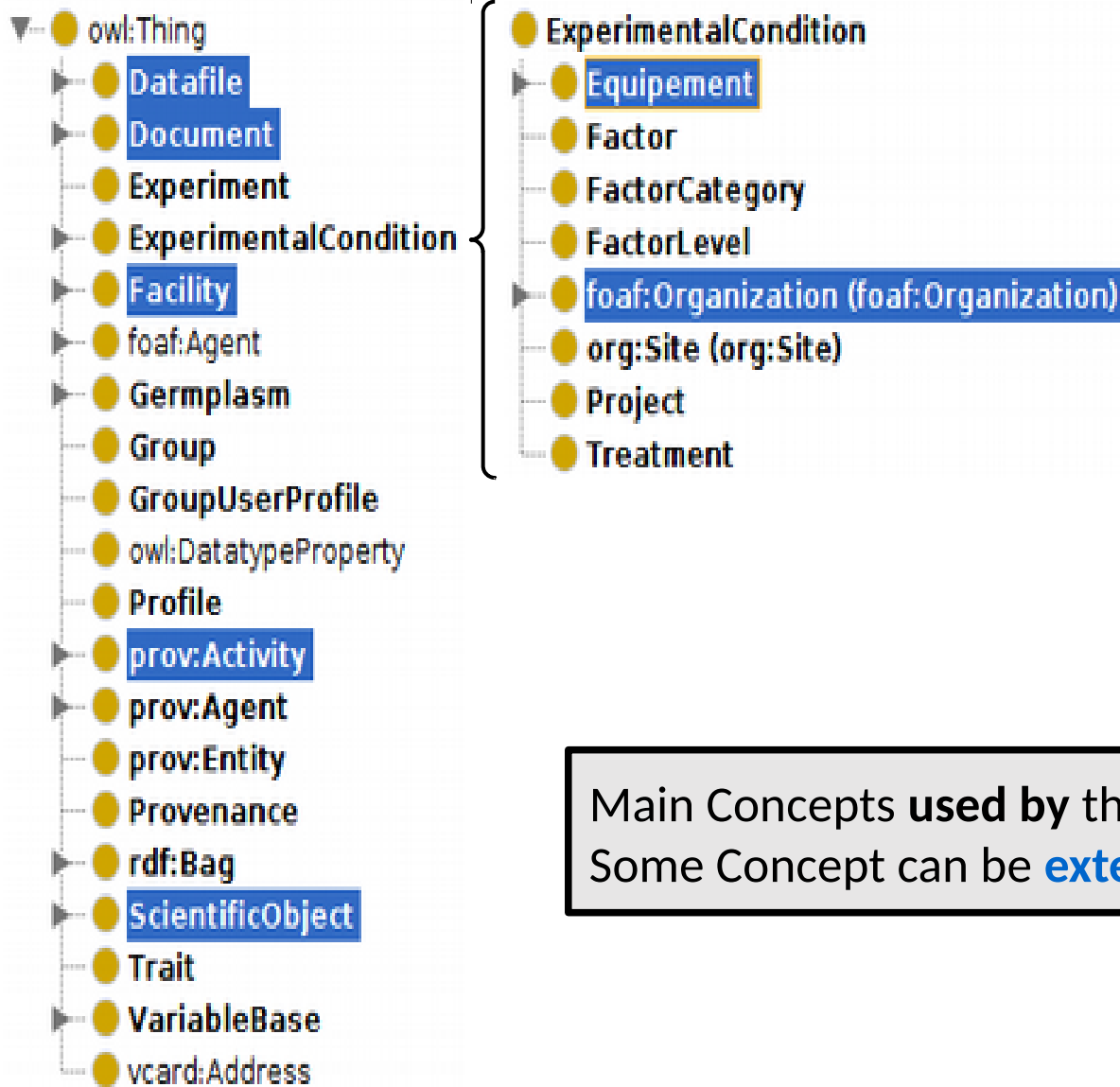
Ontology driven Information System

❖ Set of ontologies

- **Standard ontologies:** time, OA, DC, FOAF, PROV
- **Upper ontologies:** Dolce & BFO (used as a basis for conception)
- **Core ontology (OESO):** main concepts of OpenSILEX
- **Domain application ontologies:** specific to a domain or a community

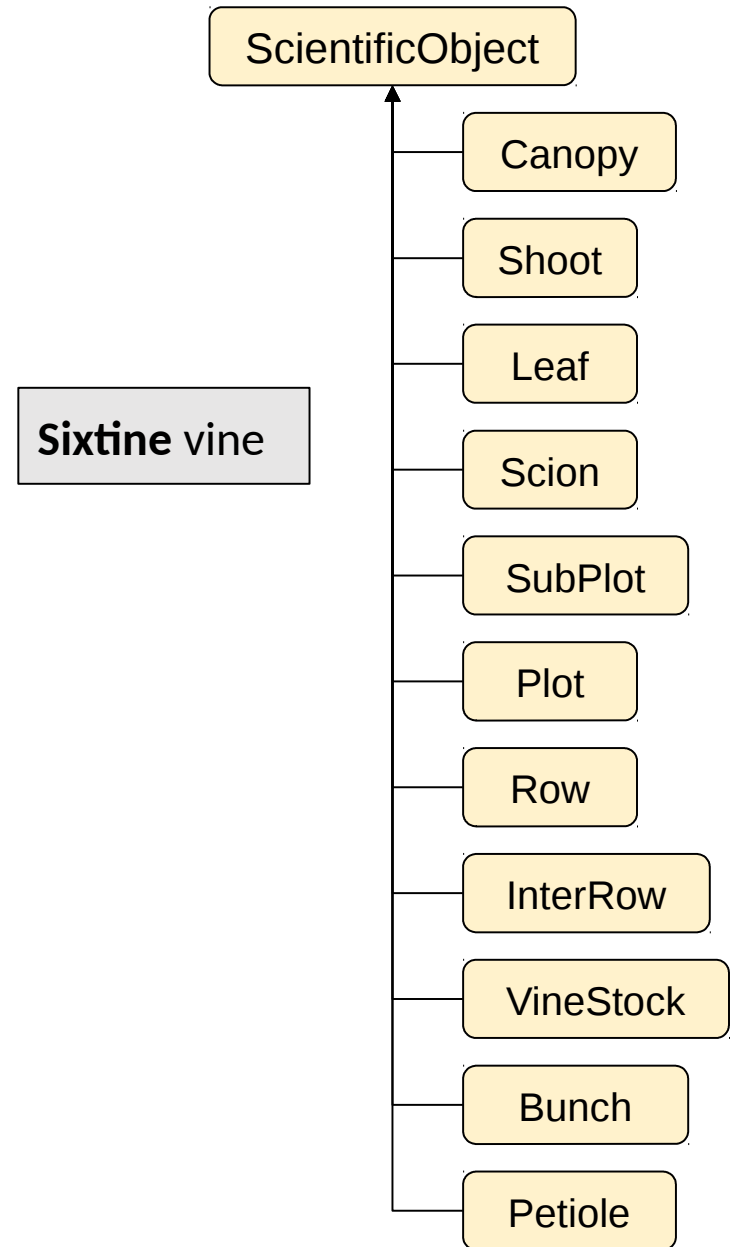
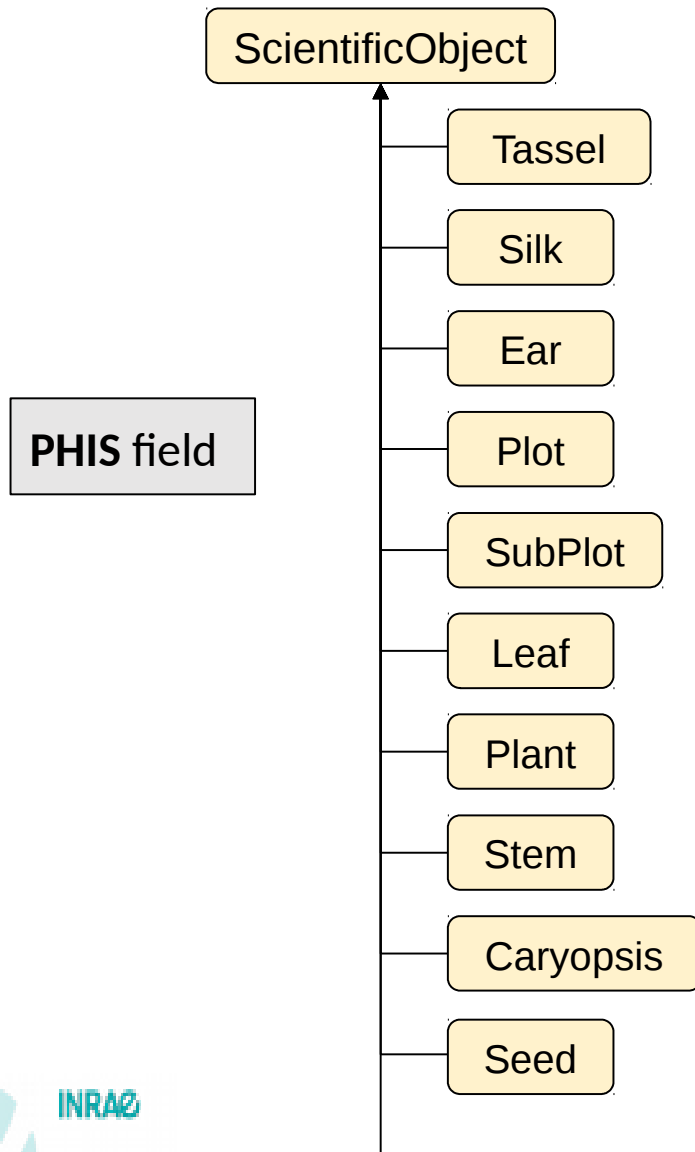


Core ontology OESO (OWL)



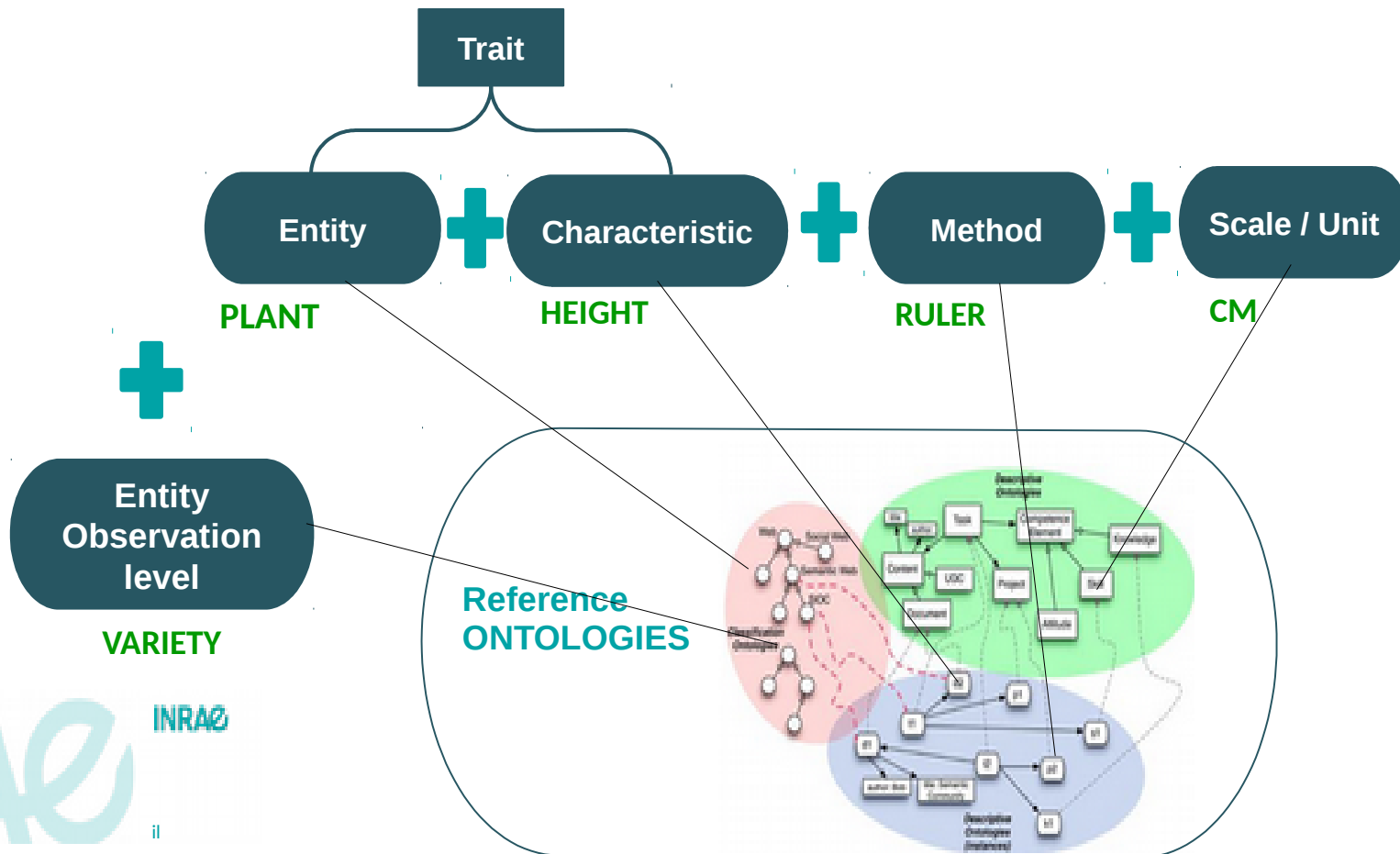
Main Concepts **used** by the application.
Some Concept can be **extended**

Example: PHIS and Sixtine



Variable Representation Model

- Enabling semantically precise descriptions
- Decomposing description into standardized elements
- Link to existing vocabularies/ontologies
- Make description machine readable





Structuring elements, contextual data, annotations, links are managed with Semantic Web tools and technologies. Named graph for internal structuration

Series of values, observation sets, spatial data are in JSON and stored in MongoDB

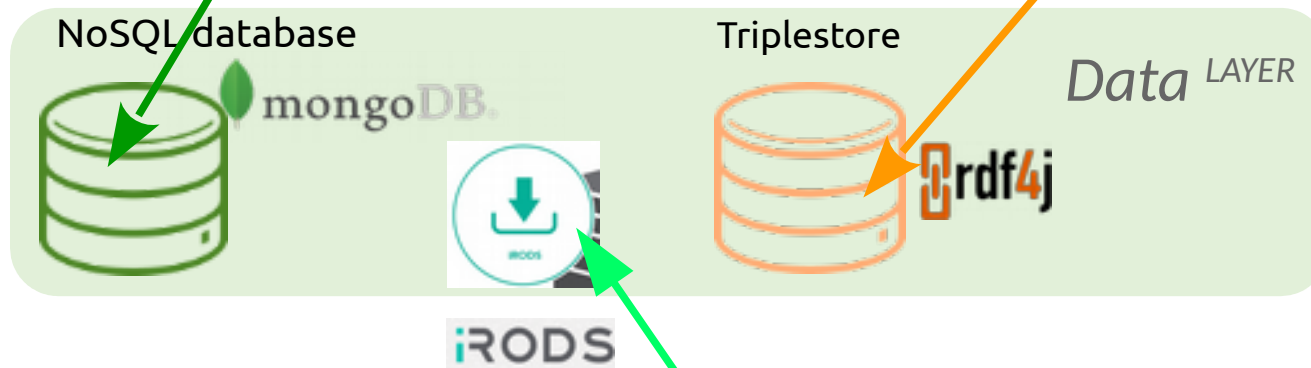
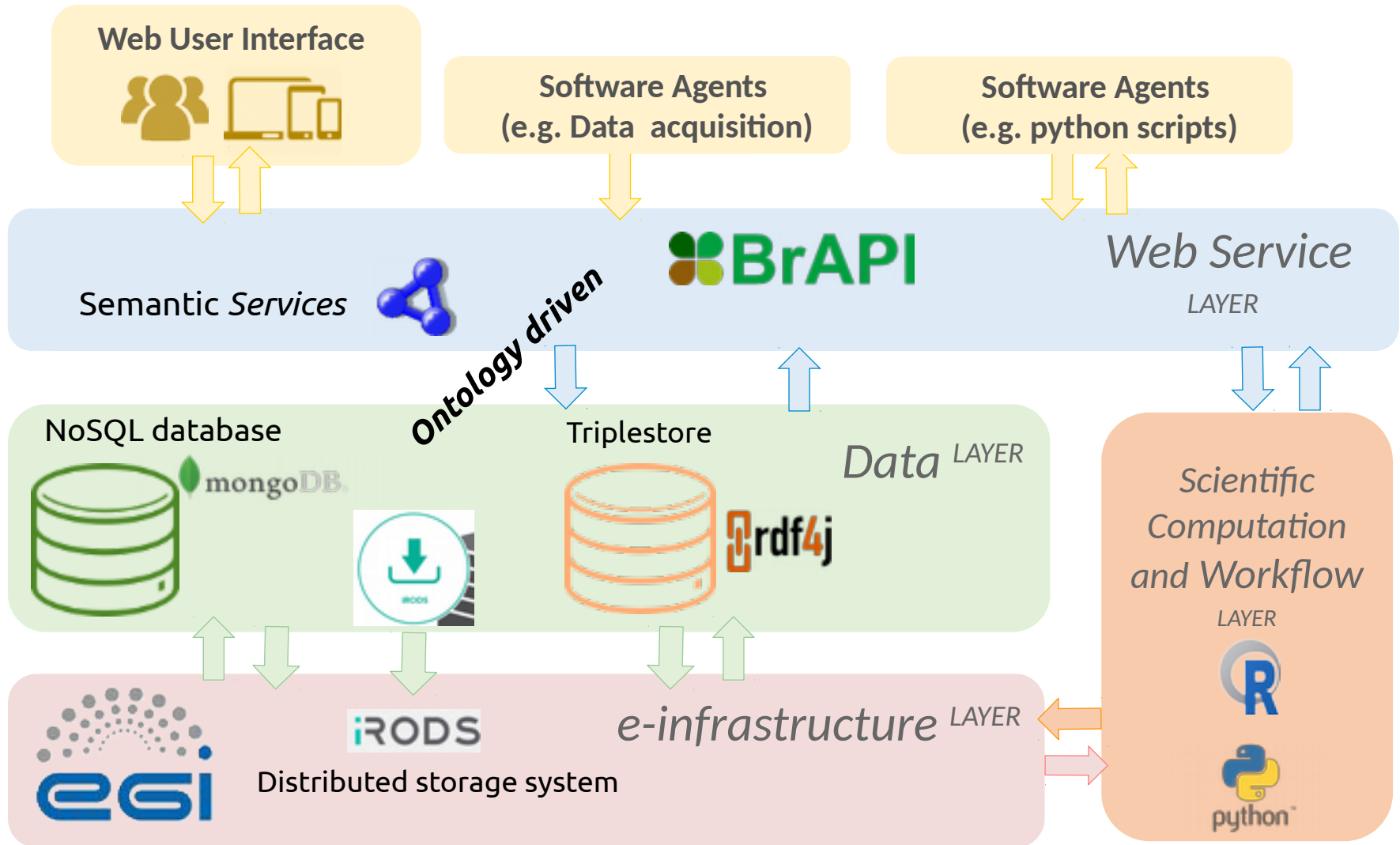


Image sets are stored on IRODS data system



OpenSILEX - Technologies



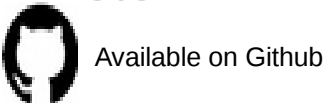
Tomcat



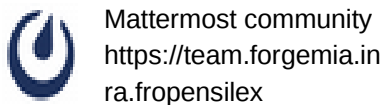
Modul architecture



CI/CD



Available on Github



Mattermost community
<https://team.forgemia.in>
ra.fropensilex



Mantis Bug Tracker



Protege Editor

This box contains the following technologies: Analyse (document icon), Geofolia (green leaf logo), a scale icon, OLGA (nature image), a red flower icon, Sensor (thermometer icon), Data INRAE (grid logo), Scripts (code icon), and Python (blue and yellow logo).

Web Interface

This box contains the following technologies: Vue.js (green V logo), a person icon, and swagger (green box with curly braces).



API, Web Services
REST / JSON

This box contains the following technologies: Java (coffee cup icon) and a purple cube icon.

Data

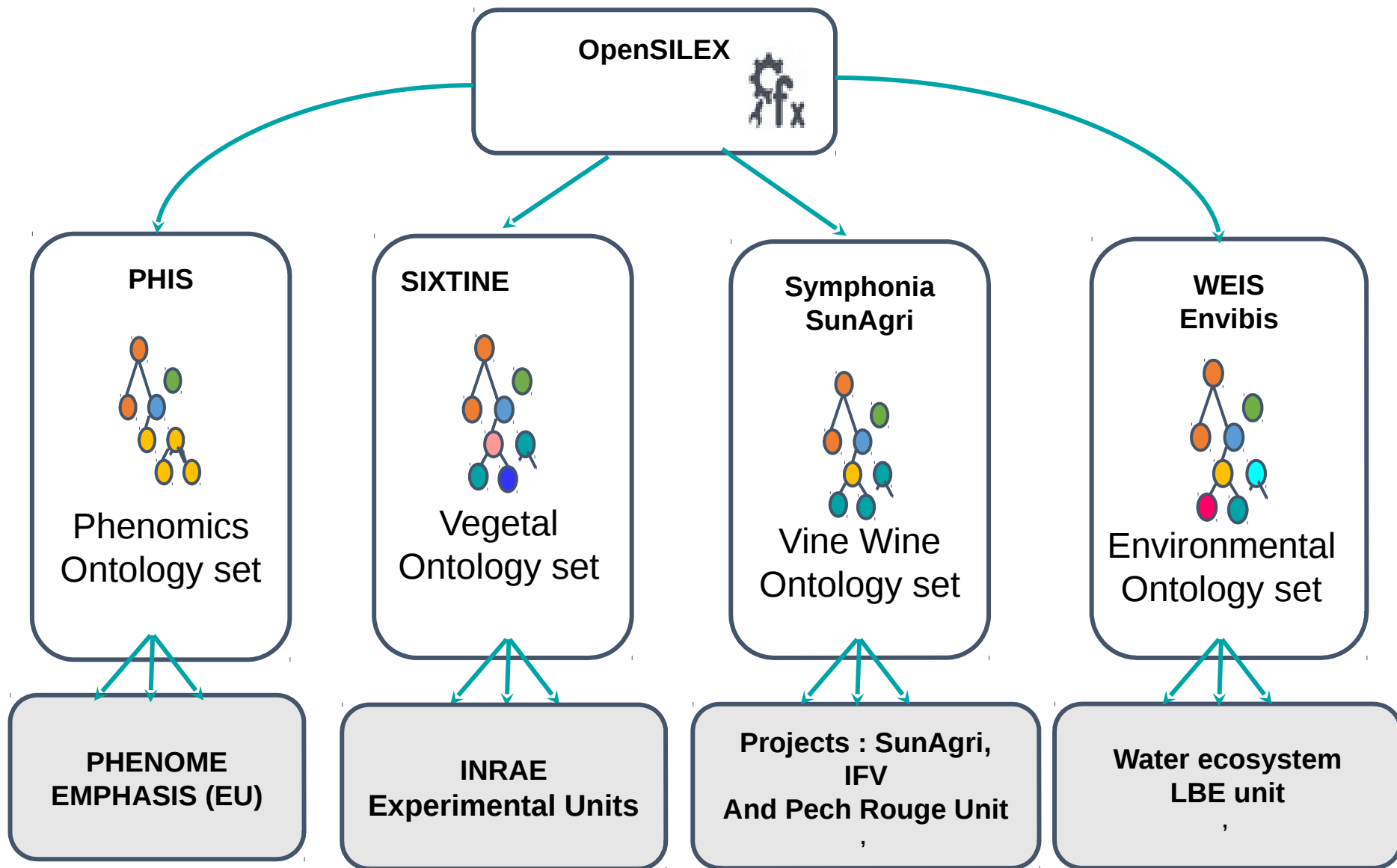
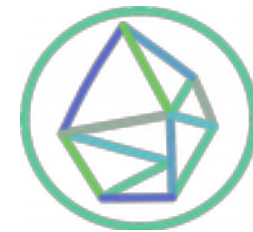
This box contains the following technologies: NOSQL (MongoDB) (cylinder icon), Triplestore (RDF4J, GraphDB) (cylinder icon), and Fichiers (document icon).

Infrastructure

This box contains the following technologies: EGI (sun icon) and a server rack icon.

OpenSILEX

Scientific Information Systems





PHIS Web Interfaces for the management of:

- Project information
- Experiment
- Facilities
- Devices
- Scientific objects
- Germplasms
- Experimental factors
- Data
- Data visualization
- Data provenance

The screenshot shows the PHIS web interface. The navigation menu on the left includes categories like Scientific Organization, Scientific Information, Germplasm, and Administration. The main content area displays a table of scientific objects.

Name	Type	Parent	Actions
maize	Species		[Red square icon]
Pearl millet	Species		[Red square icon]
poplar	Species		[Red square icon]
rice	Species		[Red square icon]
sorghum	Species		[Red square icon]
teosintes	Species		[Red square icon]
upland cotton	Species		[Red square icon]
accPoplar	Accession	poplar	[Blue pencil icon] [Red square icon]
B73	Variety	maize	[Blue pencil icon] [Red square icon]
banana	Species		[Red square icon]
barley	Species		[Red square icon]
BC-seedlot-normais	Seed Lot	banana	[Blue pencil icon] [Red square icon]
bread wheat	Species		[Red square icon]
CRAZI	Variety	maize	[Blue pencil icon] [Red square icon]
DKC4590	Variety	maize	[Blue pencil icon] [Red square icon]

Example of flexibility

- Variable declaration
- Interoperability

Add variable

URI
autogenerated URI

Entity * ⓘ Quality * ⓘ

Method ⓘ Unit *

Name * Alternative name

Data type ⓘ

Time interval ⓘ Sample interval ⓘ

Description

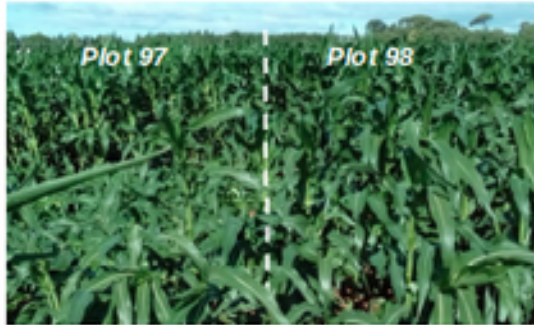
Ontologies References

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

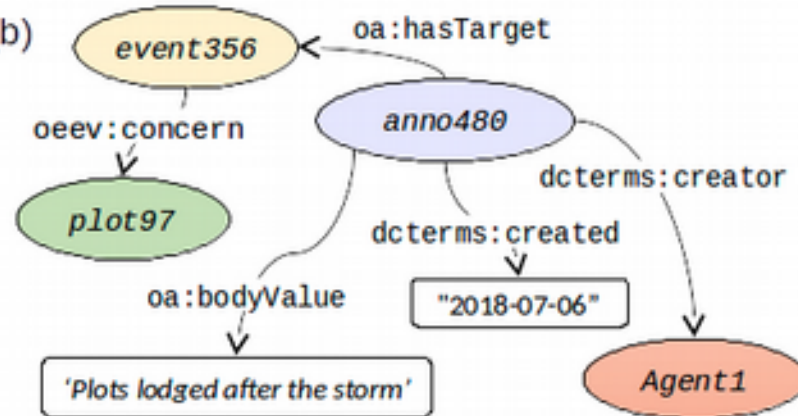
- [AGROPORTAL](#) ⓘ
- [AGROMOC](#) ⓘ
- [PLANT ONTOLOGY](#) ⓘ
- [PLANTEOME](#) ⓘ
- [CROP ONTOLOGY](#) ⓘ
- [UNIT ONTOLOGY](#) ⓘ



(a)



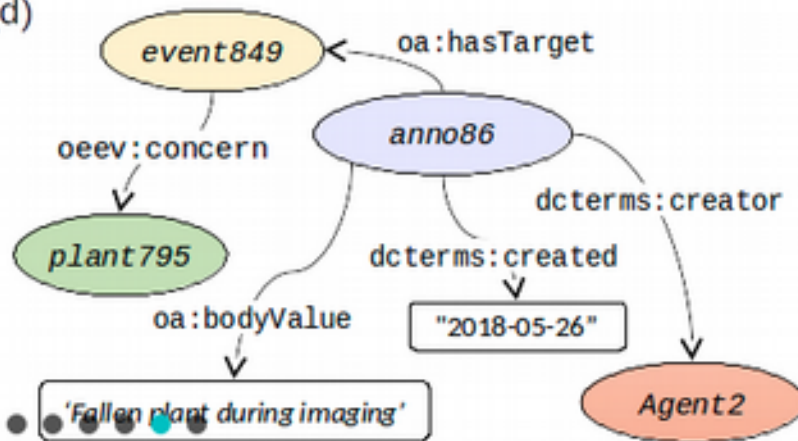
(b)

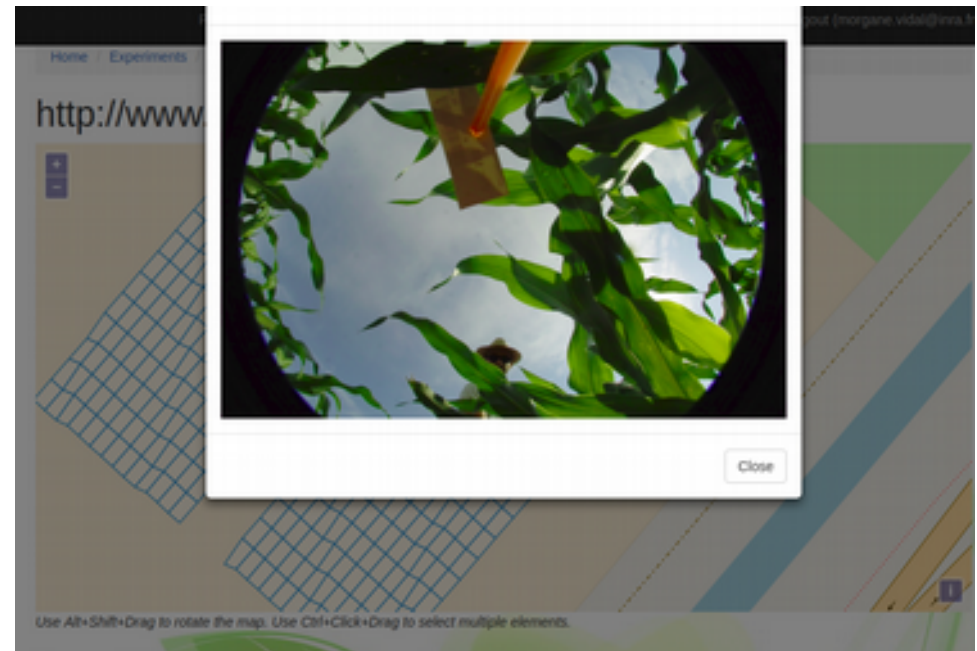
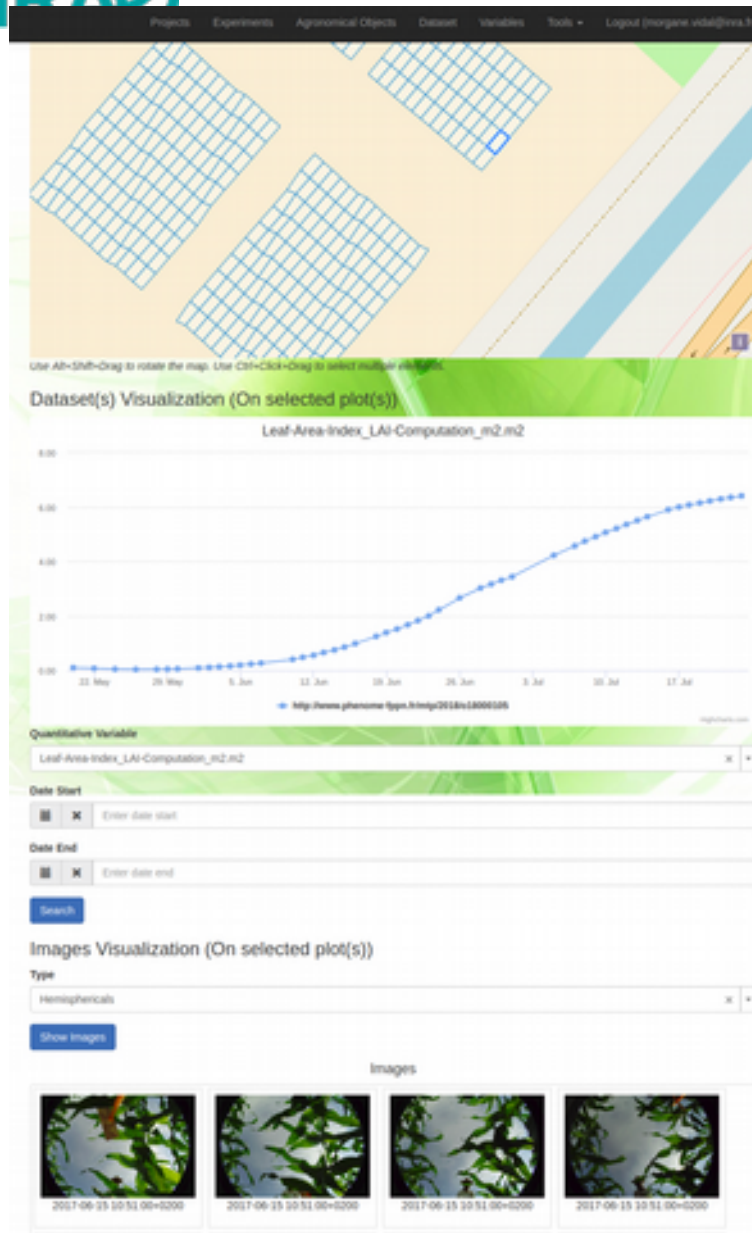


(c)



(d)







- ✓ Allows management of huge and complex data (petabytes)
- ✓ Enables and facilitates cloud computing (data center, EGI)
 - distributed computing, distributed storage, backup
- ✓ **Manages semantics** (ontologies, standardized vocabularies)
- ✓ Provides a flexible design
- ✓ **Provides provenance** and reproducibility for data processing
- ✓ Different Implementations:
 - PHIS, Sixtine, Symphonia, SunAGRI, WEIS, Envibis
- ✓ **Open Software** <https://github.com/OpenSILEX>
- ✓ **Web site:** <http://www.opensilex.org/>

Conclusion

- ✓ **Giving value to complex data requires structuring according to FAIR principles**
- ✓ **A better formalization of concepts (using ontologies) and data is required interdisciplinary research**
- ✓ **To deal with data complexity a new generation of information systems (e.g. Ontology driven) is needed**
- ✓ **Advanced data management makes data available for AI and data analytics**

Remerciements



INRAE

la science pour la vie, l'humain, la terre



PHENOME
EMPHASIS FRANCE



Sun'Agri



Département
MathNum



EMPHASIS



Lbe
Laboratoire de
Biotechnologie de
l'Environnement



東京大学
THE UNIVERSITY OF TOKYO



Mistea
Mathématiques, Informatique et Statistique
pour l'Environnement et l'Agronomie



IFV
INSTITUT FRANÇAIS
DE LA VIGNE ET DU VIN



WAGENINGEN
UNIVERSITY & RESEARCH



UNIVERSITY OF
CANBERRA



Laboratoire d'Ecophysiologie des Plantes
Sous Stress Environnementaux
lepse
Montpellier



WAGENINGEN
UNIVERSITY & RESEARCH



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI



SPo
UMR

Sciences pour
l'œnologie

