INRAE

PHIS

Ontology Driven Information Systems For Agriculture and Environment

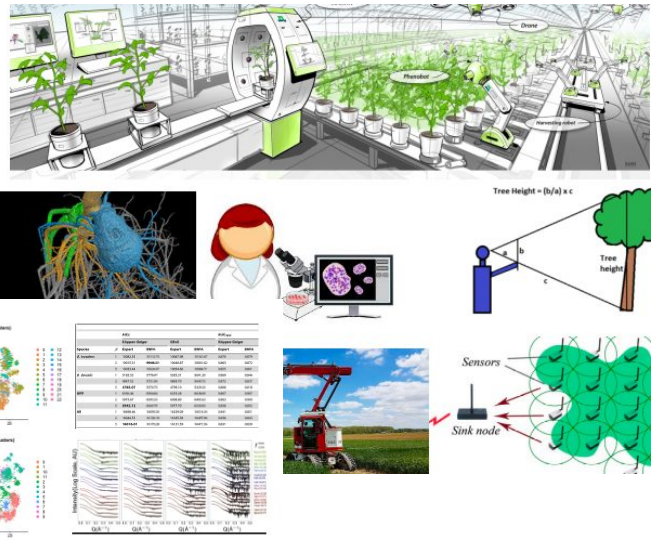
Pascal.Neveu@inrae.fr

Département MathNum

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

Data Challenge For Plant Phenotyping

- **Experimentations or Observations**
  - Complex and cannot be reproduced
  - A lot of various resource requirement
  - Huge and **very complex datasets**
  - **Interdisciplinary** context



- Strong needs of **transparence** and **reproducibility** of data processing

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




What is hard?

- Number and diversity of data sources are increasing
- Stronger complexity of data pipelines

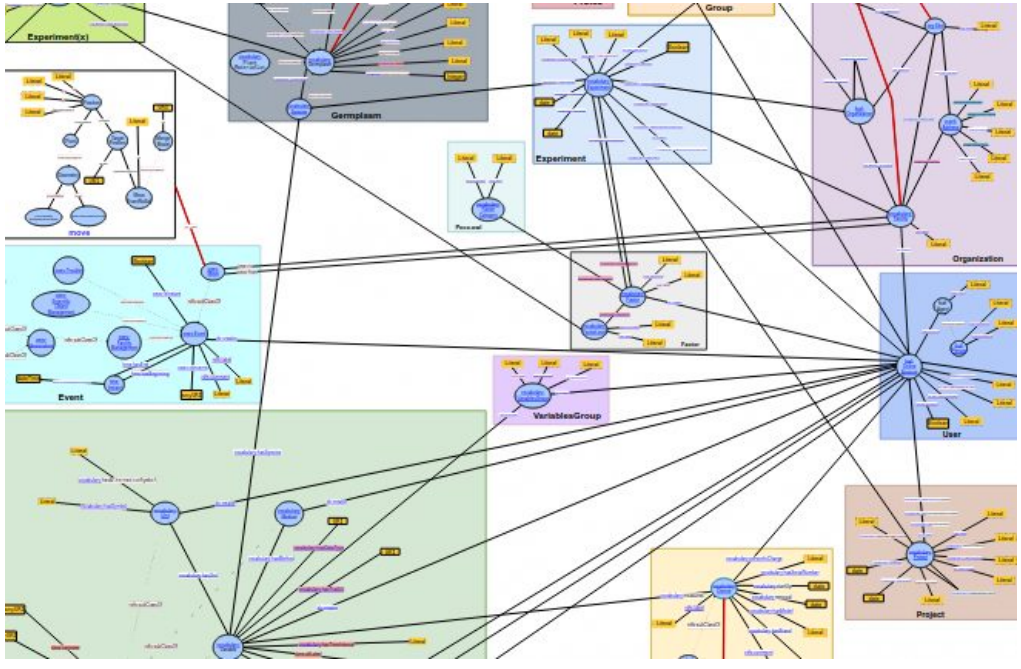
- Increasing the difficulties for a reproducible and open science
- Make harder to understand how and under what conditions the data were produced
- Data integration and data analytics may be impossible

Machine and human need to know about of data production!

Global Approach for Data Management

Producing FAIR data and avoiding data silos

- **Structuring of data**
- **Use standards**
- **Cloud computing and distributed systems**
- **Linking Data** (machine readable) using ontologies: allows to build data sets
  → data integration, knowledge discovery, offer services, validate results, prescriptive analytics, etc.

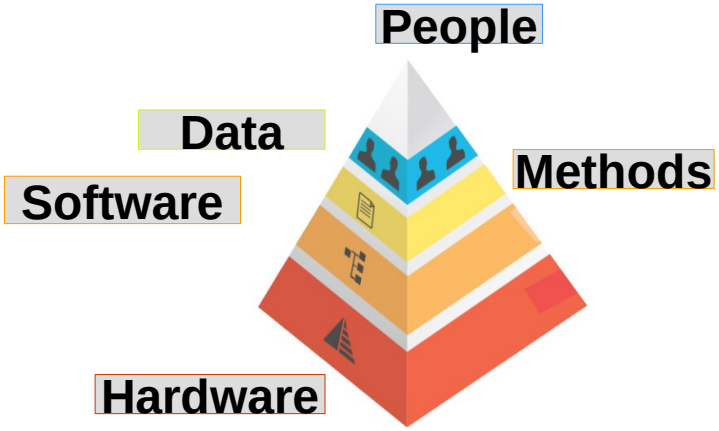


PHIS is based on OpenSILEX

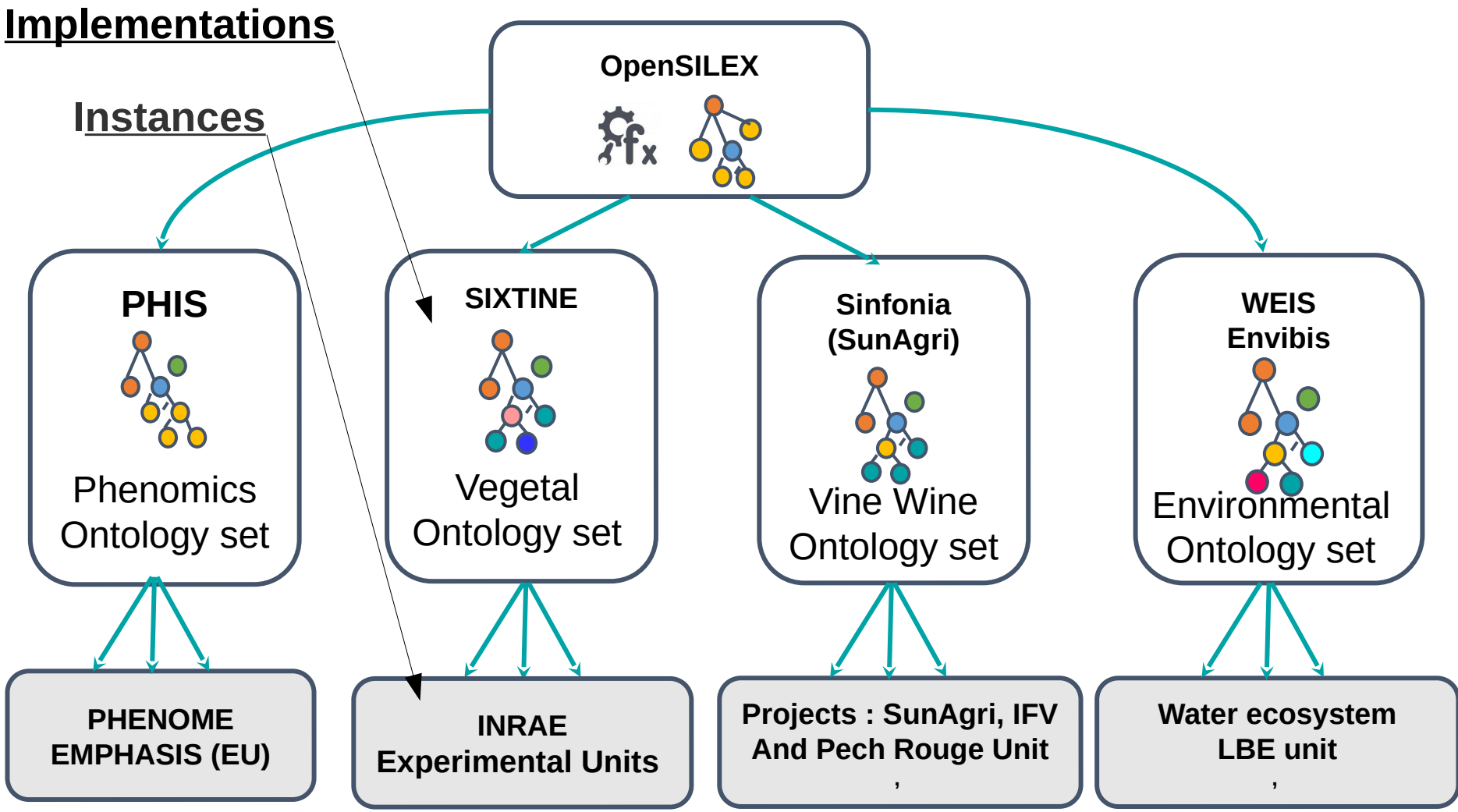
PIPA, PhenomIS and PHIS are Information Systems of Emphasis

OpenSILEX an Open Source software set

- Methods and components to implement information systems for experimental data in agriculture and environment
- Tools and services for the collection, organisation, storage, **exchange, explore and treatment of information**
- **Various communities** → Ontology driven



OpenSILEX based Scientific Information Systems



Structuring of Data

Structuring data to implement good practices:

- **Make it FAIR**
- **Avoid data silos**
- **Have flexible and scalable systems**
- **Integrate Data, combine data**
- **Apply a DMP**

Two key elements

Identification	Semantics
	

OpenSILEX: Structuring of Data

- **Standardized and unambiguous Identification** of entities:
  - Studying objects (plants, plots, canopy, germplams, etc)
  - Experimental organizations (projects, experiments, studies)
  - Experimental resources (devices, facilities, vectors, etc)
  - Events (management, faults, meteo, etc.)

- **Semantics** (based on ontology set) provide:
  - Data understanding
  - Schemas for data
  - Controlled an standardized vocabulary
  - Knowledge representation models with formalized relationships between entities (→ reasoning)
  - Data annotation and enrichment (e.g. search engine friendly)
  - A frame for reproducible data processing

Data Structuring: approach

OpenSILEX → Ontology driven Information System

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

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>

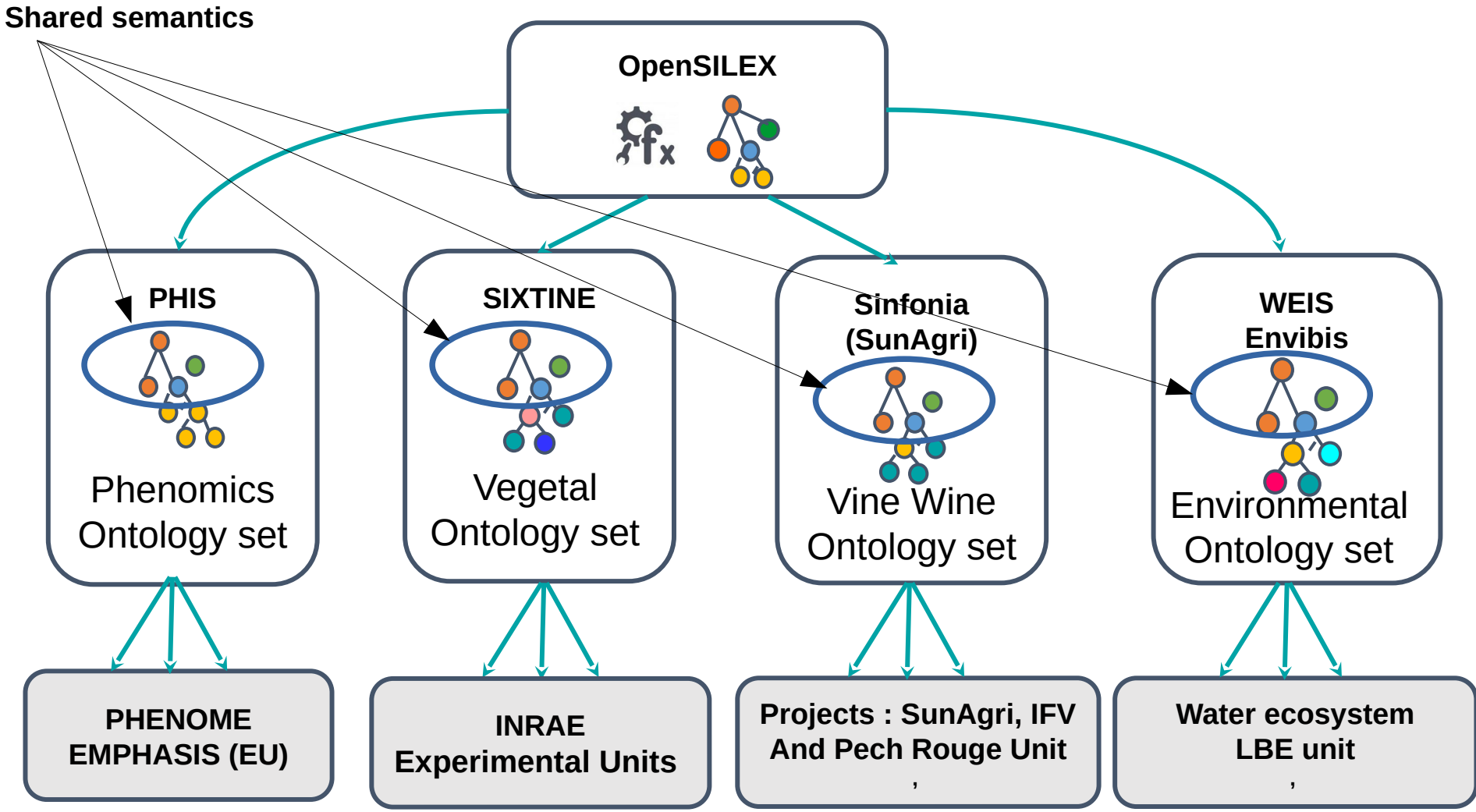
URI of cart:
<http://phenome.fr/arch/2013/ct1300123>

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>

OpenSILEX based Scientific Information Systems



Ontology driven Information System

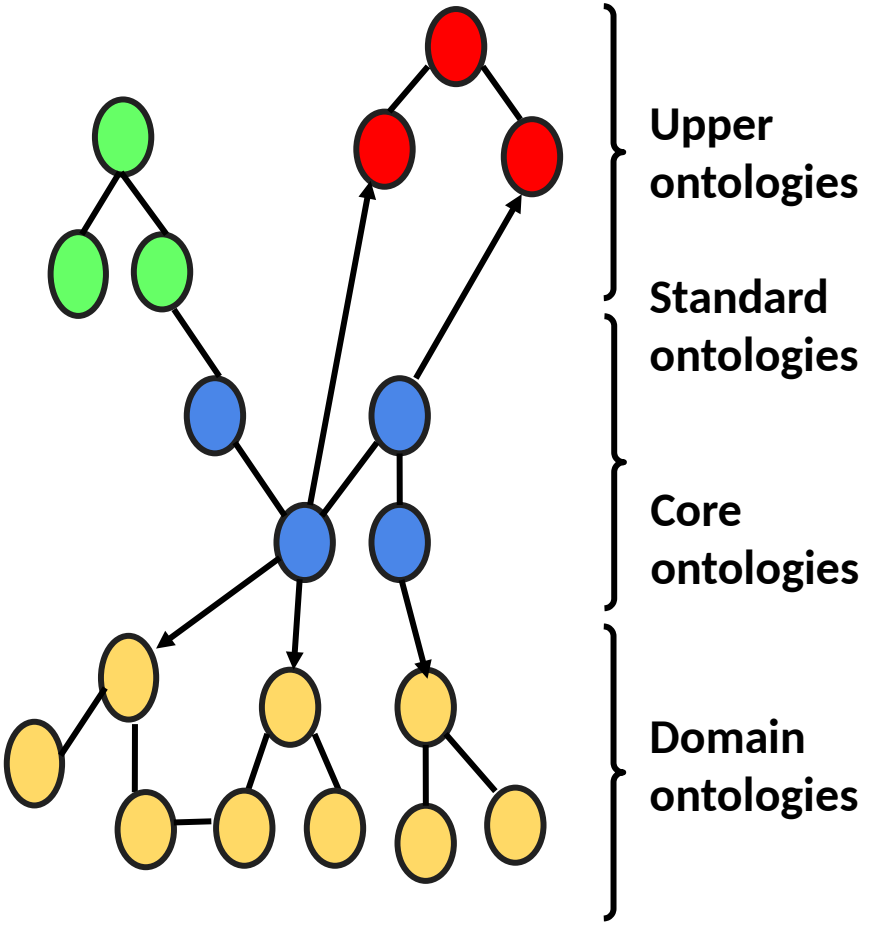
Set of ontologies

Upper ontologies: Dolce & BFO (used as a basis for conception)

Standard ontologies: time, OA, DC, FOAF, PROV-O, SOSA, etc

Core ontologies (OESO & OEEV): main concepts of OpenSILEX

Domain application ontologies: specific to a domain or a community

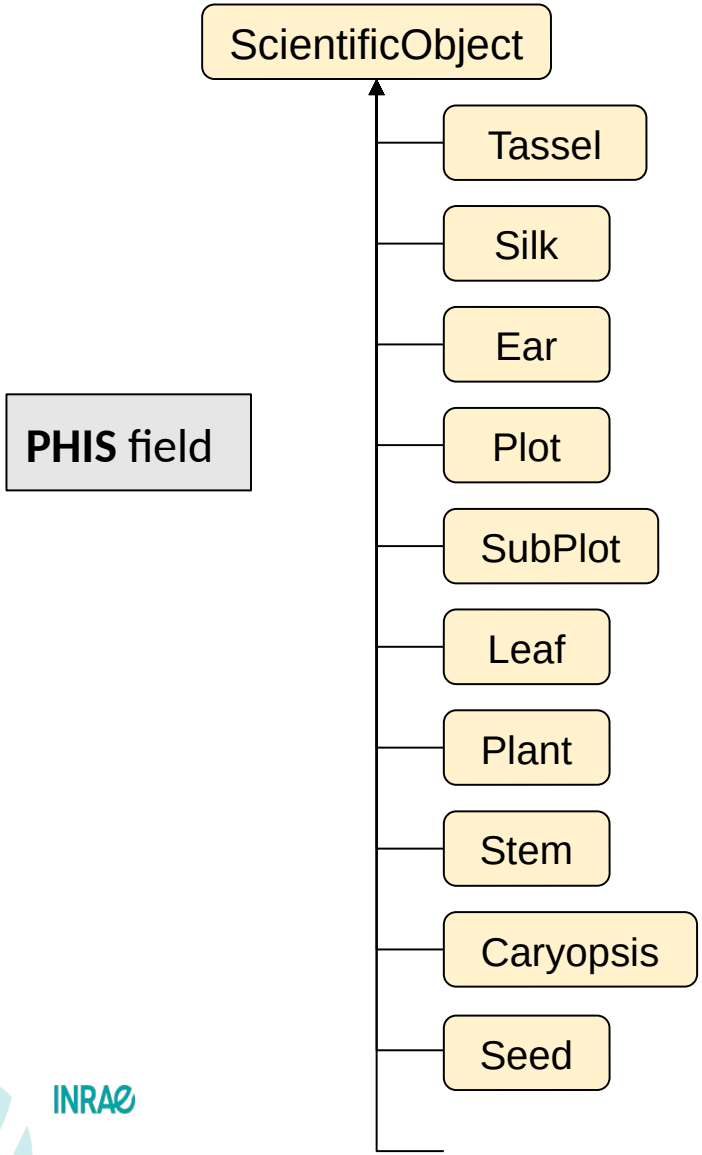


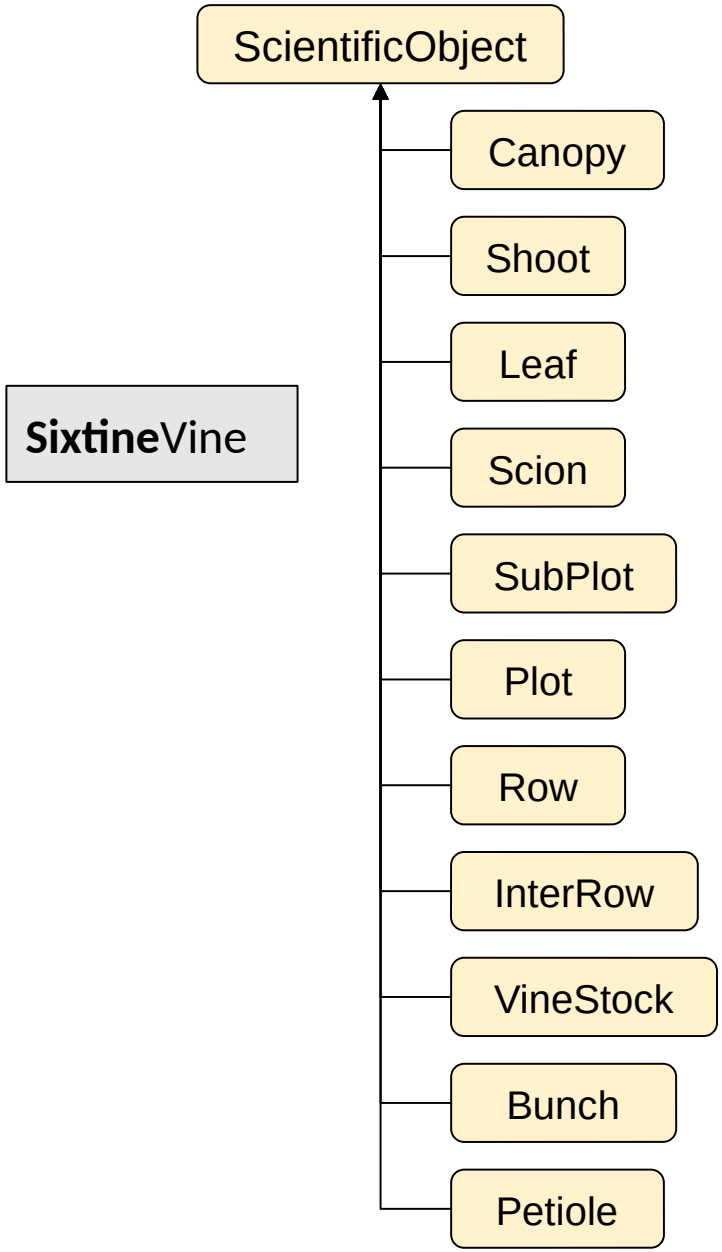
Definitions of Core Concepts

Scientific objects: *Elements (plants, plots, etc.) characterized and observed individually within an experimental framework that enable to verify a hypothesis or to be better understood a phenomenon; in particular by varying factors (situational parameters such as treatment or temperature) associated with the scientific objects in order to observe the effects induced by these changes.*

Events: *Events can be processes, actions or facts that occur (or even precede) scientific experiments and have an influence on the experiment. Events are identified and characterized. Main category are controlled (irrigation, fertilizer) or uncontrolled (hail, frost, pests).*

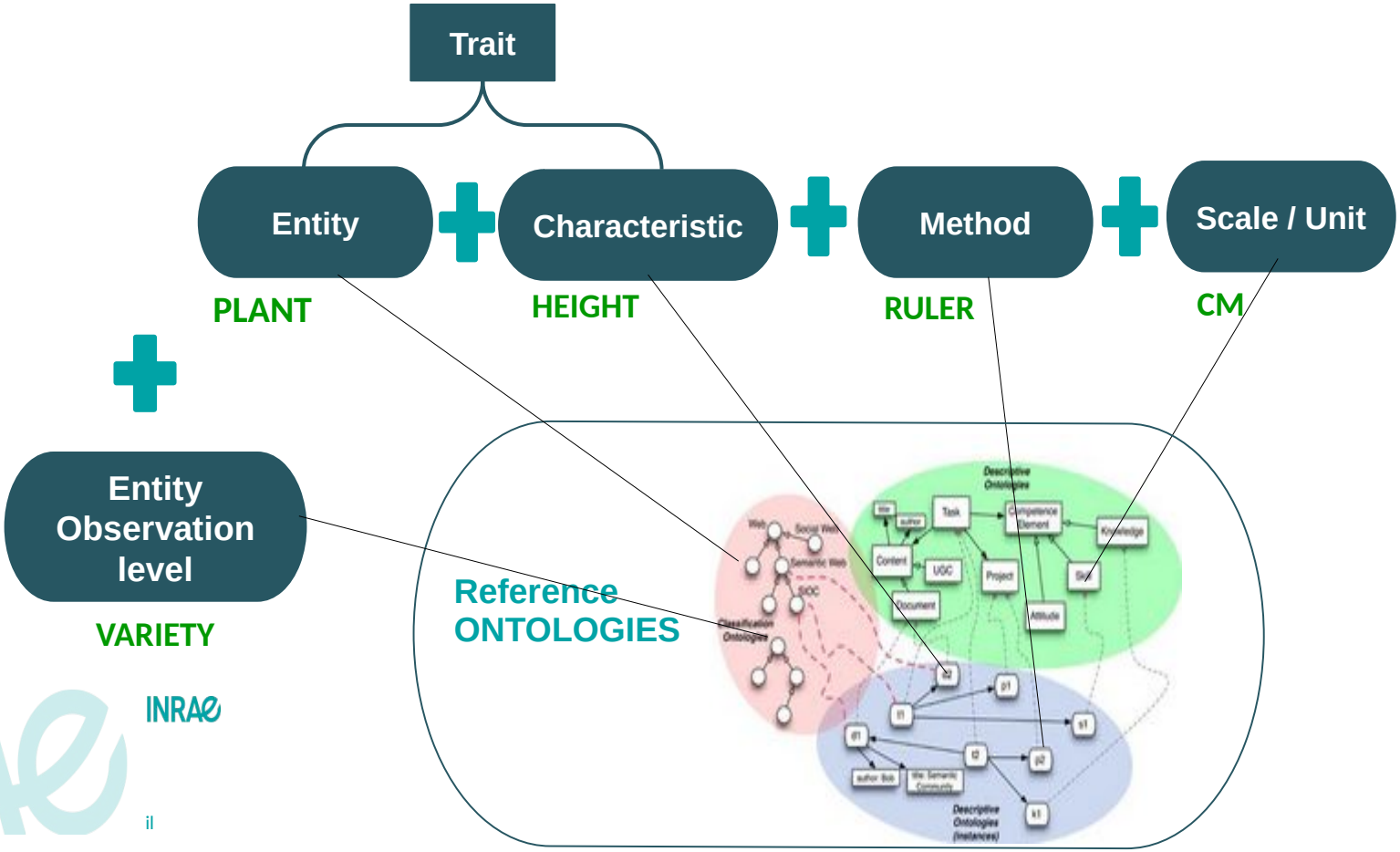
Example: PHIS and SixtineVine





Variable Representation Model

- Enabling semantically precise descriptions
- Decomposing description into standardized elements
- Link to existing vocabularies/ontologies
- Make description machine readable
- Strongly improve data integration process



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

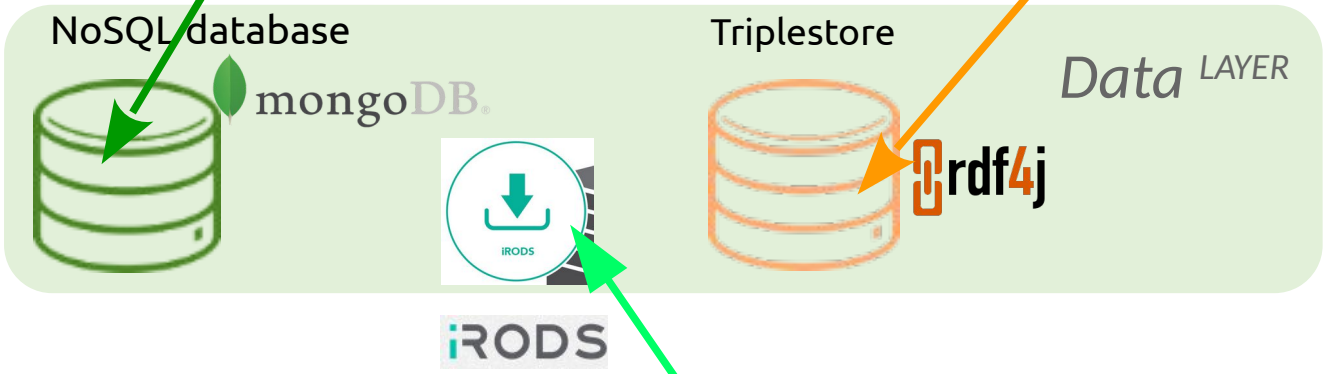
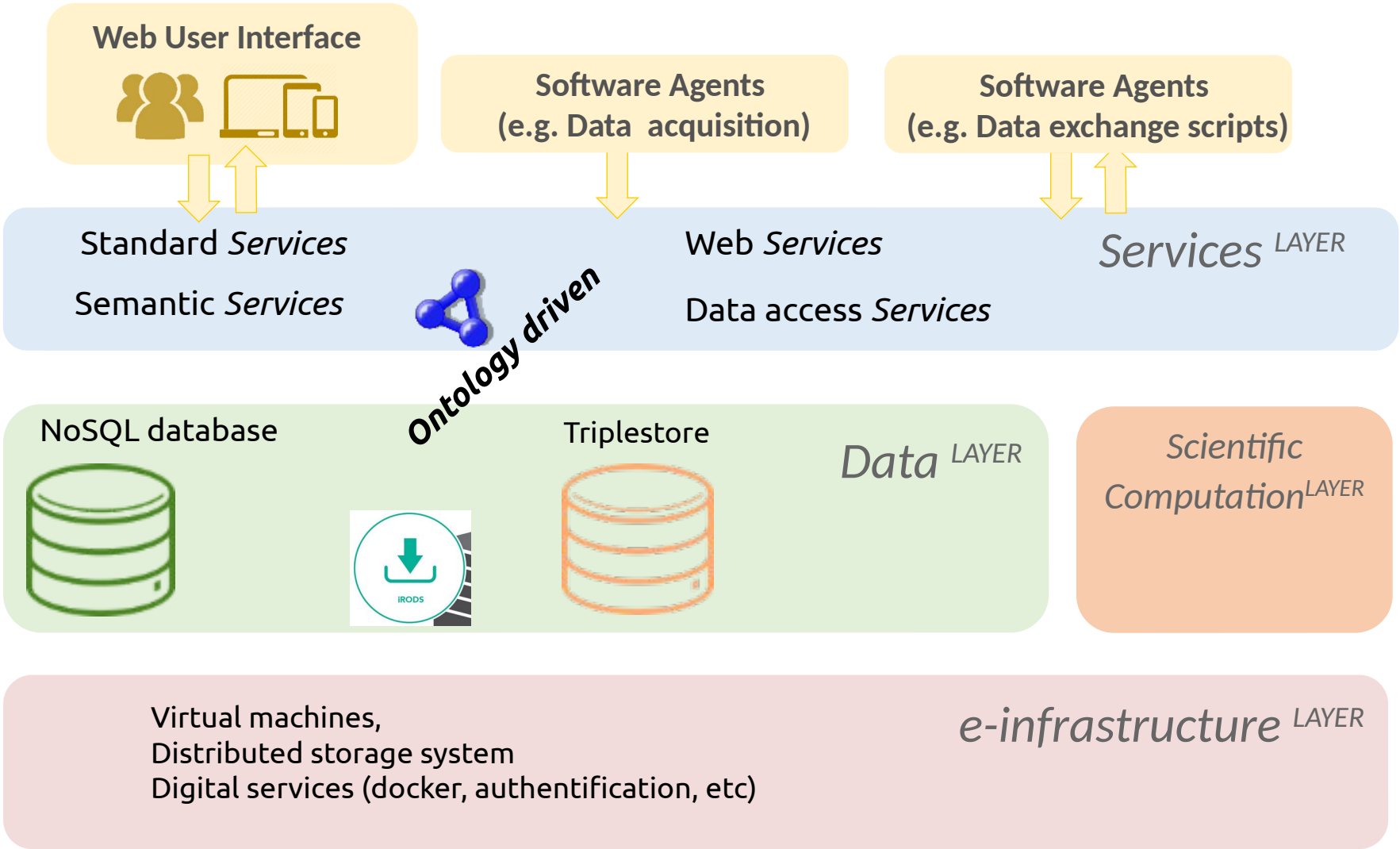
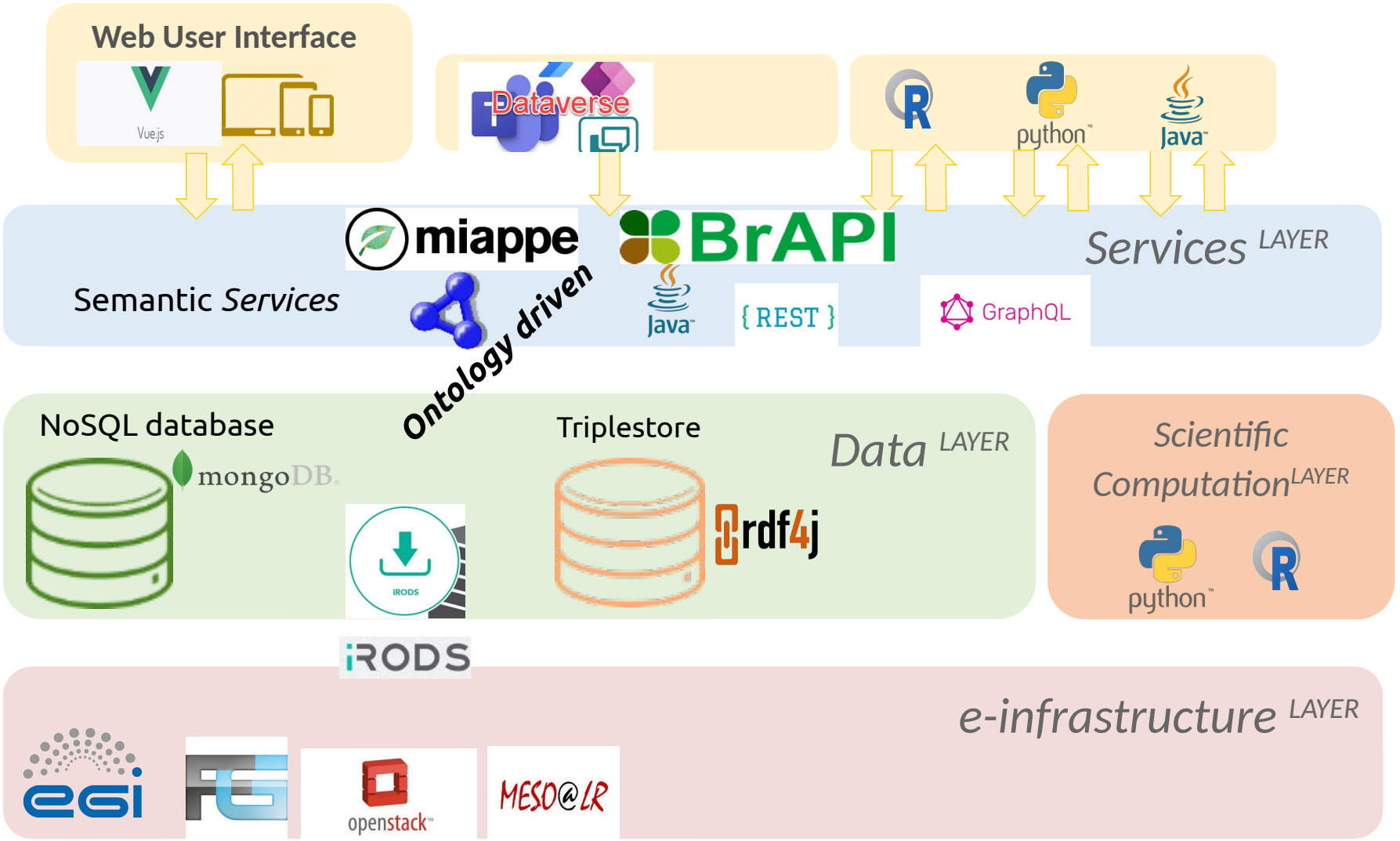


Image sets are stored on IRODS data system

OpenSILEX Architecture



PHIS Specific Architecture



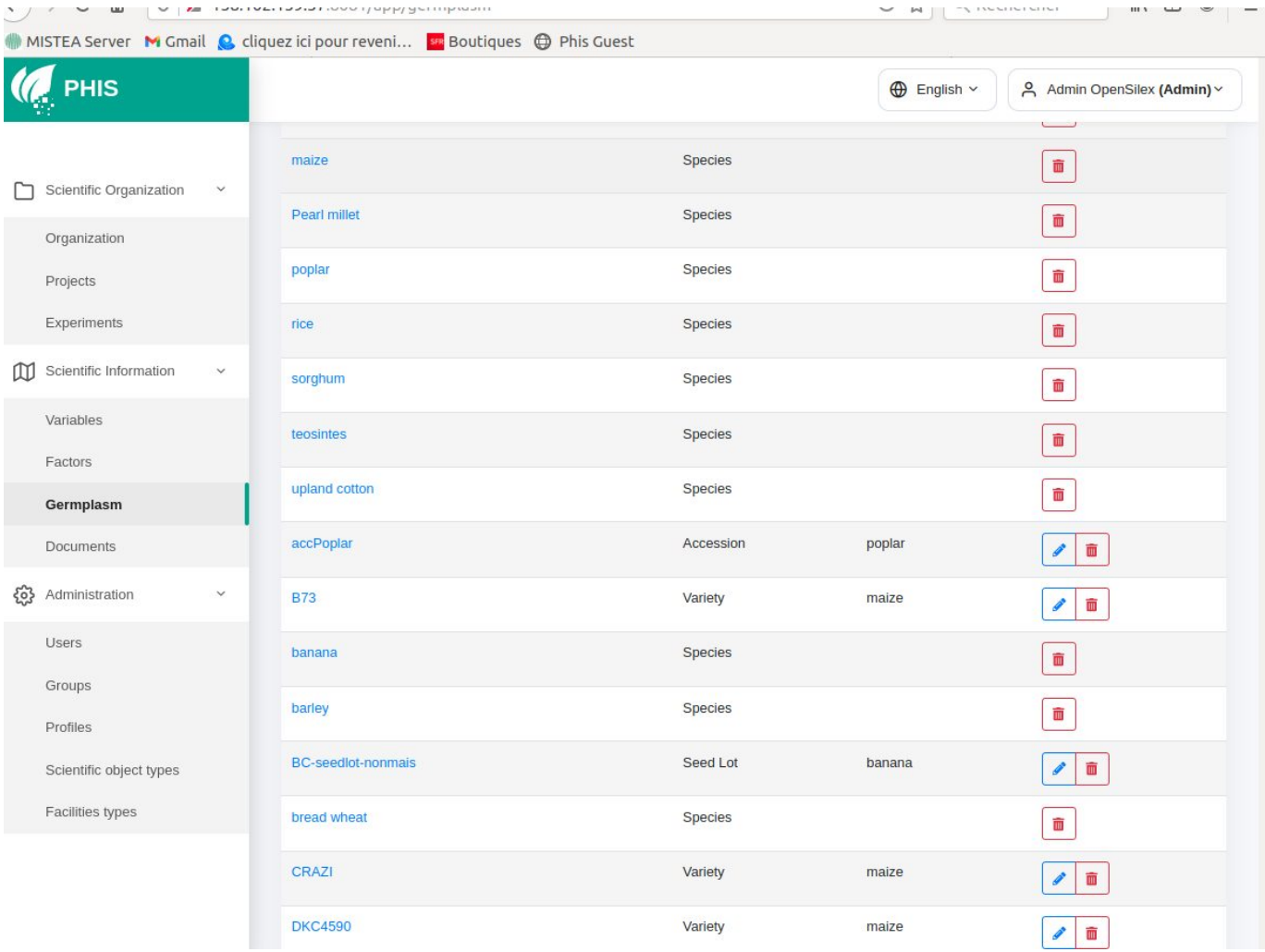
PHIS User Interface

PHIS Web Interfaces for the management of:

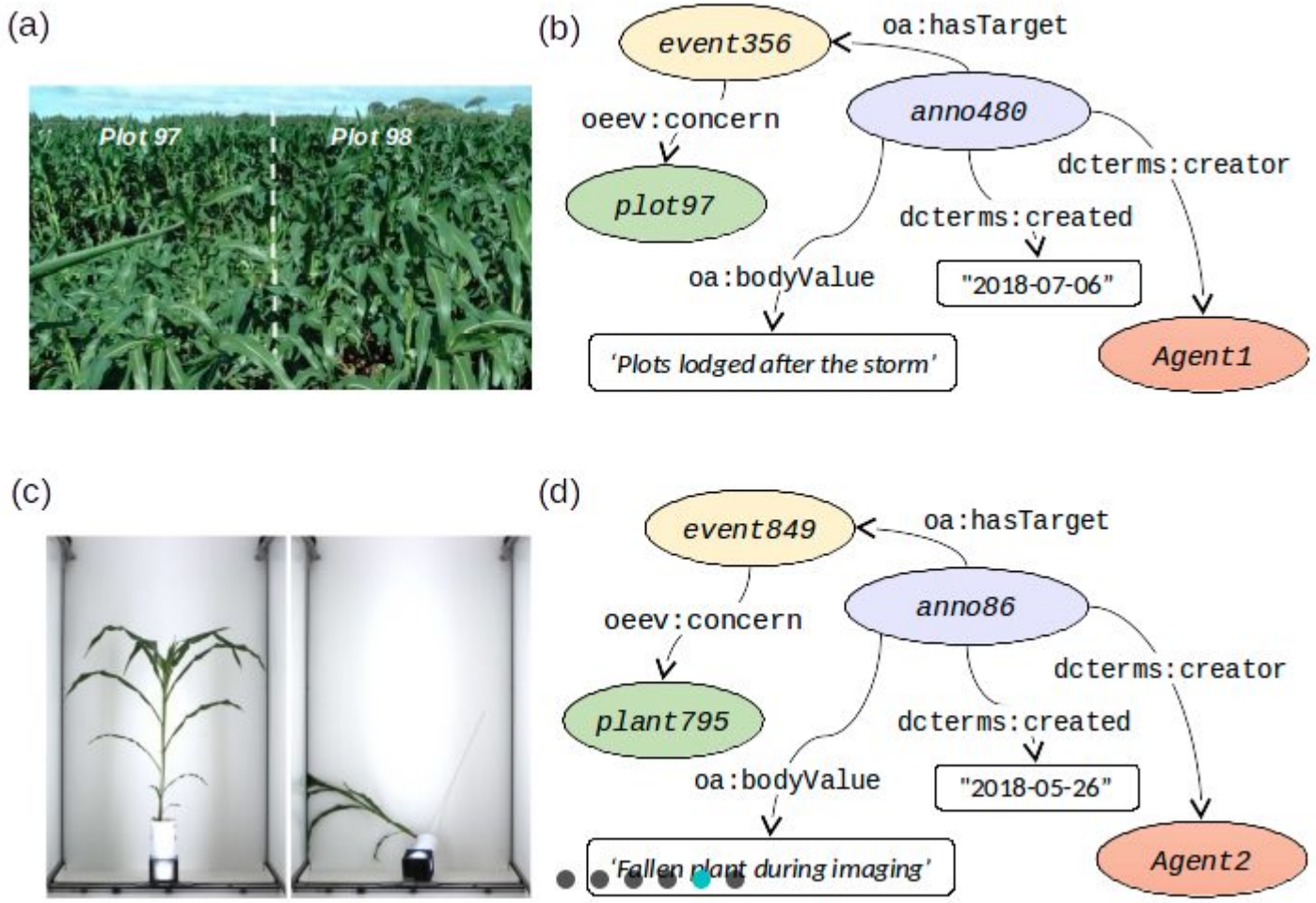
- Project information
- Experiment
- Facilities
- Devices

- Scientific objects
- Germplasms
- Experimental factors

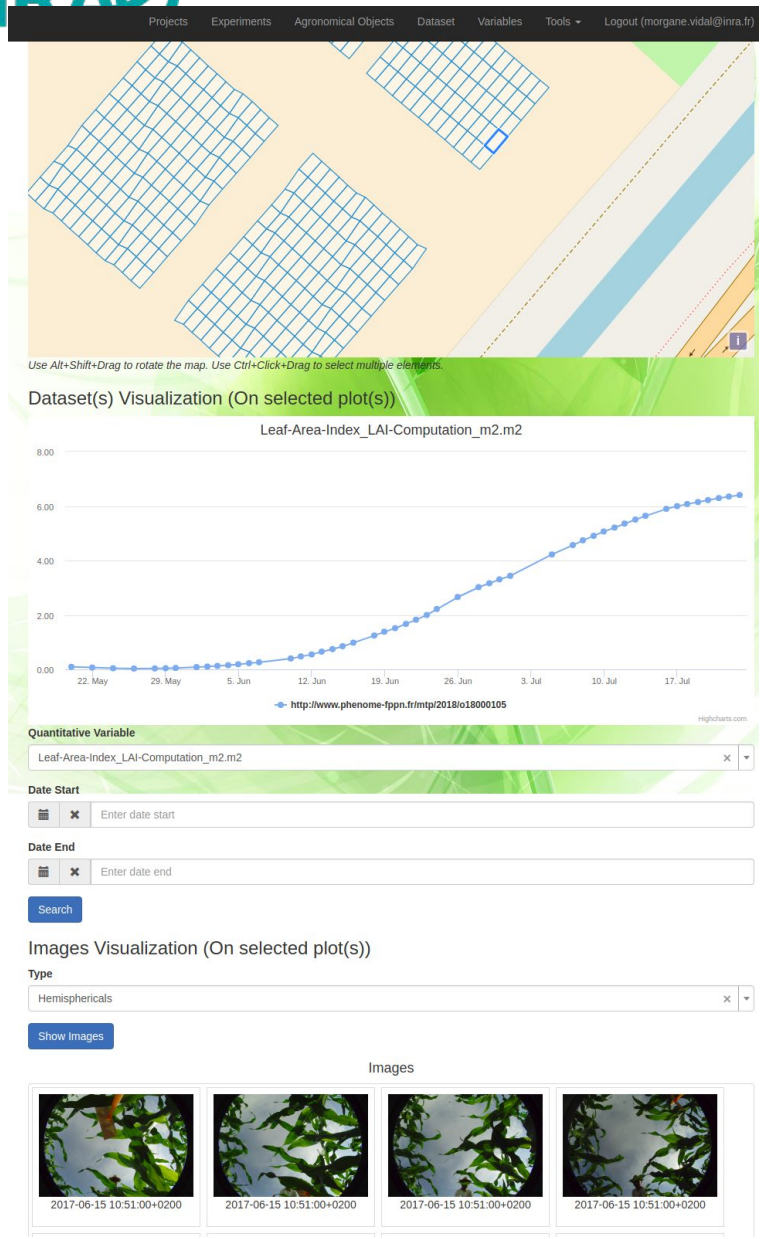
- Data
- Data visualization
- Data provenance

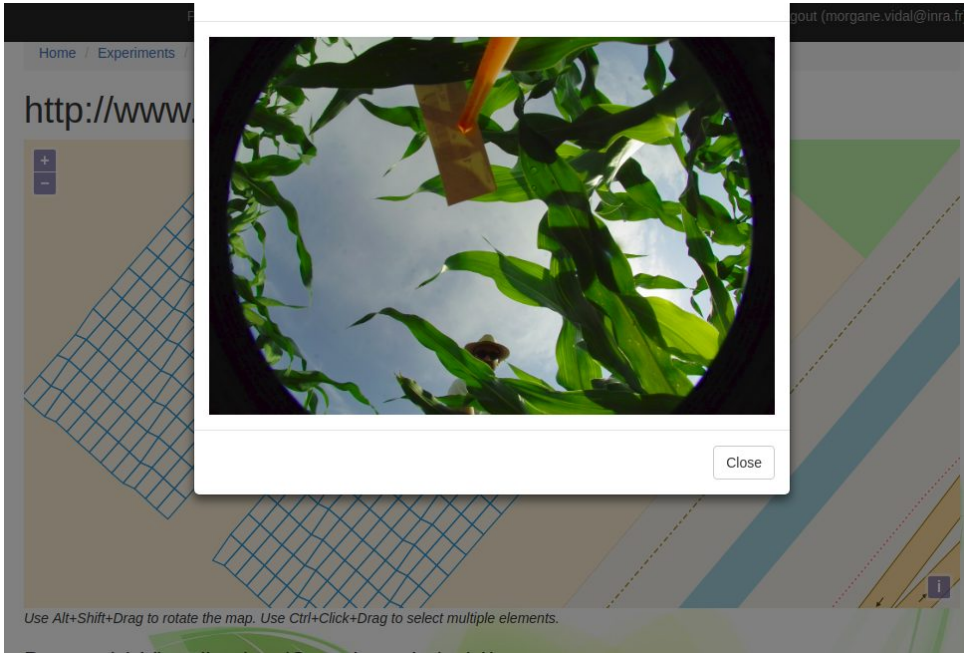


Ontology driven Information System



Trait – Provenance





In short OpenSILEX

- ✔ **Allows management of huge and complex data (petabytes)**
- ✔ **Enables and facilitates cloud computing (data center, EGI)**
- ✔ **Manages semantics (ontologies, standardized vocabularies)**
- ✔ **Provides a flexible design**
- ✔ **Provides provenance and reproducibility for data processing**
- ✔ **Different Implementations:**

  PHIS, Sixtine, Simphonia, 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 for 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

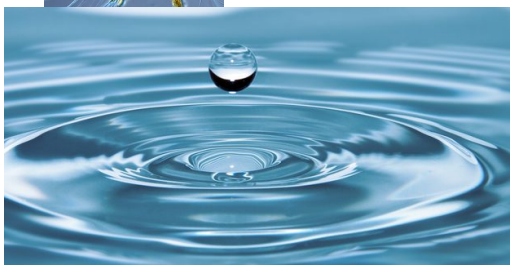
Complex Data

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

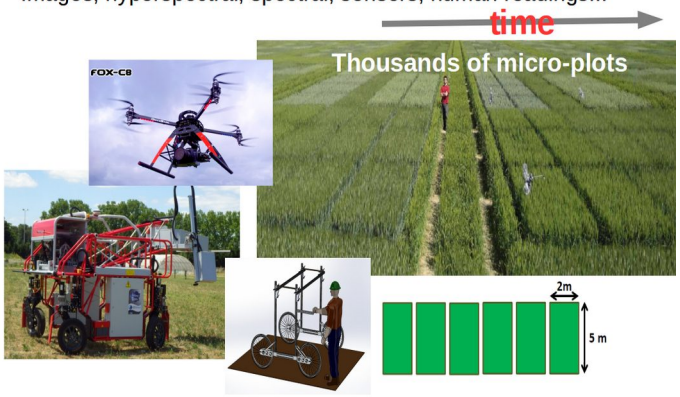
Composition and the structure of biopolymers

Quantification of metabolites and enzyme activities

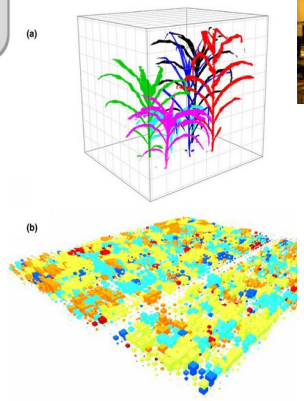
Field Platforms

Various scales and data types

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

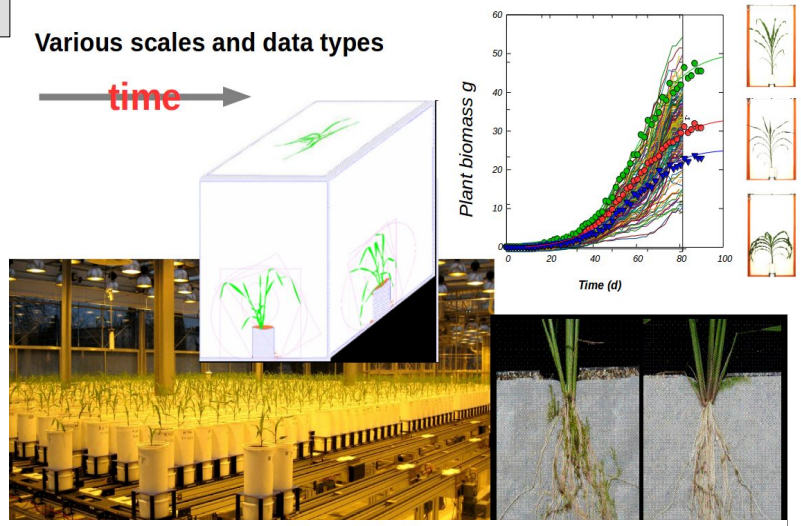


Virtual experiments



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...