

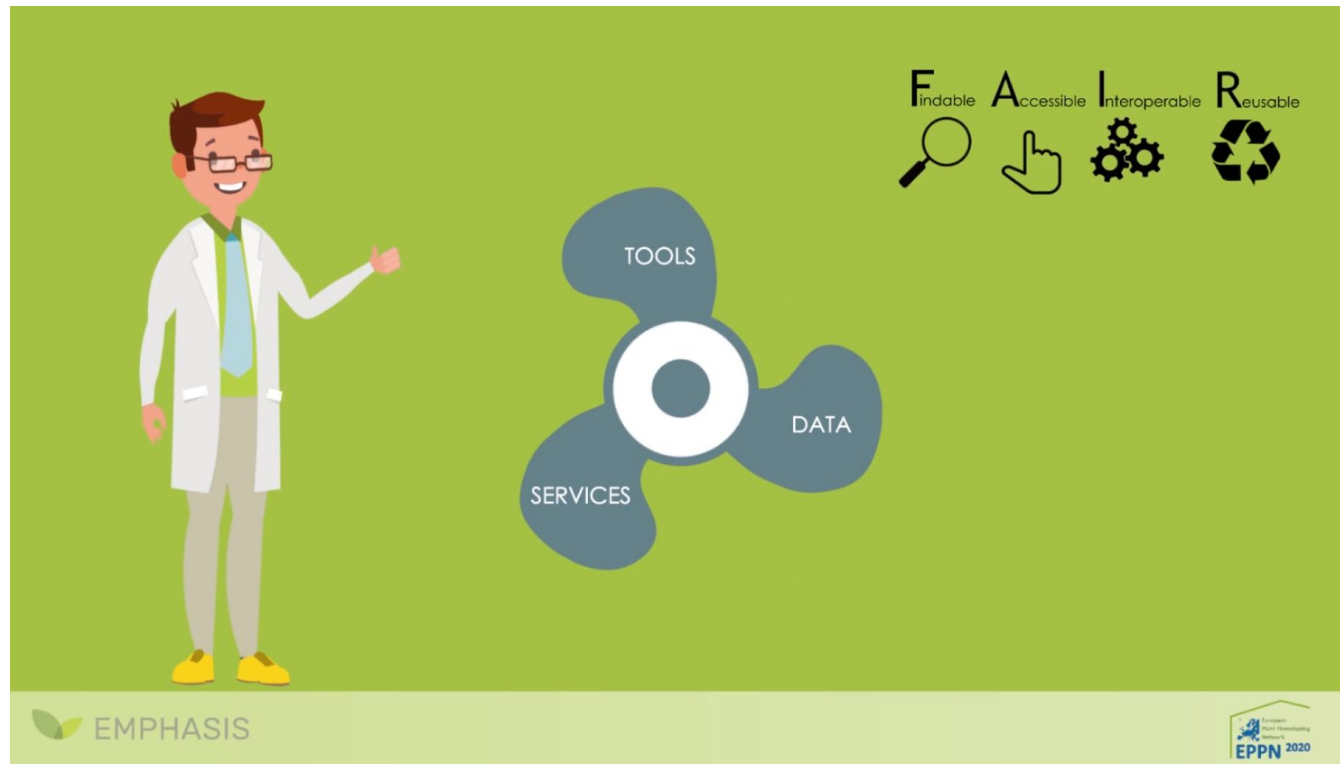
Exchange event between the European
phenomic community and industry

Towards implementation of **FAIR principles** for **Plant Phenotyping**

Pascal Neveu/Isabelle Alic (INRAE, FR)

Let's watch a short video...

<https://www.youtube.com/watch?v=rXKBy7DOPrY>



- **Plant Phenomics Experimentations**

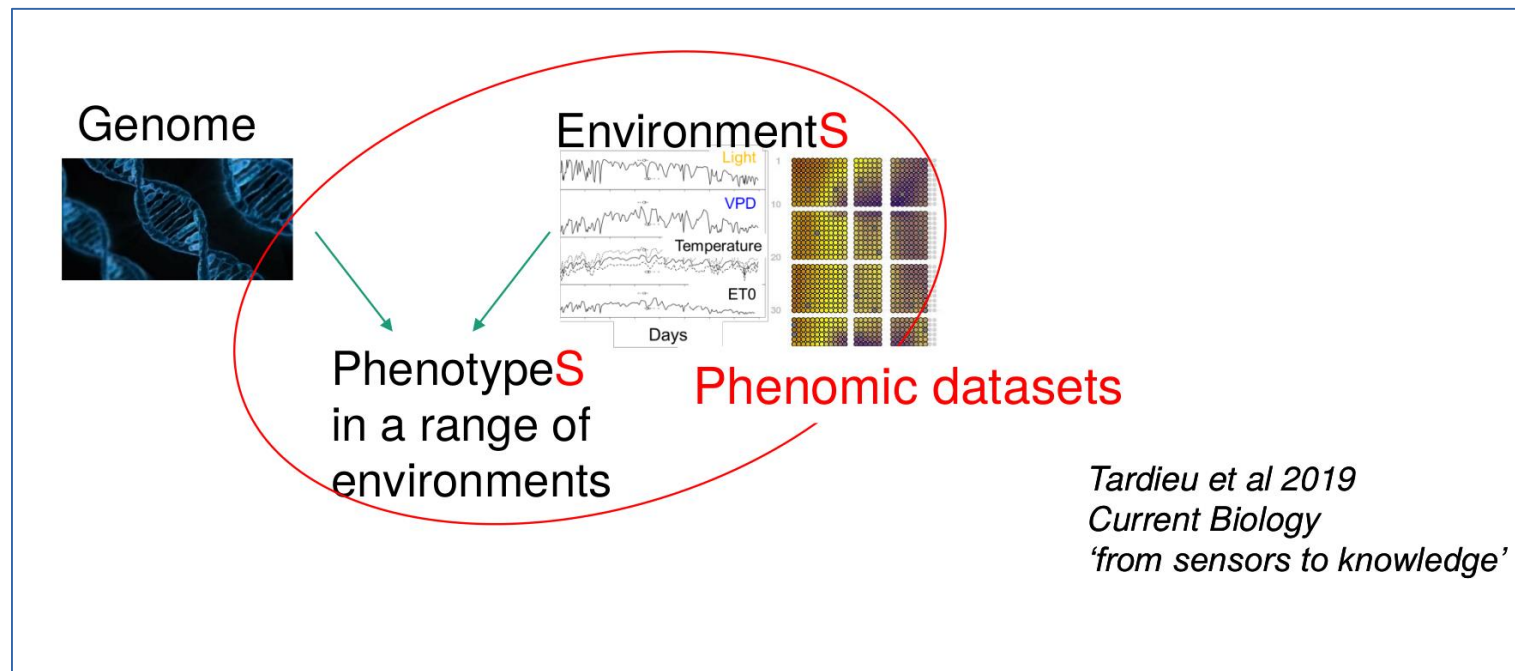
- Expensive, require a lot of resources and often very hard
- Cannot be reproduced
- Huge and complex datasets
- Strong needs of transparency: reproducibility for data analytics

- **Save time, make data valuable!**

But re-analyses, meta-analyses and new analyses
→ impossible without rich metadata

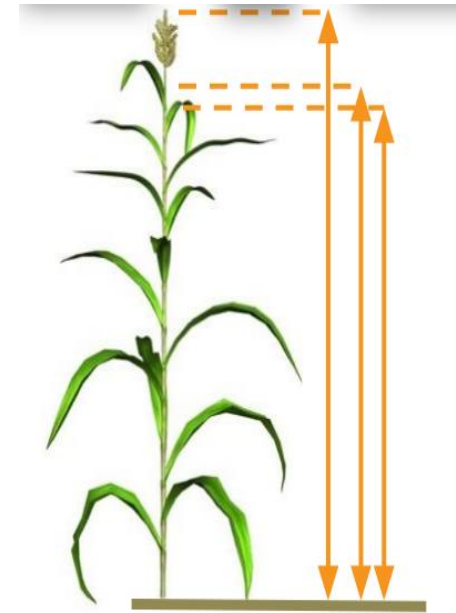
- **EMPHASIS Information systems**

- **Environmental conditions** are PART of the phenomic information
- We need the **time courses of environmental variables**, at the **exact time** (minutes) when data were collected
- And the **precise x-y-z position** of sensors of measured plants/plots



Some common mistakes we do

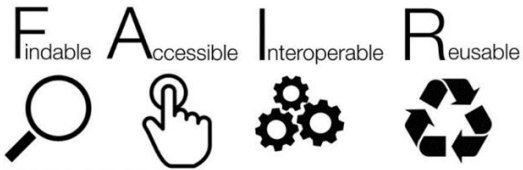
- **Metadata in file names** (low and very poor metadata)
- **Ambiguous ID**
- **Variable naming**
same name for several variables, not well defined, no ID, no schema,...
- **Data are stored on personal computer**
- **Unstable data files** (non-automatisable processing)
- **Context, faults are not described**
- **No data links**
- **Missing data representation**
- **No licence**



Plot566
in 2016

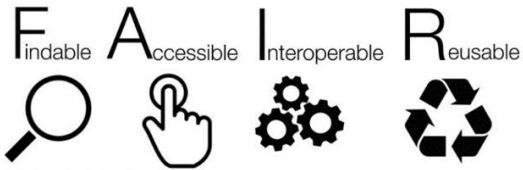
Plot566
in 2017





Findable: Persistent ID, indexed in portals, standardized and relevant metadata

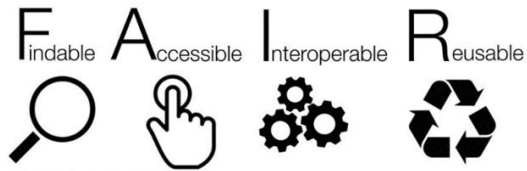




Findable: **PID**, indexed in portals, standardized and relevant metadata

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



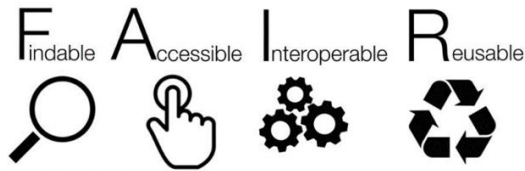


Findable: **PID**, indexed in portals, standardized and relevant metadata

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

Interoperable (technology, syntax, **semantic**): shared standardized formats, vocabularies and **formal languages for knowledge representation**





Findable: **PID**, indexed in portals, standardized and relevant metadata

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

Interoperable (technology, syntax, semantic): shared standardized formats, vocabularies and **formal languages for knowledge representation**

Reusable: provenance, relevant metadata for understanding **across disciplines**



How ?

Based on 2 key elements:

→ Identification and Naming convention

- Objects: plants, plots, experiments, sensors, events, etc
- Persistent, **unambiguous**, resolvable, globally unique

→ Semantic and tagging (based on ontologie set)

- Controlled vocabulary
- Formalized relationships between entities
- Data annotation and enrichment (search engine friendly)

Identification

URI : Uniform Resource Identifier

- **Standardized, Unambiguous, Actionable**
- **Generated by tools** under responsibility of scientific coordinator
- Use URI for every objects

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

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

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

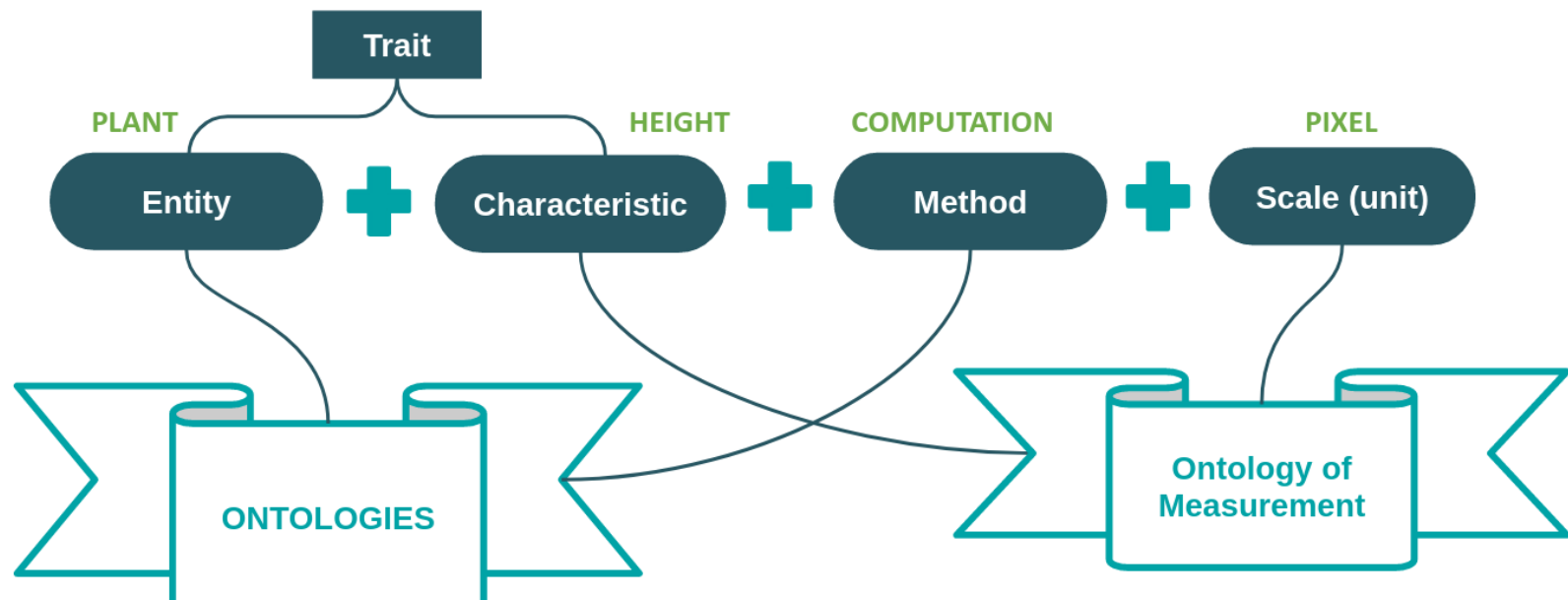
URI of cabin:
<m3p:arch/2018/ac180015>

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



Variable naming

- Use **URI** for unambiguous name (in global context)
- **Reuse existing variable** if available
- Use **standardized/shared representation schema** for formalisation of new variable



Identification


- Local infrastructure use **URIs for all objects**
- Information systems implemented in **nodes**
- Automatic generation of URIs : **ID Generator**

Variable naming


- **Environmental and phenotypic** variables named
- Variable names with **quadruple definition**
Entity – Characteristic – Method – Unit
- Web Interface to **share and declare variables**

Edit variable

URI
os:variable#variable.air_temperature_50cm


Entity  *

Air

Characteristic  *

Temperature

Trait already existing in an ontology

Method  ?

Thermocouple

Unit/Level *


Degree Celsius

Name *

air_temperature50cm_thermocouple_degree_Celsius

Alternative name

air_temperature50cm

Data type  *

Decimal Number

GENERATE NEW URI

ENRICH EXISTING ID

LAST DATABASE URI

LOG OUT

Generate new URI

Provide a csv table of your data. Expected format is the [following](#) .
Connected as test

1

Import your file
A file with one row for each resource you want to generate a URI for.
Parcourir... Aucun fichier sélectionné.
Details: v

2

Host Name
test
Installation name
your installation
Object Type
plant

3

Data to put in the URI
Project related
aProject
Year
2020
Generate URI
You can find the file in your Download folder (default settings)
Generate QR-Codes ?

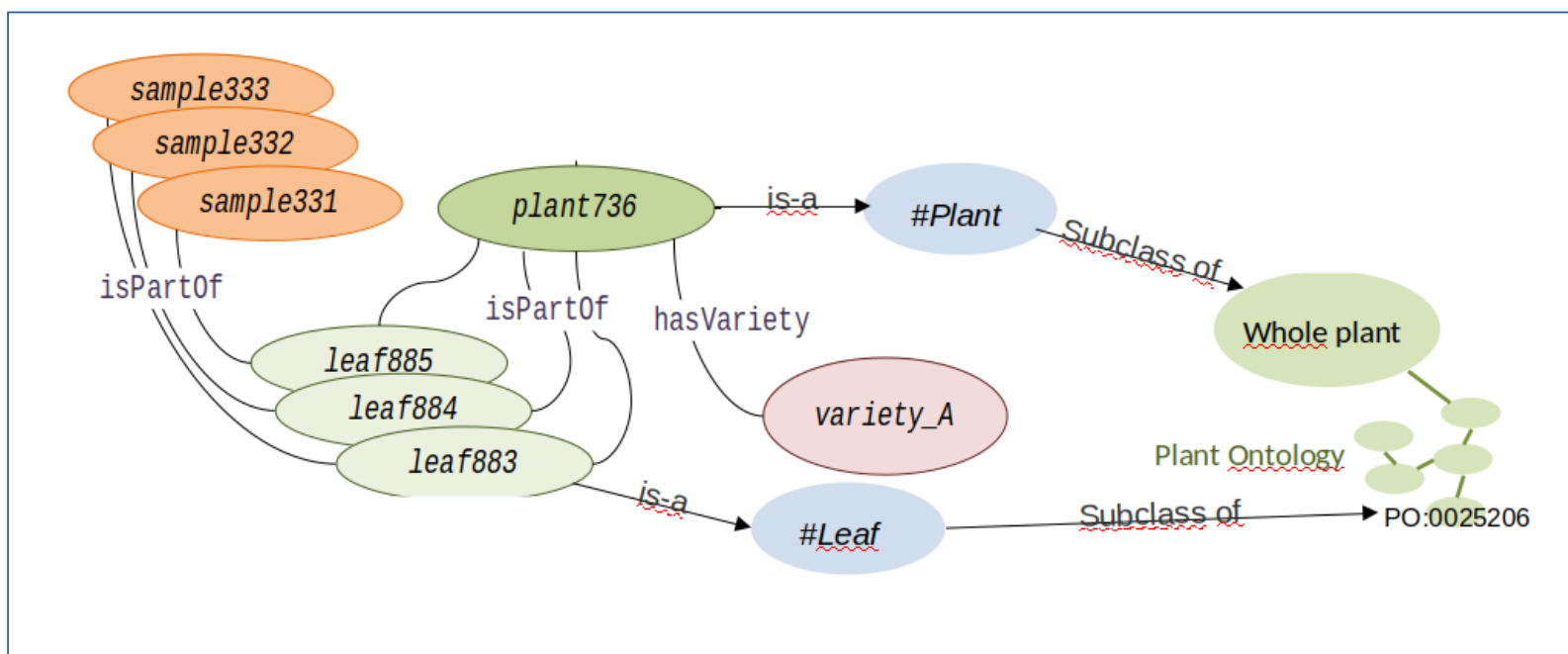
<http://138.102.159.36:8082/app/>

- 19 March 2021

12

Metadata / ontologies provide the meaning of data:

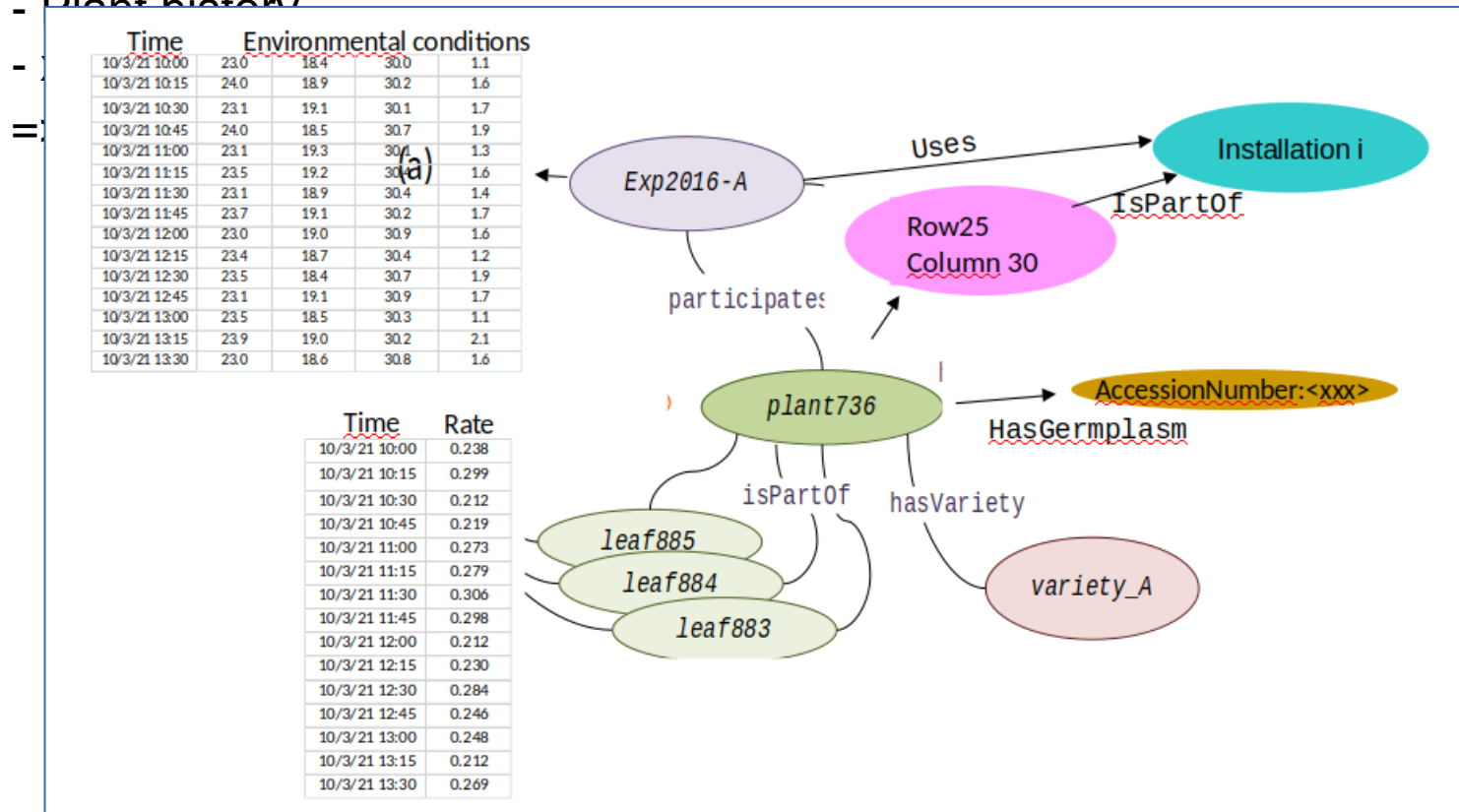
- Link each data element with a controlled, shared and machine readable vocabulary
- Structure the data in a graph
- **Reference ontologies** (*Agrovoc, Plant Ontology, PATO, etc.*)
- **Application ontologies** (*scientific objects, events, ...*)

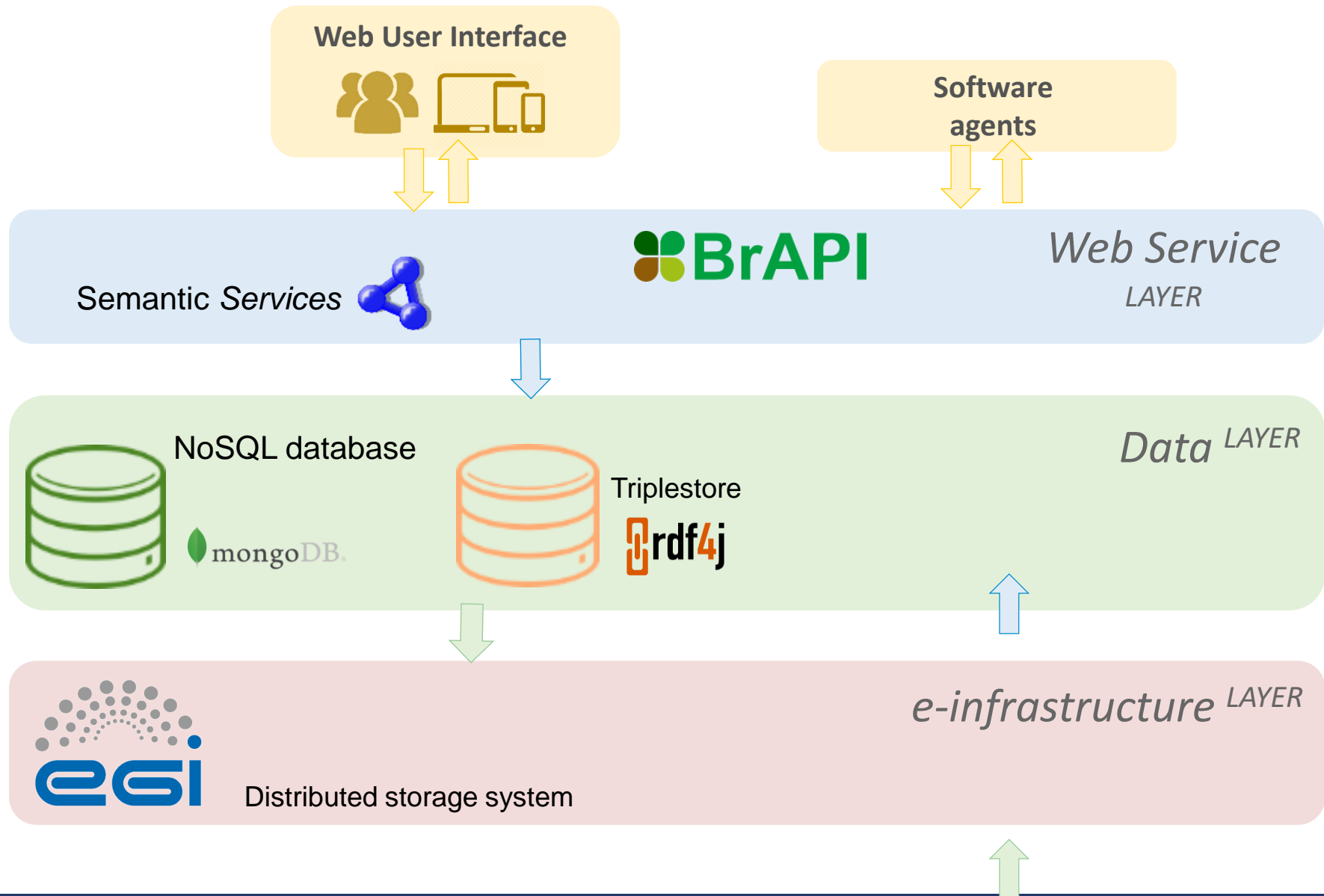


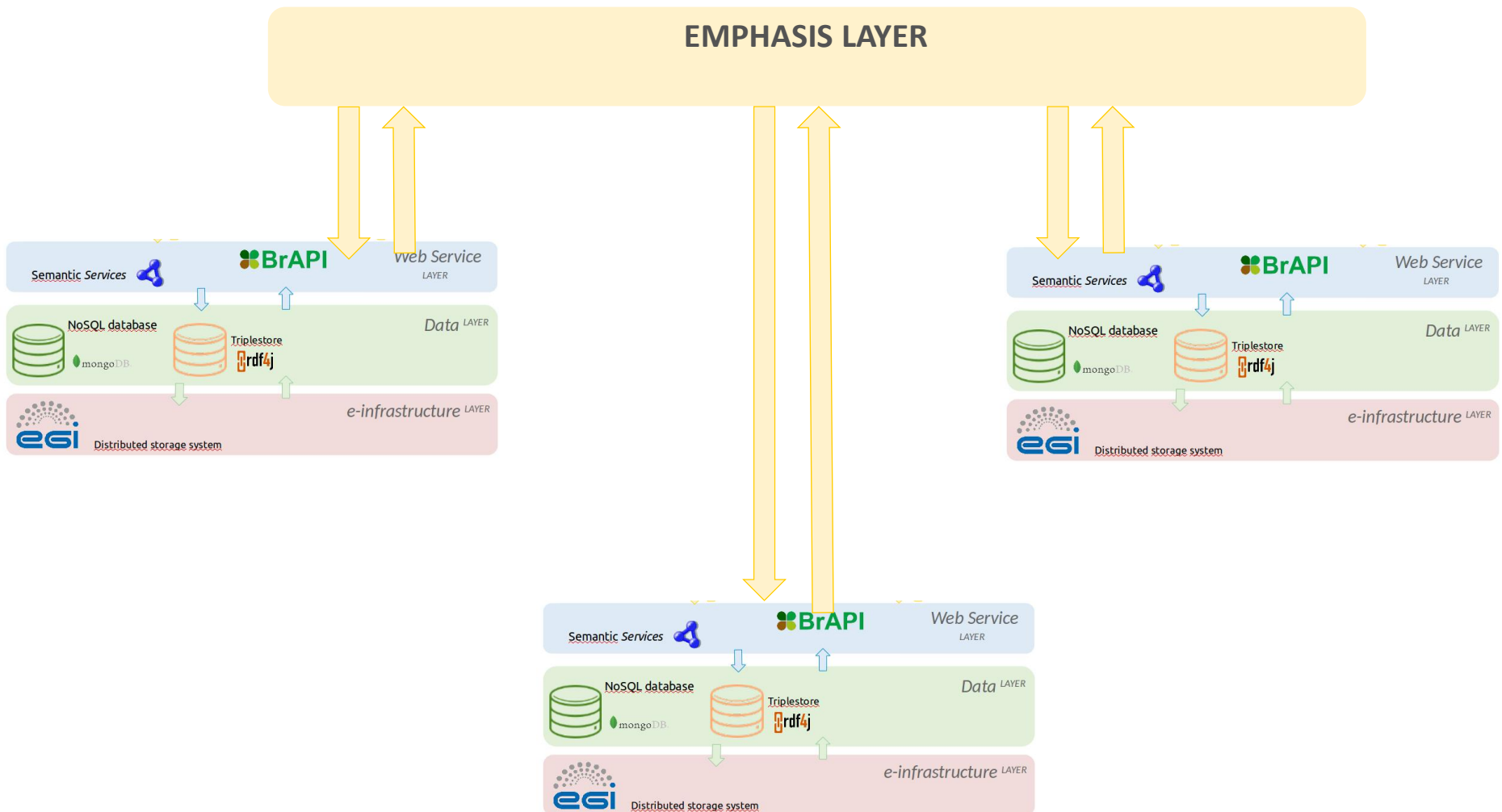
Data are structured by a set of ontologies

What is kept in **EMPHASIS-compatible** information systems (e.g. PHIS) :

- Genotype,
- Accession,
- Sampling time,
- Plant history,







PHIS Interfaces allow management of

- **Scientific Organization**

- Project information
- Experiment
- Facilities
- Sensors

- **Scientific Information**

- Variables
- Scientific objects
- Germplasm
- Experimental factors

- **Data**

- Data visualization
- Data provenance

The screenshot displays the PHIS web interface. On the left is a sidebar menu with categories: Scientific Organization, Scientific Information, Data, and Administration. The 'Devices' option under Scientific Information is highlighted. The main content area shows a form to 'Add device' with fields for Name, Type (set to SensingDevice), and Constructor model. Below this is a table of existing devices. A modal window is open for editing a device named 'udpate', showing fields for URI, Type (Humidity Sensor), Name (HS_02), Brand (Campbell), Constructor model (CS655), and Serial number.

Select	Name
<input type="checkbox"/>	HS_01_mod
<input type="checkbox"/>	HS_02
<input type="checkbox"/>	HS_05
<input type="checkbox"/>	MSC_03

- ✓ **Data management needs time and work** at first but...
 - Will **save you time** !
 - Allows you to handle **large and heterogenous data sets**
 - Allows you to **do analyses previously impossible**
 - Allows teams and communities a **better formalization of concepts and data**

- ✓ **FAIR data** requires **training and support**