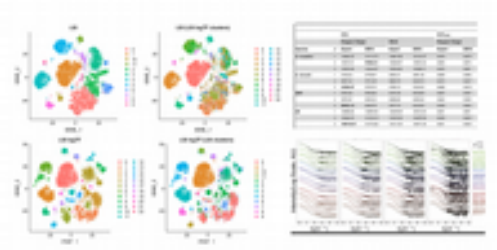
OpenSILEX

Pour des Systèmes d'Information pilotés par des ontologies

Contact : Pascal.Neveu@inrae.fr

CONTEXT

- **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 **transparence** and reproducibility

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

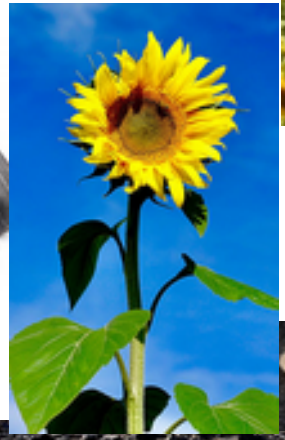
Complex Data

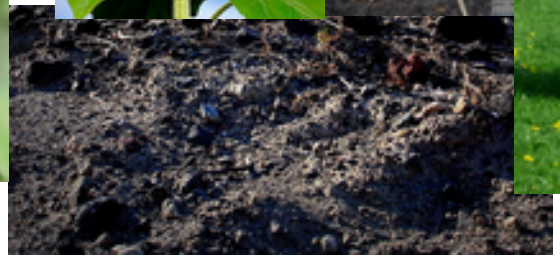
Different environments and ecosystems

Different stages

Different scales

Different interactions





Complex Data

Instrumentation that evolves and used at various frequencies

In heterogeneous facilities

By Different teams

« omics » Platforms

Various data complex types

Genomics

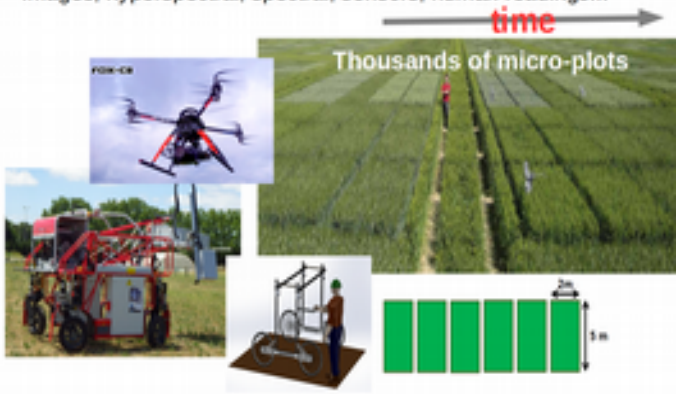
Composition and the structure of biopolymers

Quantification of metabolites and enzyme activities

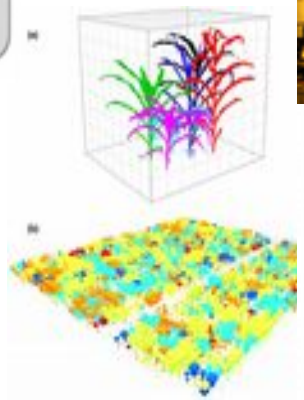
Field Platforms

Various scales and data types

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

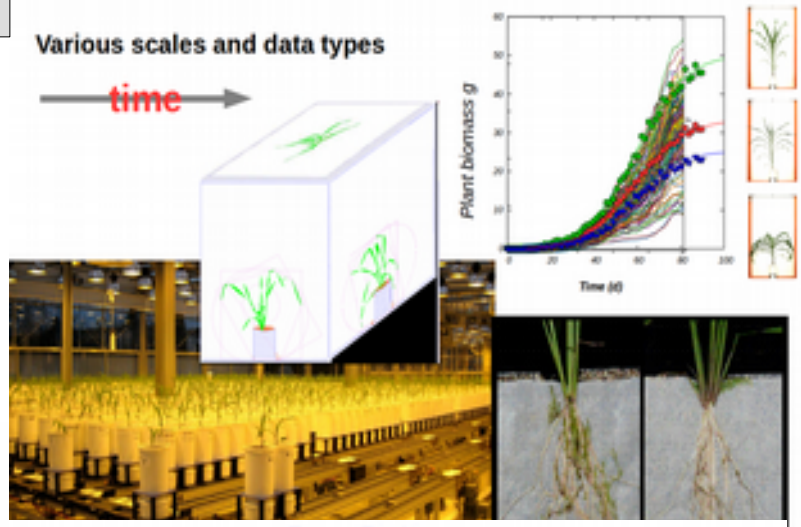


Virtual experiments



Green house Platforms

Various scales and data types



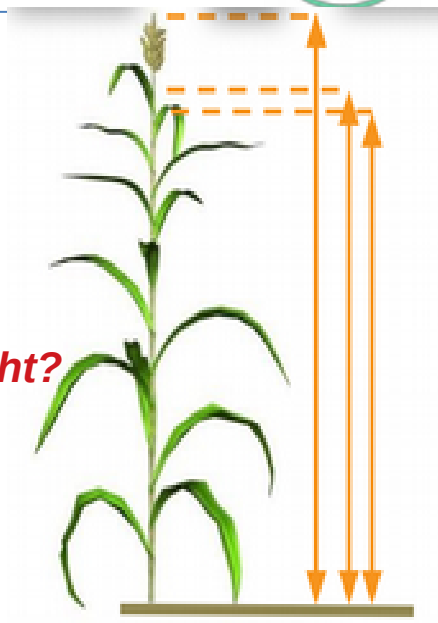
Farm Platforms

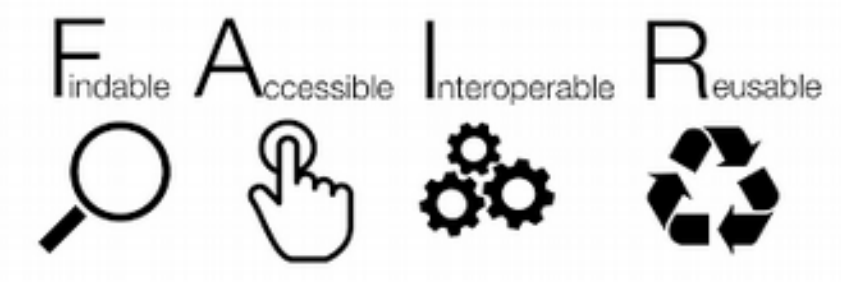
Various scales and data types from thousands of farms

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

Some common mistakes we do

- **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 uncompatible organisations)
- **Ambiguous ID**
- **Context, faults are not described**
- **No data links**
- **No missing data representation**
- **No licence**
- **Etc.**





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***

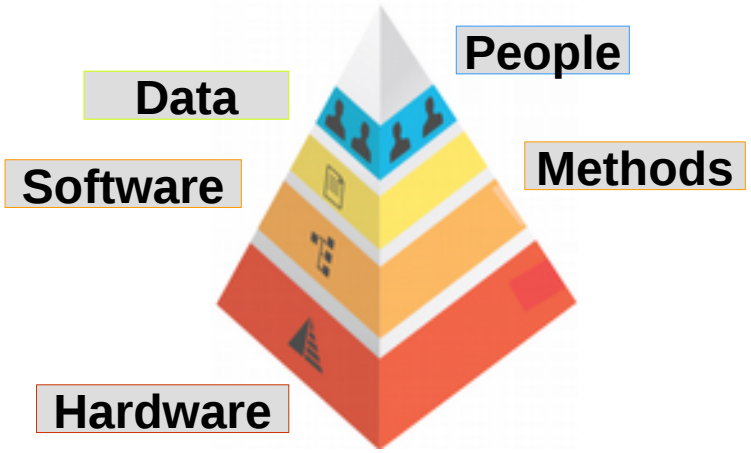
Complex Data Challenge

A need for a new generation of Information Systems

OpenSILEX

OpenSILEX: an open source software set

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



OpenSILEX: Structuring of Data

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 an standardized vocabulary
  - Representation models
  - Formalized relationships between entities
  - Data annotation and enrichment (search engine friendly)

Data Structuring: approach

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)*

*Semantic Web Languages

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>

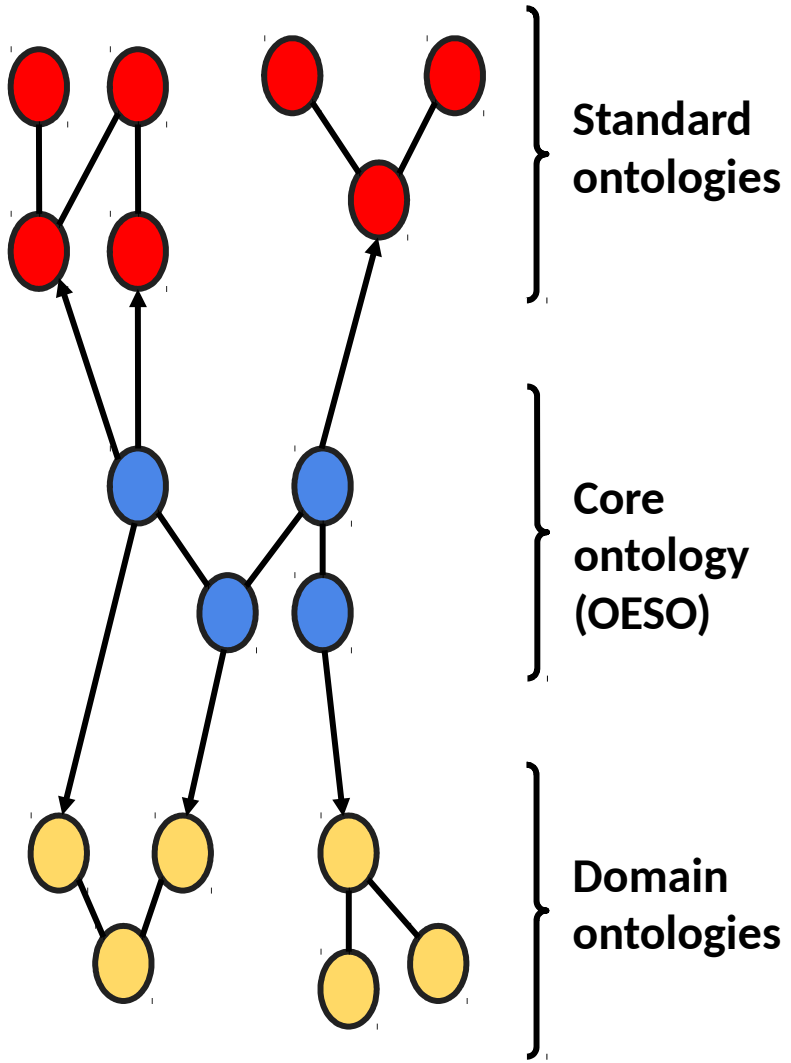
URI of cabin:
<http://phenome.fr/arch/2018/ac180015>

URI of camera:
<http://phenome.fr/arch/2018/ac180019>

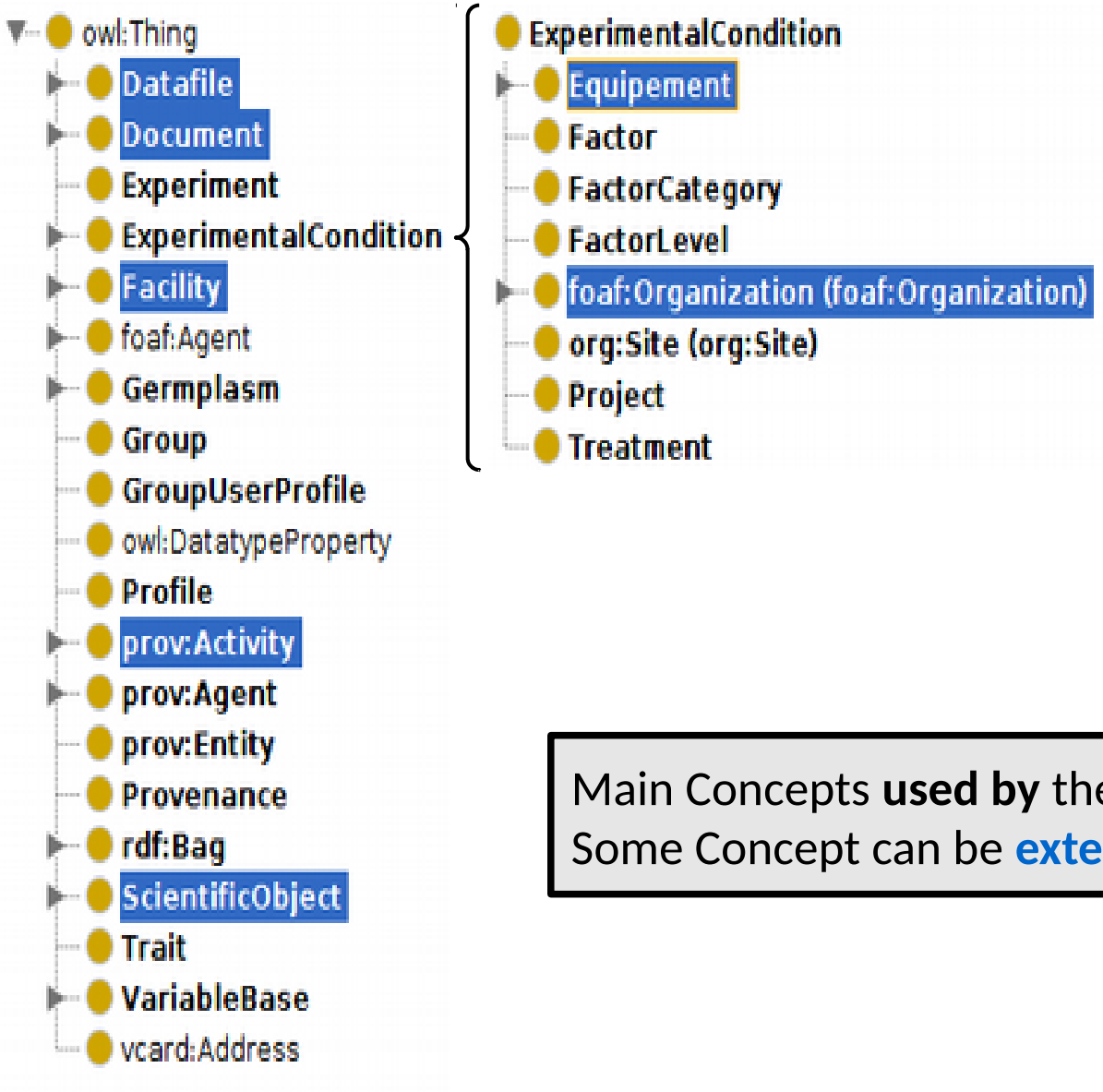
URI of image:
<m3p:arch/2017/ic17002295855>

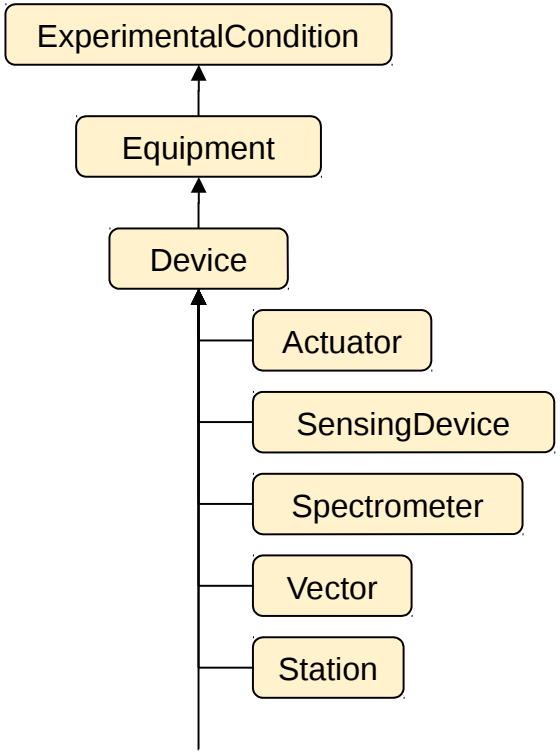
Ontology driven Information System

- 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



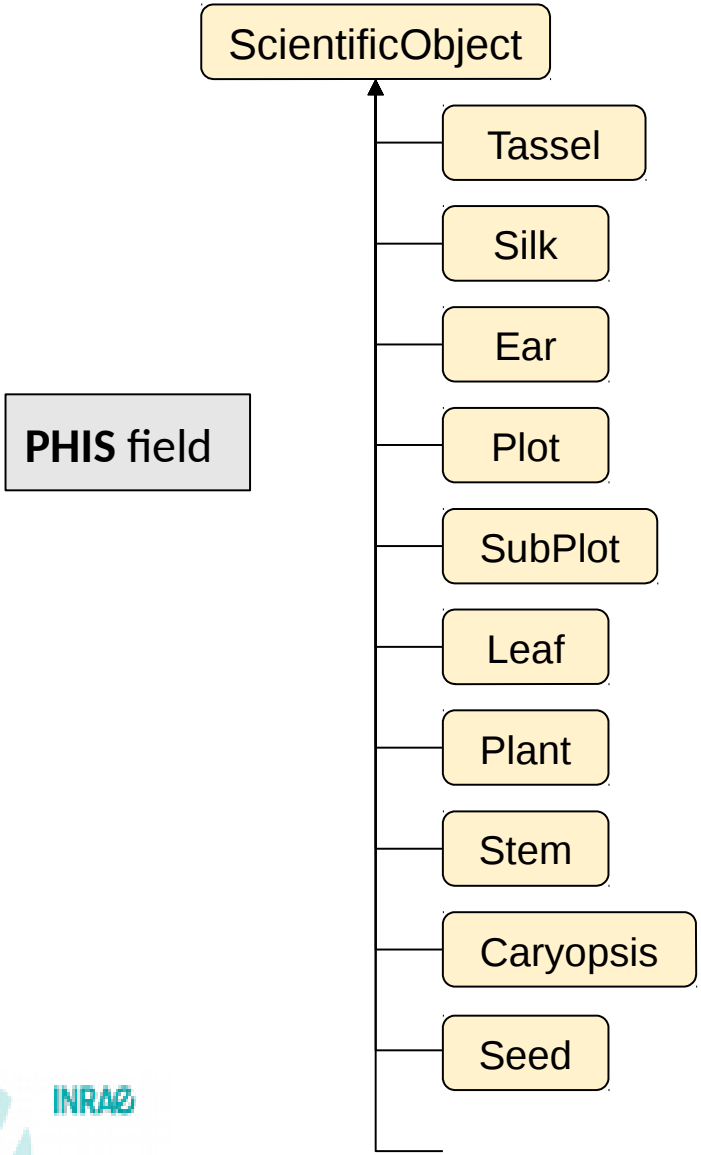
Core ontology OESO (OWL)

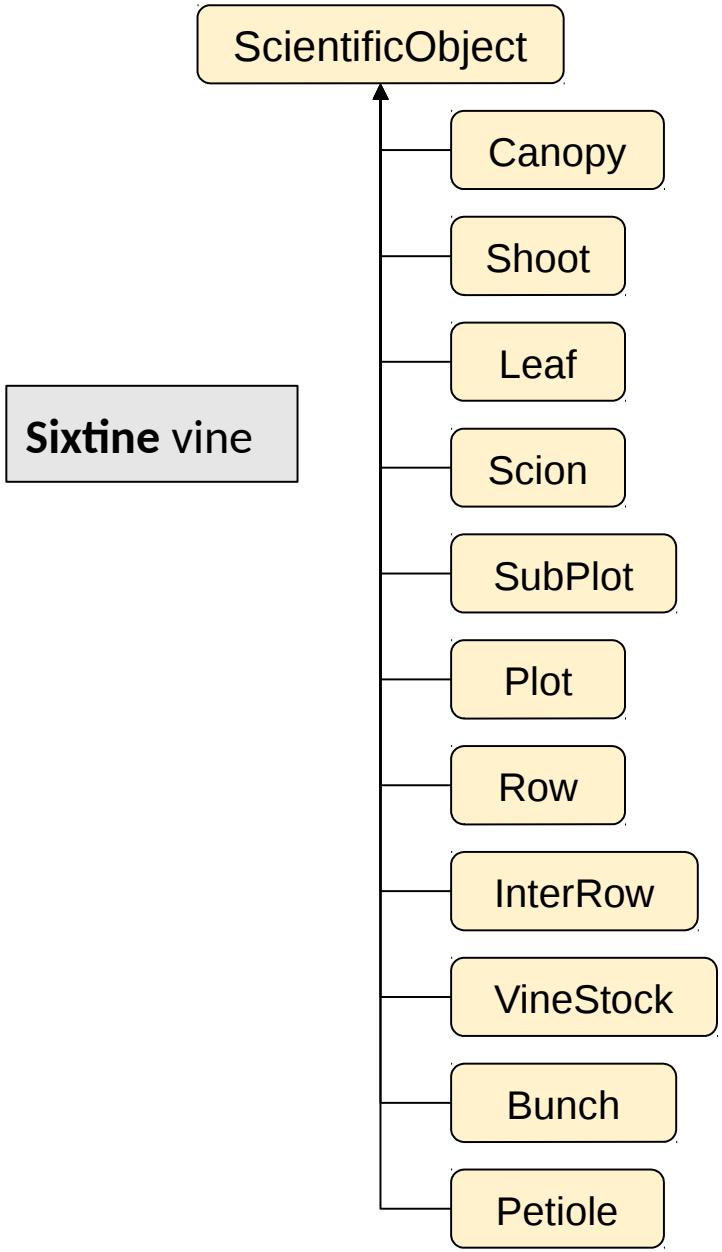




Main Concepts **used by** the application.
Some Concept can be **extended**

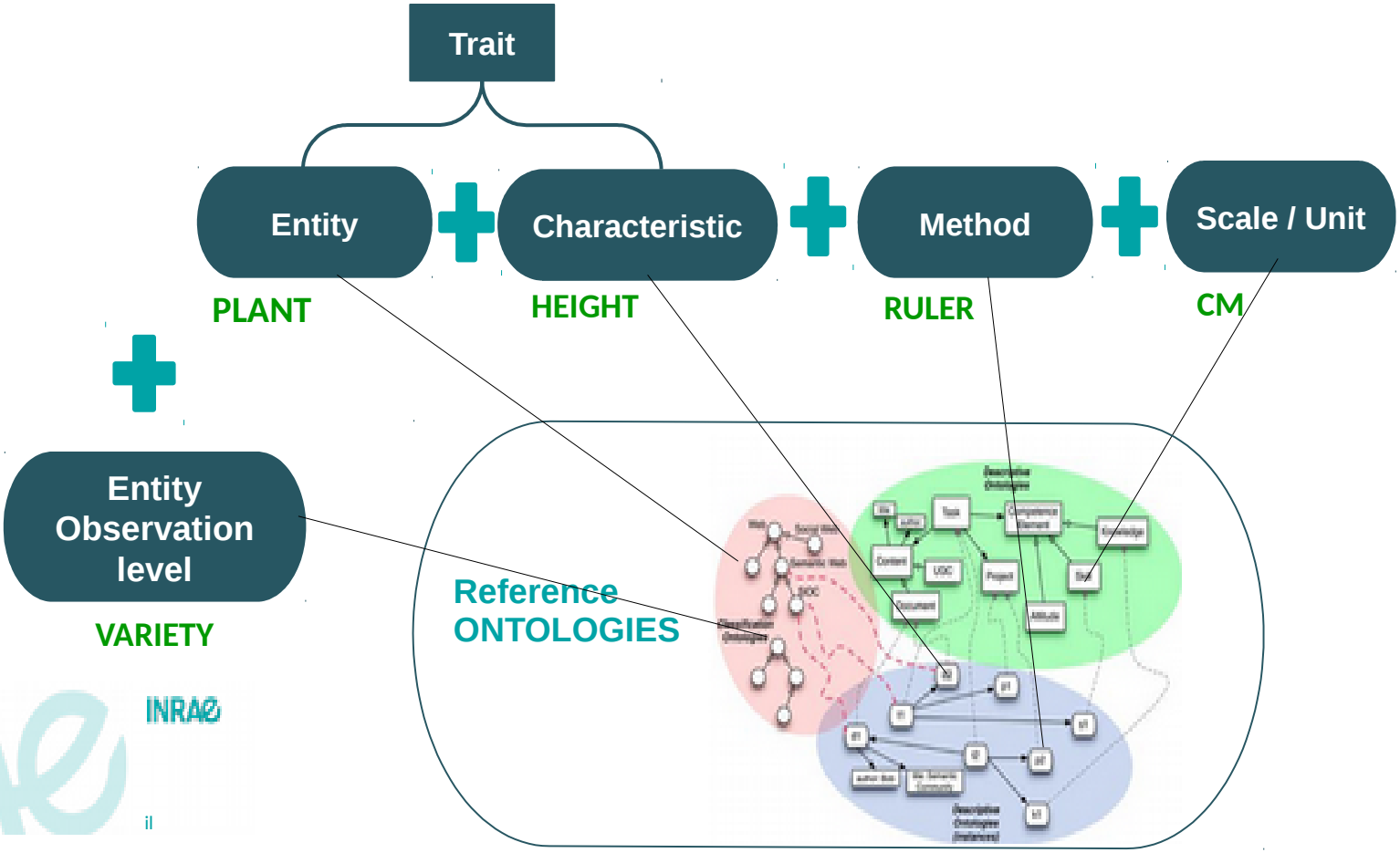
Example: PHIS and Sixtine





Variable Representation Model

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



Data Organisation

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

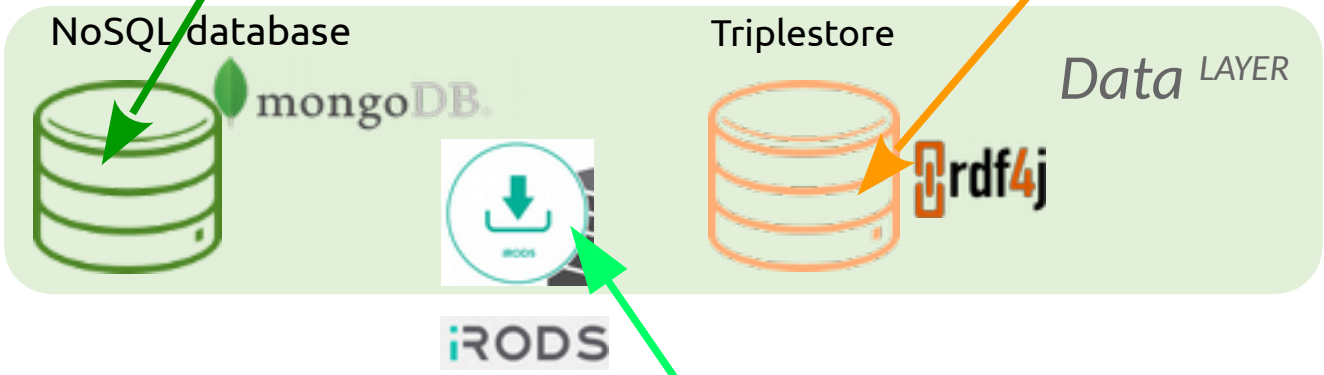
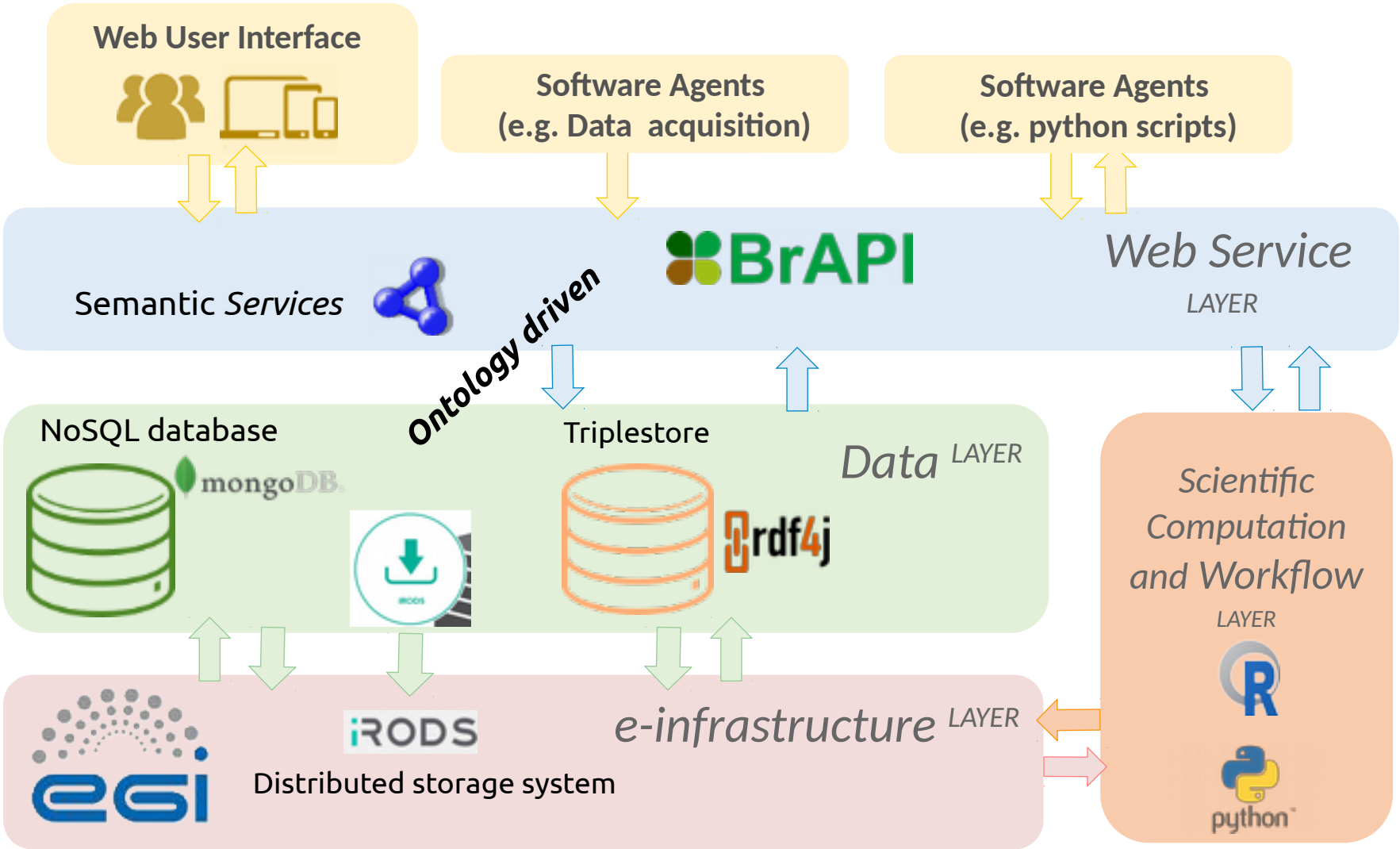


Image sets are stored on IRODS data system

OpenSILEX Architecture



OpenSILEX - Technologies

Tomcat

Maven
Modul architecture

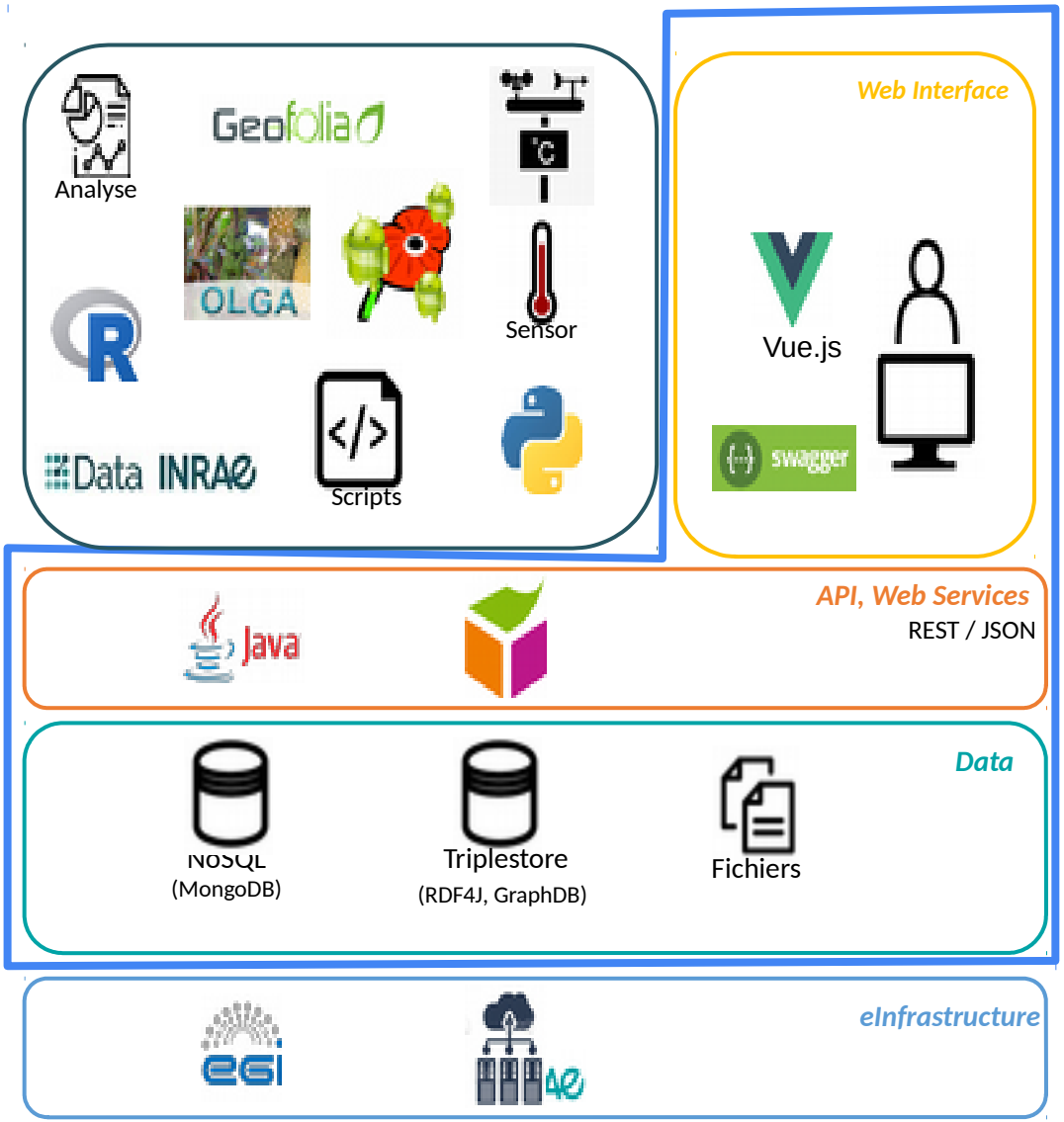
GitLab
CI/CD

Available on Github

Mattermost community
https://team.forgemia.inra.fropensilex

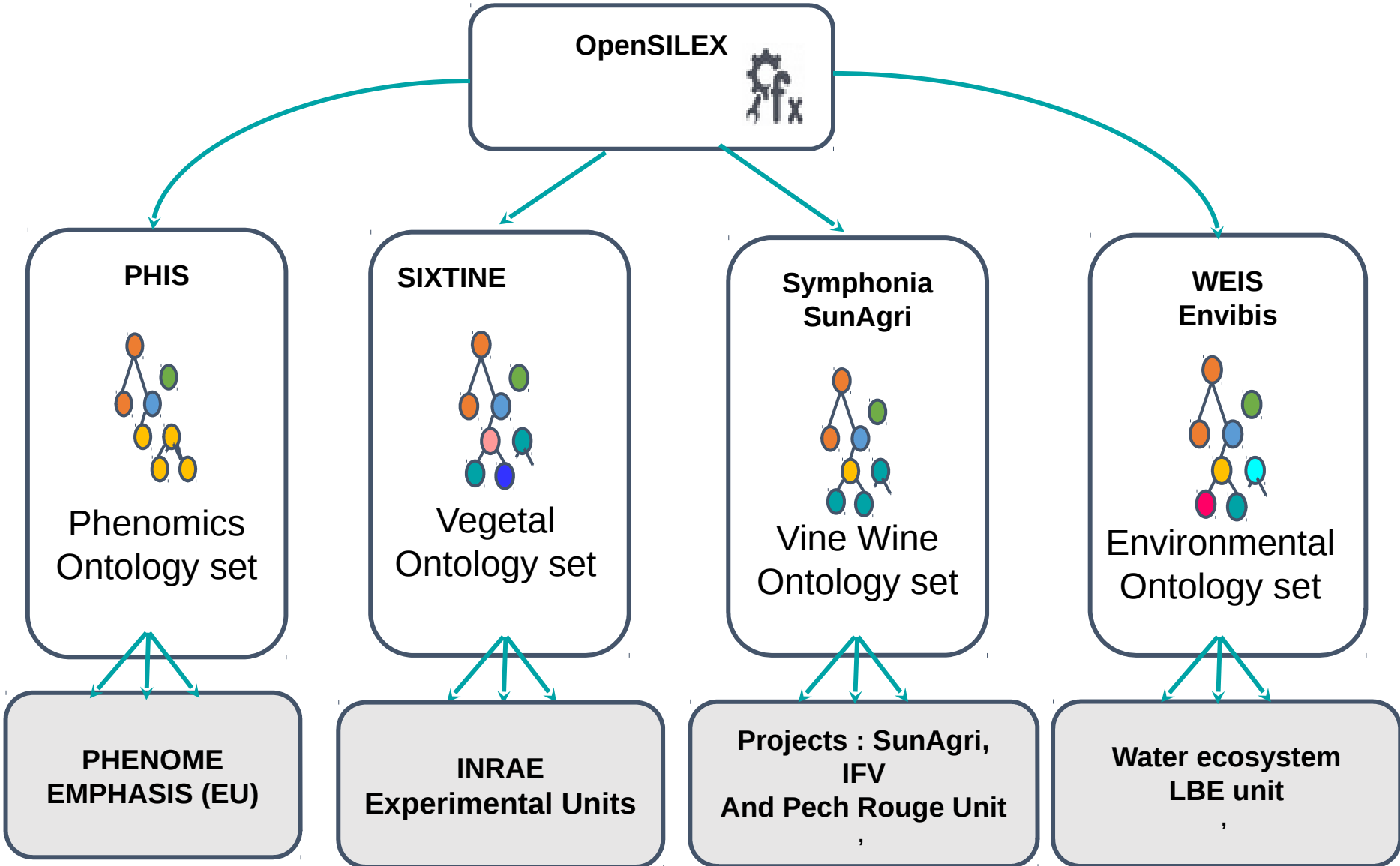
Mantis Bug Tracker

Protege
Editor



docker

OpenSILEX
Scientific Information Systems



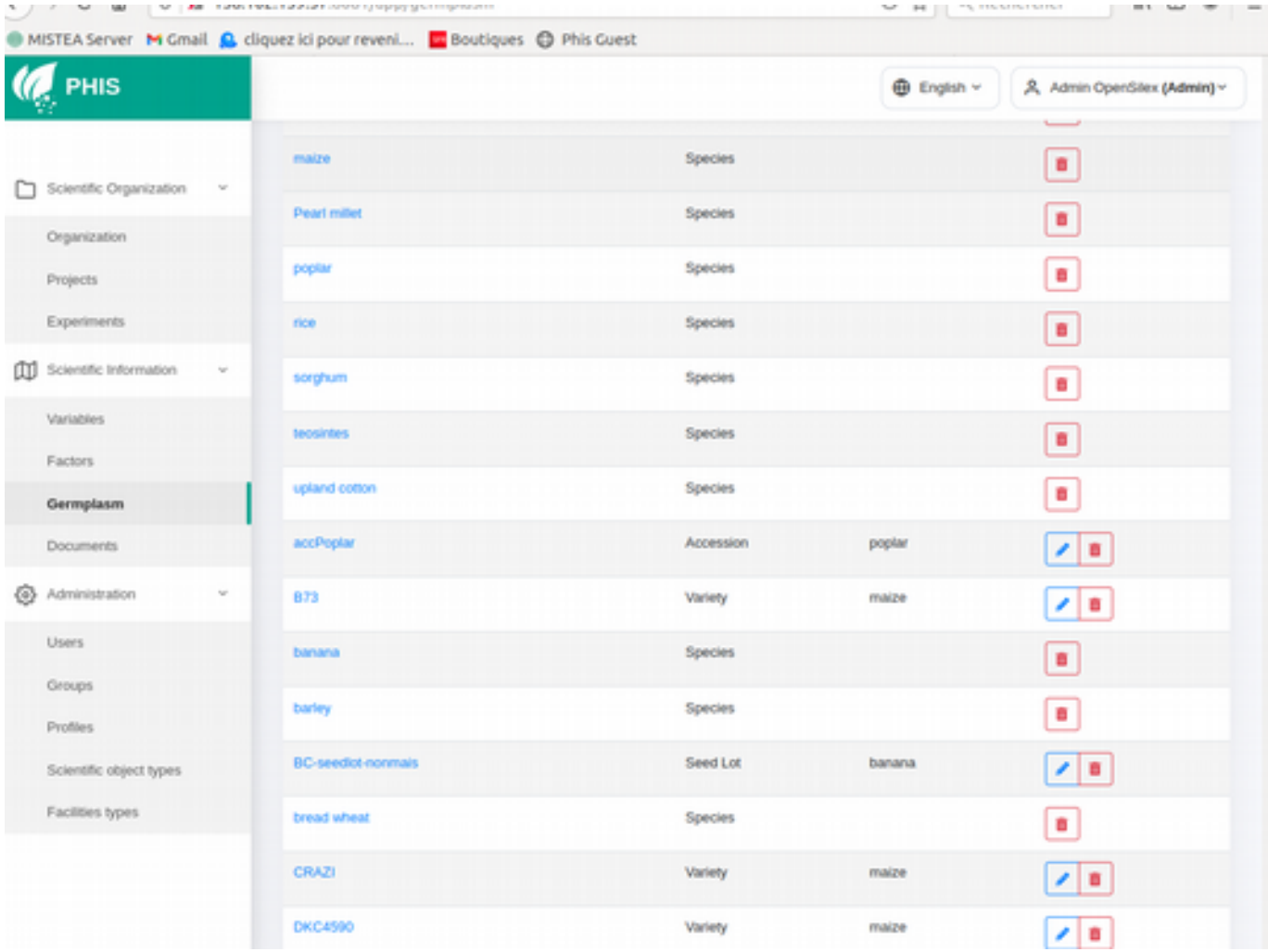
PHIS User Interface

PHIS Web Interfaces for the management of:

- Project information
- Experiment
- Facilities
- Devices

- Scientific objects
- Germplasms
- Experimental factors

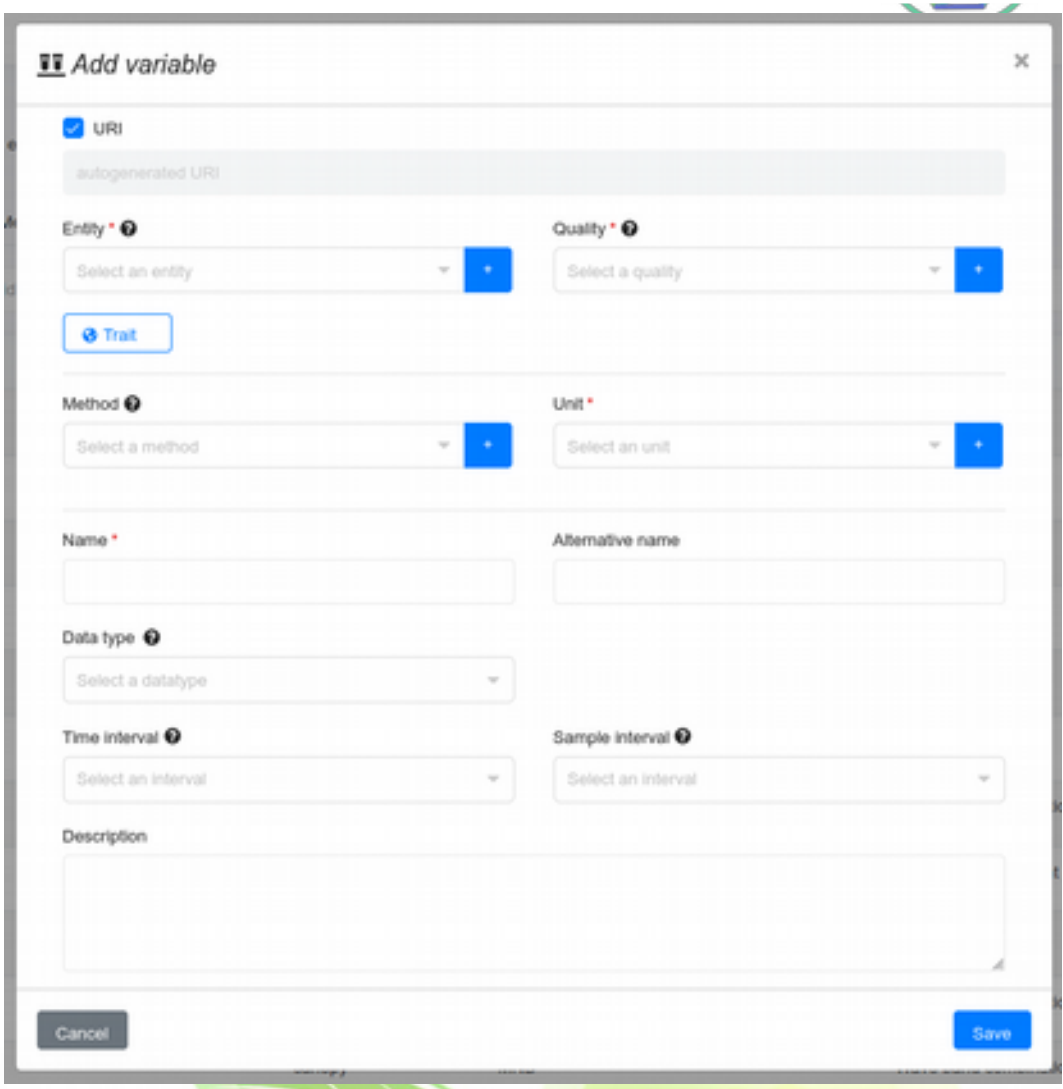
- Data
- Data visualization
- Data provenance



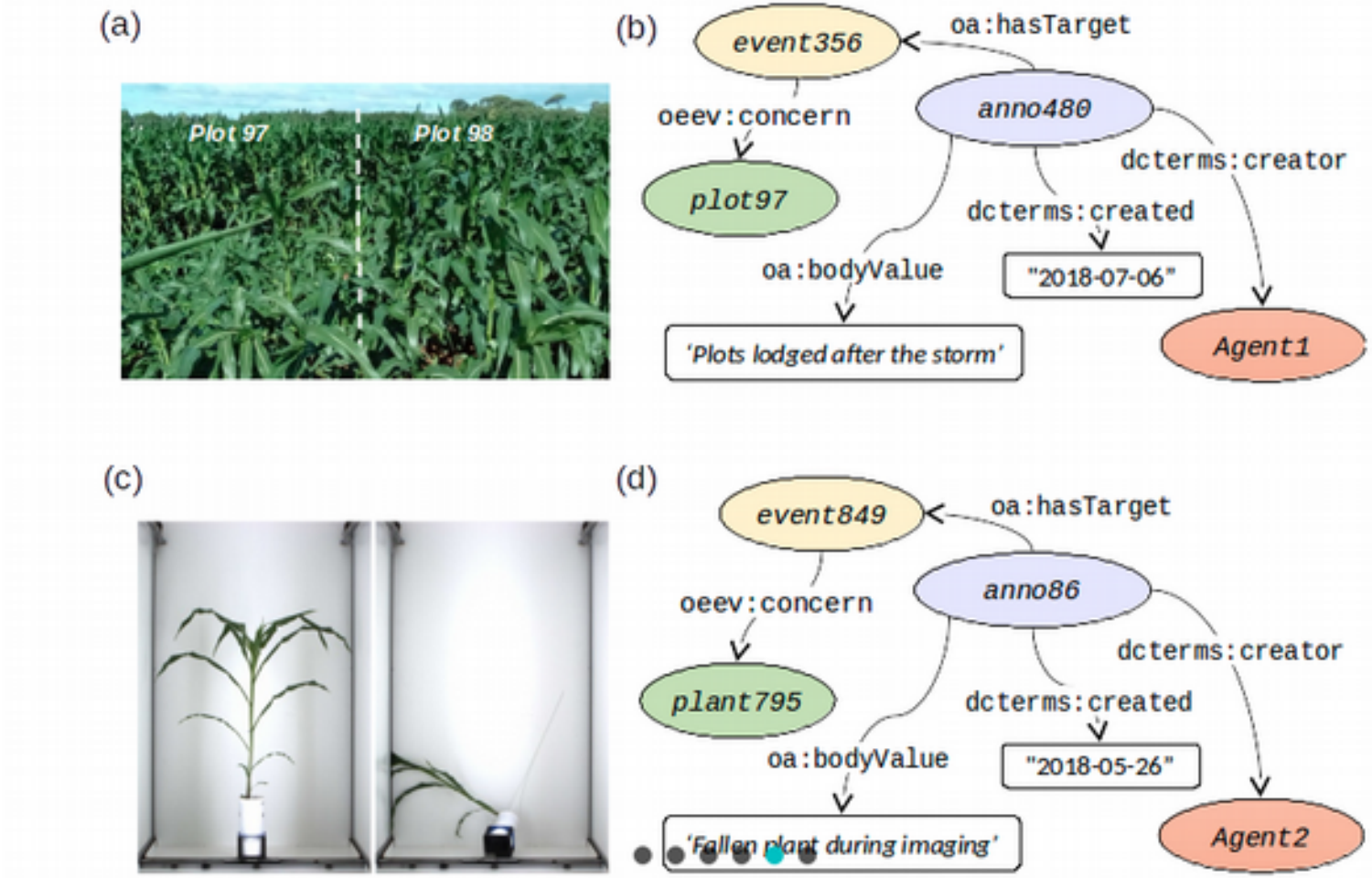
PHIS User Interface

Example of flexibilty

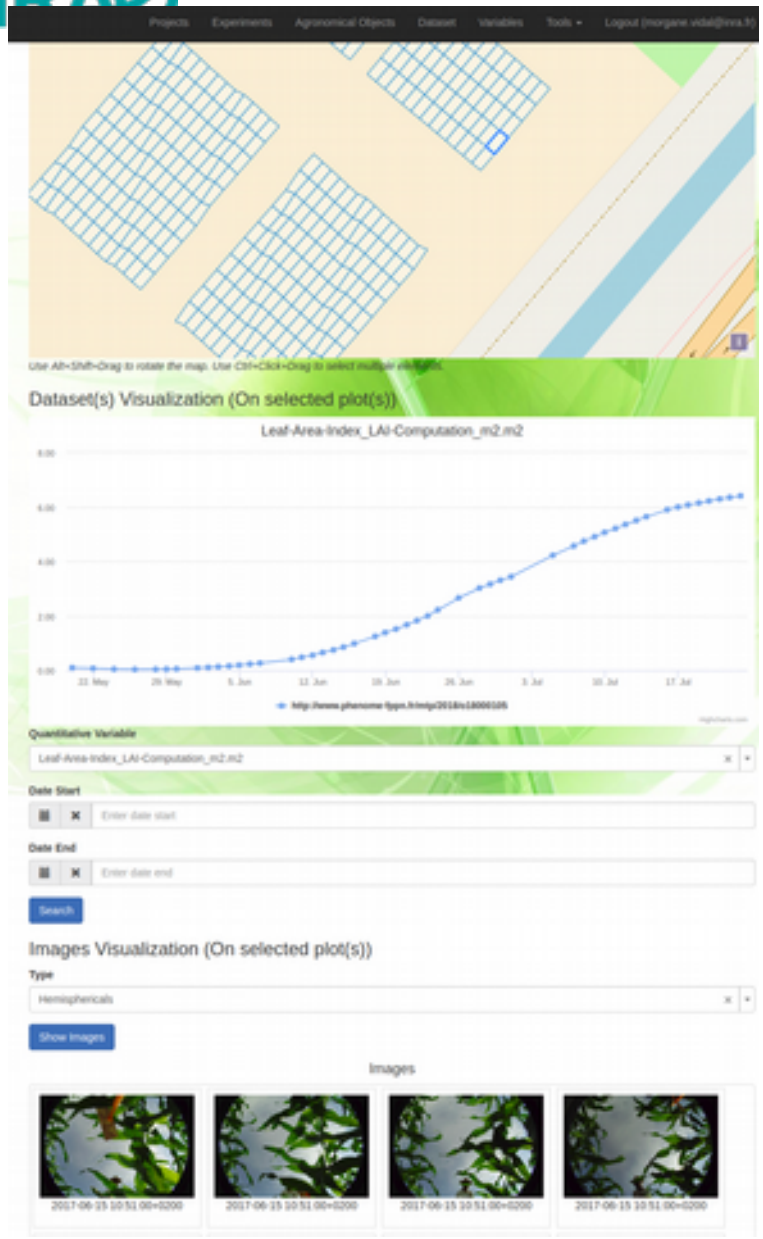
- Variable declaration
- Interoperability

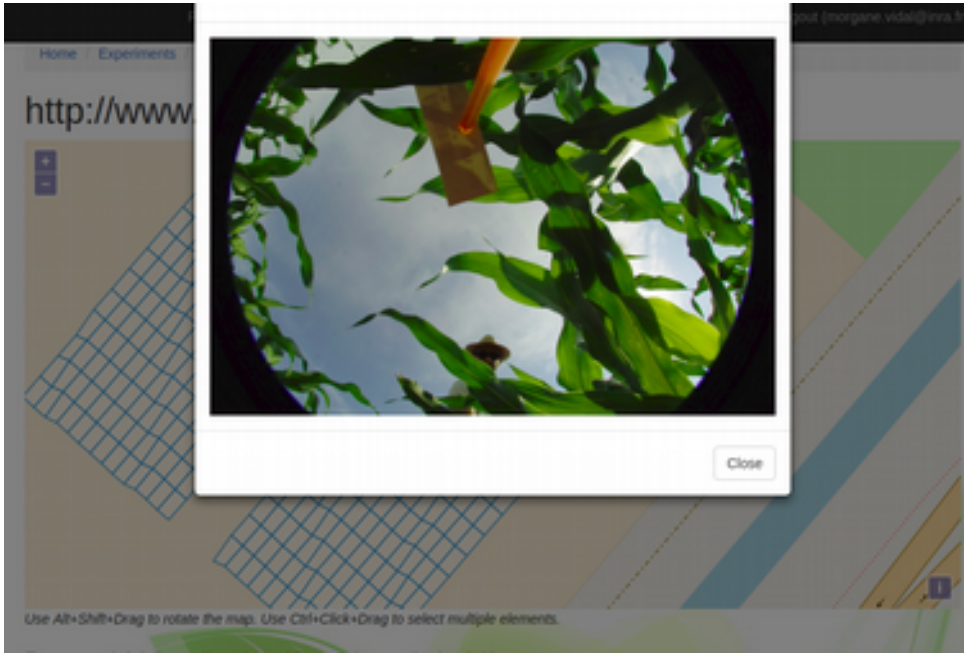


Ontology driven Information System



Trait – Provenance





In short OpenSILEX

- ✔ **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