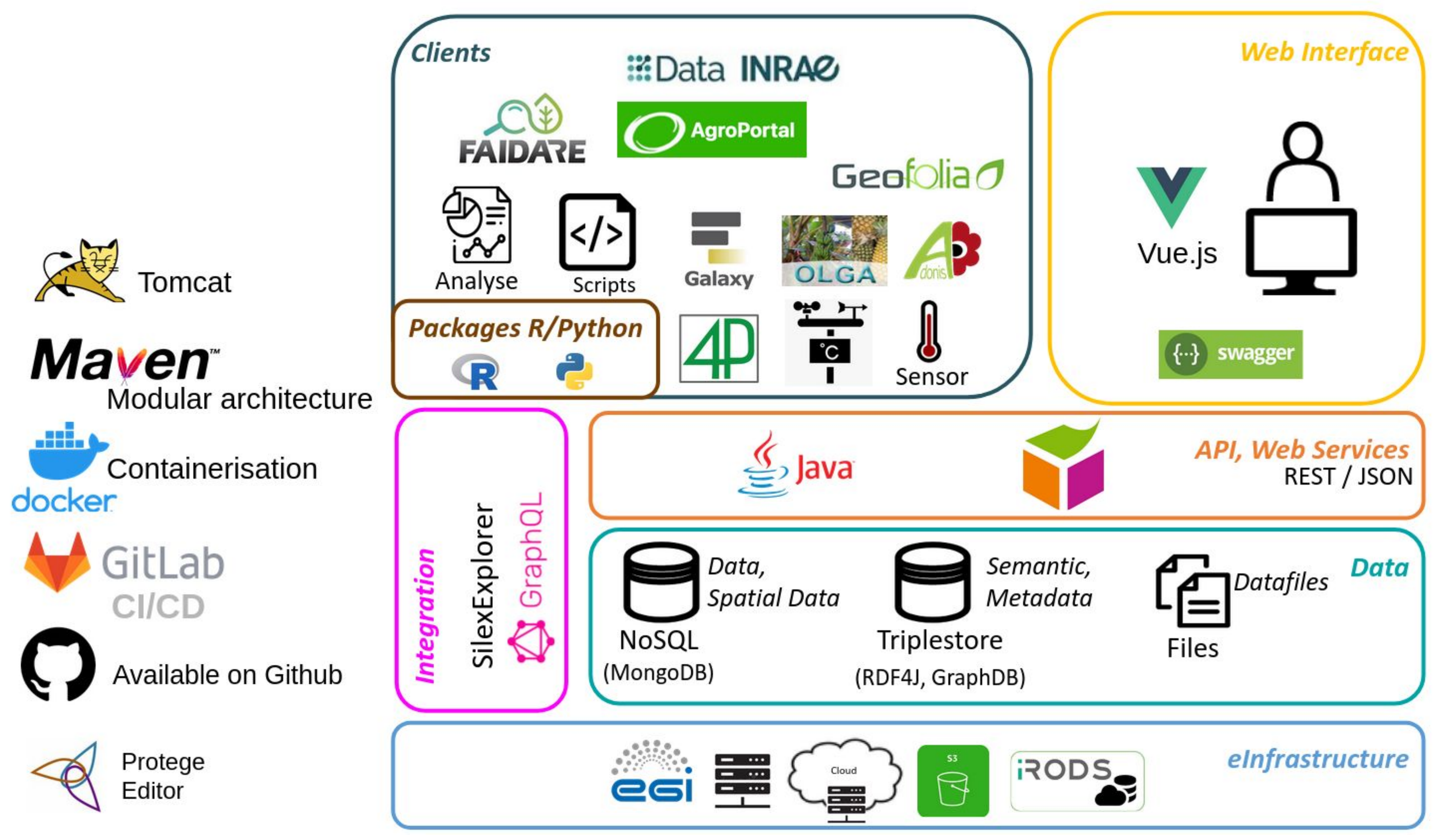
OpenSILEX software suite

Combining knowledge graphs and NoSQL databases to build scientific communities information systems

Isabelle Alic [1], Arnaud Charleroy [1], Renaud Colin [1], Farzaneh Kazemipour-Ricci [1], Pascal Neveu [1], Catherine Roussey [1], Anne Tireau [1]

1 INRAE- UMR MISTEA - 2 Place Viala, Montpellier- France



1- Hybrid Information System: ontologies with large volumes of data

- Represents all descriptive data as a data graph based on Semantic Web technologies
- Suited for experimental contexts wich are various, complex and constantly evolving.
- Uses ontology for flexible data integration

BUT:

- Triple stores are no adequate to manage big data storage
- Storage of data files in RDF format is not usable

COMBINING:

- NoSQL manages the large volumes of data (time series of measurements) in a flexible way
- NoSQL data management systems are more efficient for most queries than a triple store
- Datafiles are stored on high-performance distributed file system solutions (Grids & Clouds)

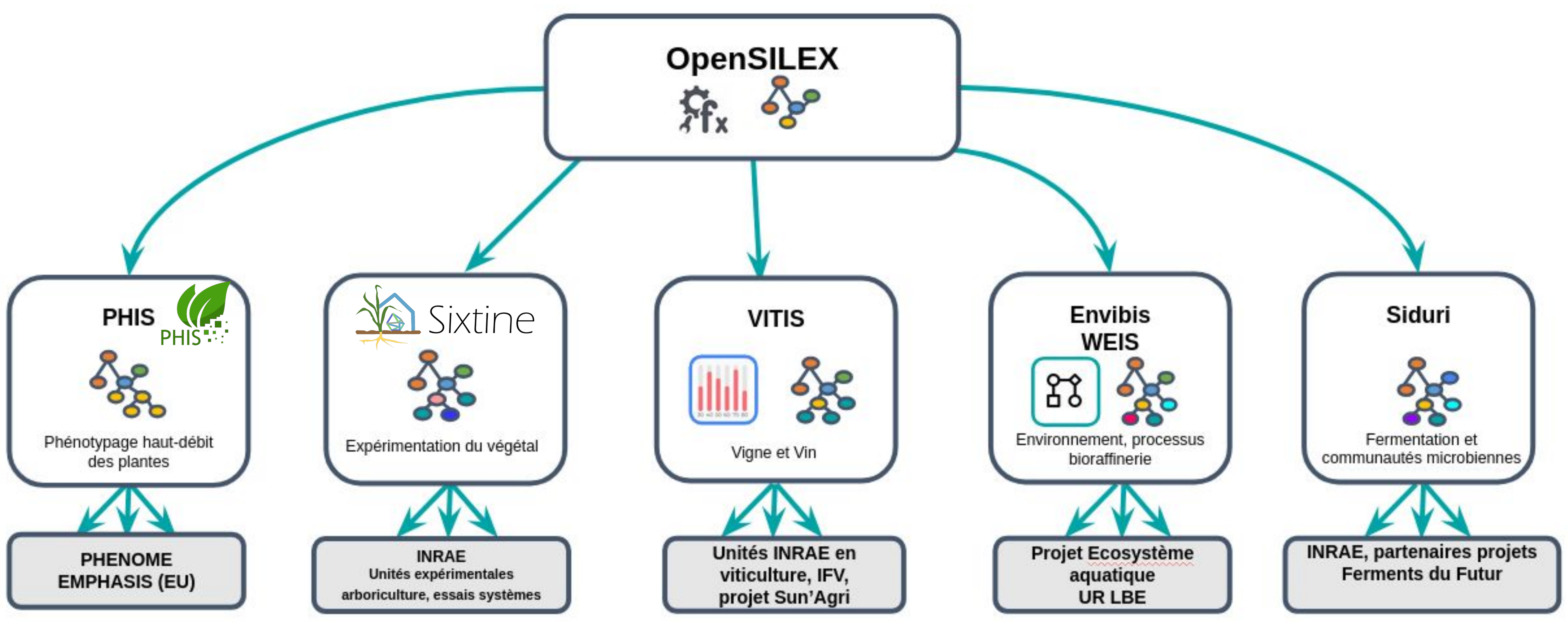
=> **OpenSILEX opted for a hybrid approach based on NOSQL and triplestore**

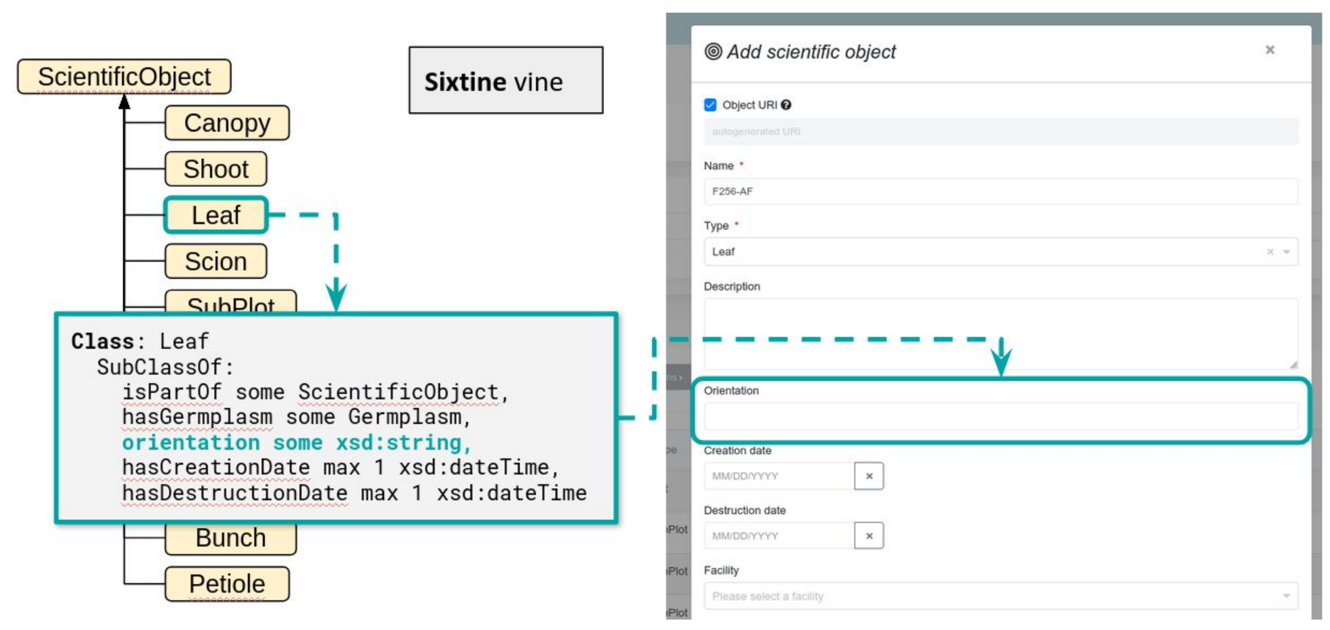
=> **Make data valuable!**

2- System-specific with community vocabularies & specific functionalities

Create Information Systems for management of heterogeneous and large-scale data:

- Based on a set of core functionalities and ontologies
- Enrich with modules for specific functionalities and vocabularies dedicated to communities





3- An adapted user interface thanks to ontology

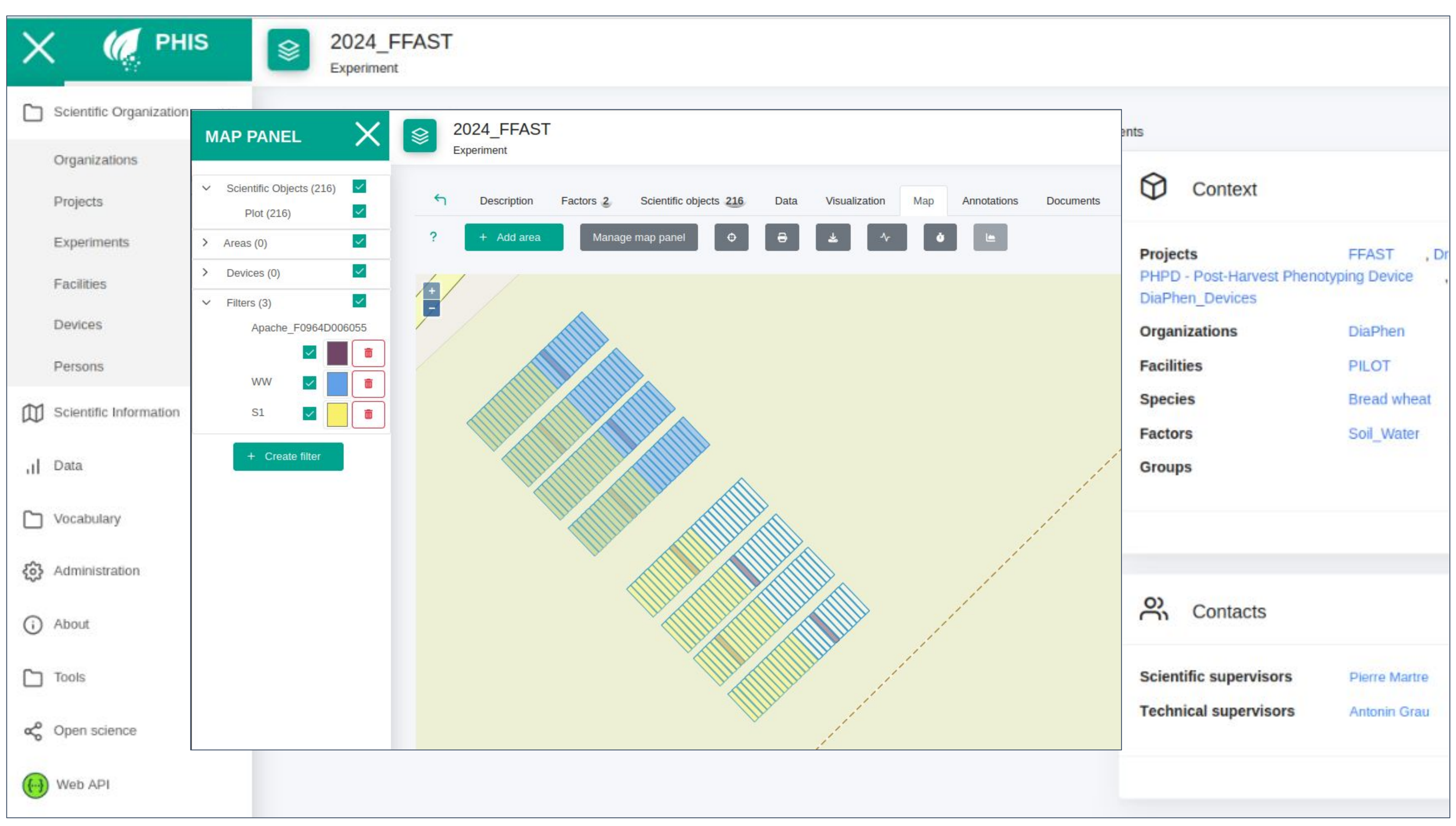
- **Ontology driven**, based on a application **Ontology for Experimental Scientific Objects (OESO)**
- Adapt context, labels, forms, templates, search, etc.
- Use **communities vocabularies**: CropsOntology or Plant Ontology

4- Describe specific experimental context

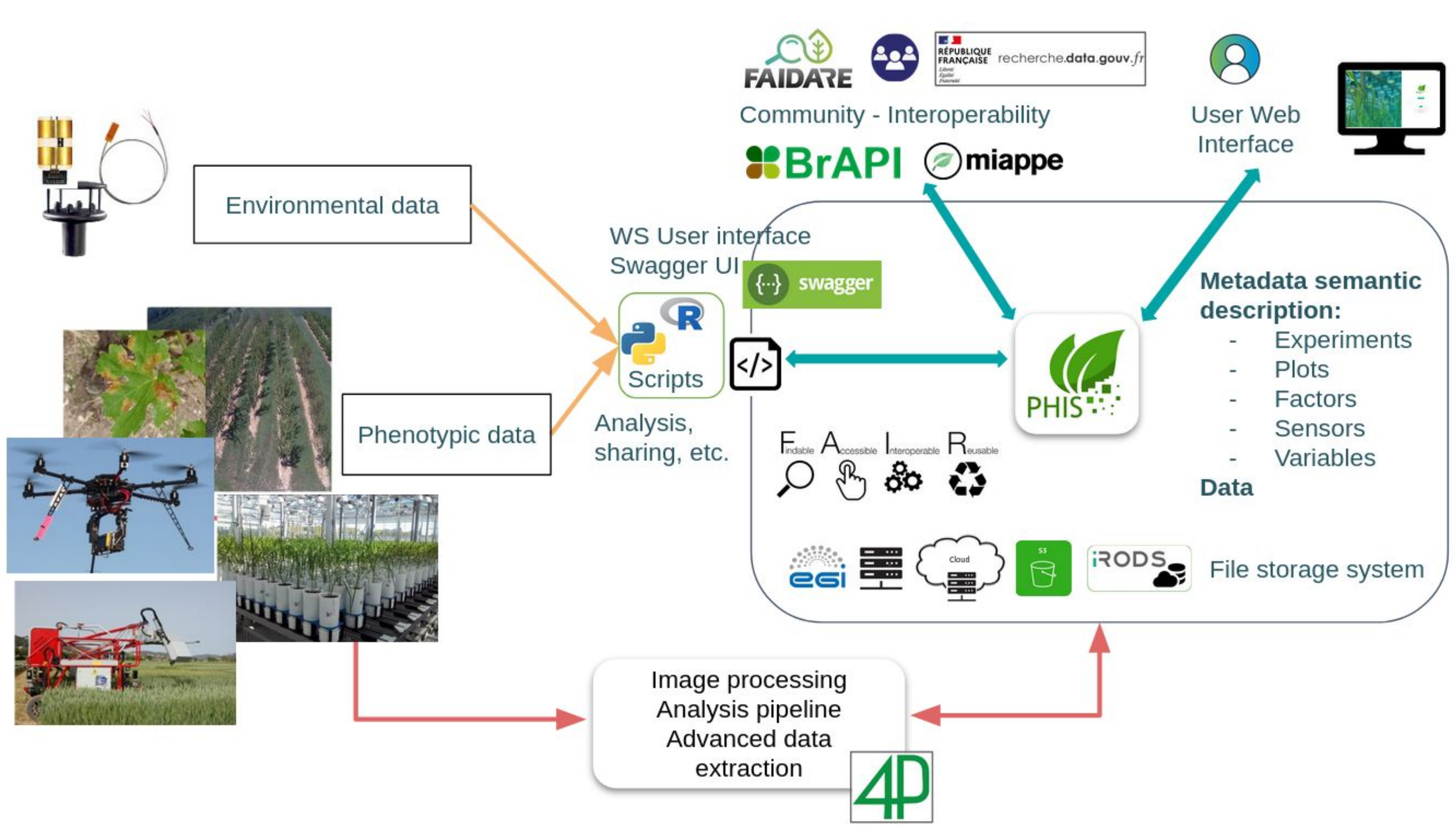
- Identify each element (project, experiment, plot, camera…) by an **URI**
- Involve studied scientific **objects** (plant, leaf, process...)
- Make **links**, improve data exploration, sharing and **Open Science**
- Enrich with provenance, documents or annotations

=> Get better **traceability** or **reproducibility** and profit expert **knowledge**

4-bis Take advantage of a geospatial visualisation of your experimental design



5- Workflow and interoperability: from acquisition to sharing

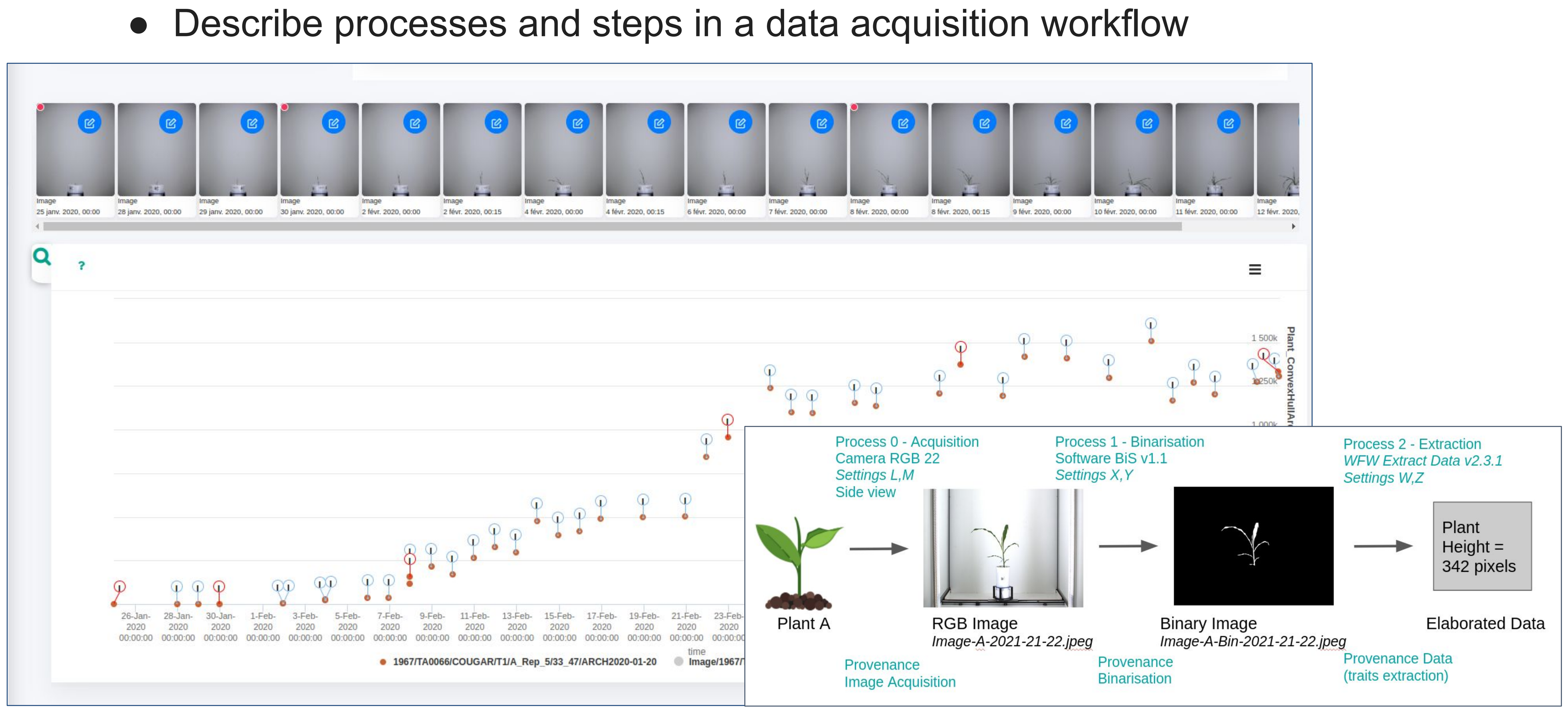


6- Digital Resources Centers: share across experimental platforms

- Centralise community resources: semantics, genetics, variables, etc.
- Share and reuse knowledge and promote Open Research.

7- Traceability image-processed: explore from trait to raw data

- Track transformations of raw data into processed and elaborated data
- Describe processes and steps in a data acquisition workflow



Website: https://opensilex.org/ - Contact: opensilex@inrae.fr - Test: http://opensilex.org/sandbox/app/ - GitHub: https://github.com/OpenSILEX/