



## **ELTerm: a terminology module for a plant data management system**

Lysiane Hauguel, Tanguy Lallemand, Rayan Eid, Fabrice Dupuis, Sylvain Gaillard, Florian Blessing, Sandra Pelletier, Julie Bourbeillon

### **► To cite this version:**

Lysiane Hauguel, Tanguy Lallemand, Rayan Eid, Fabrice Dupuis, Sylvain Gaillard, et al.. ELTerm: a terminology module for a plant data management system. Journée Ouvertes de Biologie, Informatique, Mathématiques, JOBIM 2020, Jun 2020, Montpellier (virtuel), France. . hal-03205869

**HAL Id: hal-03205869**

**<https://institut-agro-rennes-angers.hal.science/hal-03205869>**

Submitted on 22 Apr 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

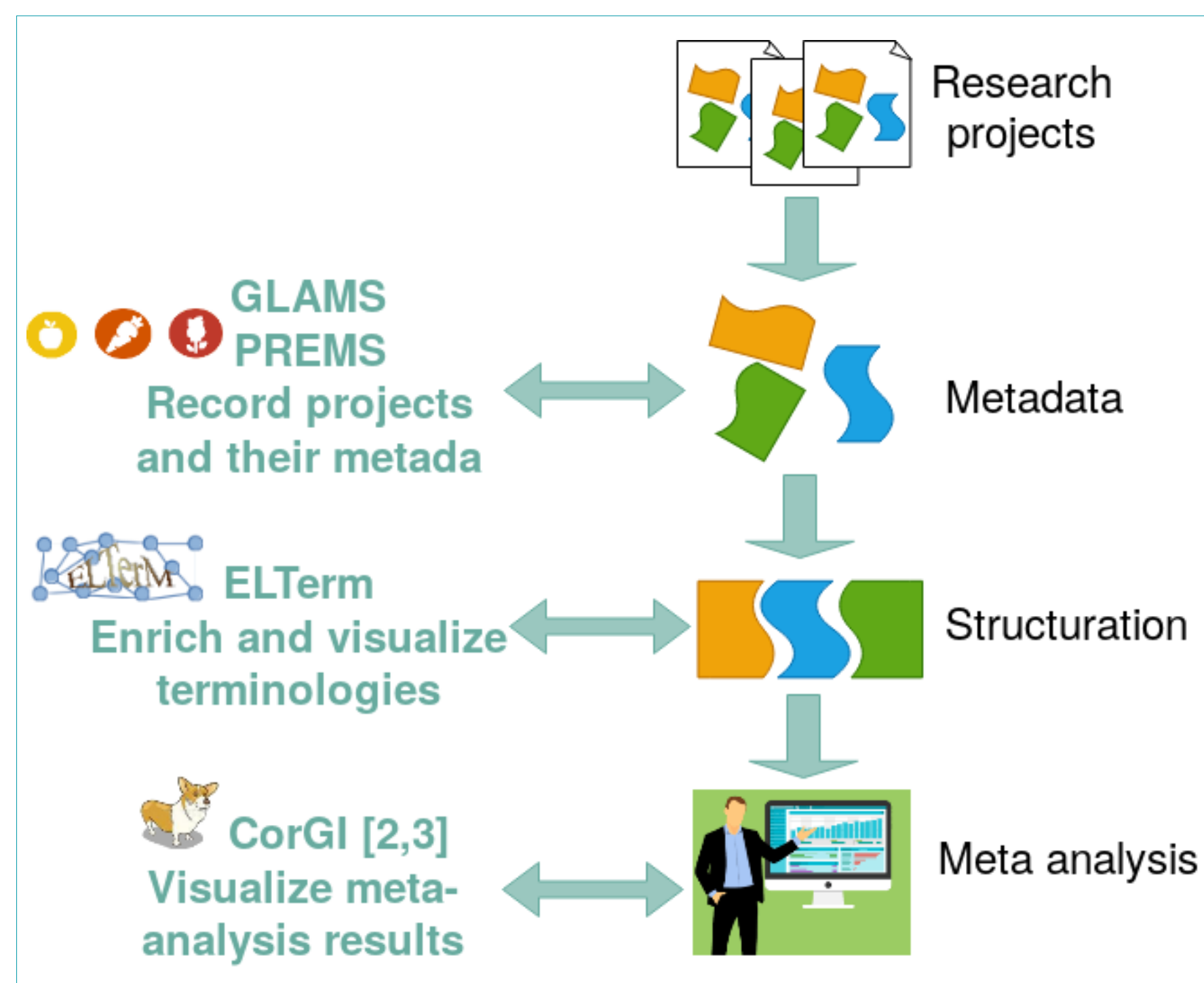


# ELTerm: a terminology module for plant experiments metadata management

Lysiane Hauguel <sup>1</sup>, Tanguy Lallemand <sup>1</sup>, Mickaël Ivanoff <sup>1</sup>, Rayan Eid <sup>1</sup>, Fabrice Dupuis <sup>1</sup>, Sylvain Gaillard <sup>1</sup>, Florian Blessing <sup>1</sup>, Sandra Pelletier <sup>1</sup> and Julie Bourbeillon <sup>1</sup>  
Institut de Recherche en Horticulture et Semences (IRHS), Université d'Angers, INRAE, AGROCAMPUS-Ouest, SFR4207 QUASAV, Université Bretagne Loire, Angers, France.

## Introduction

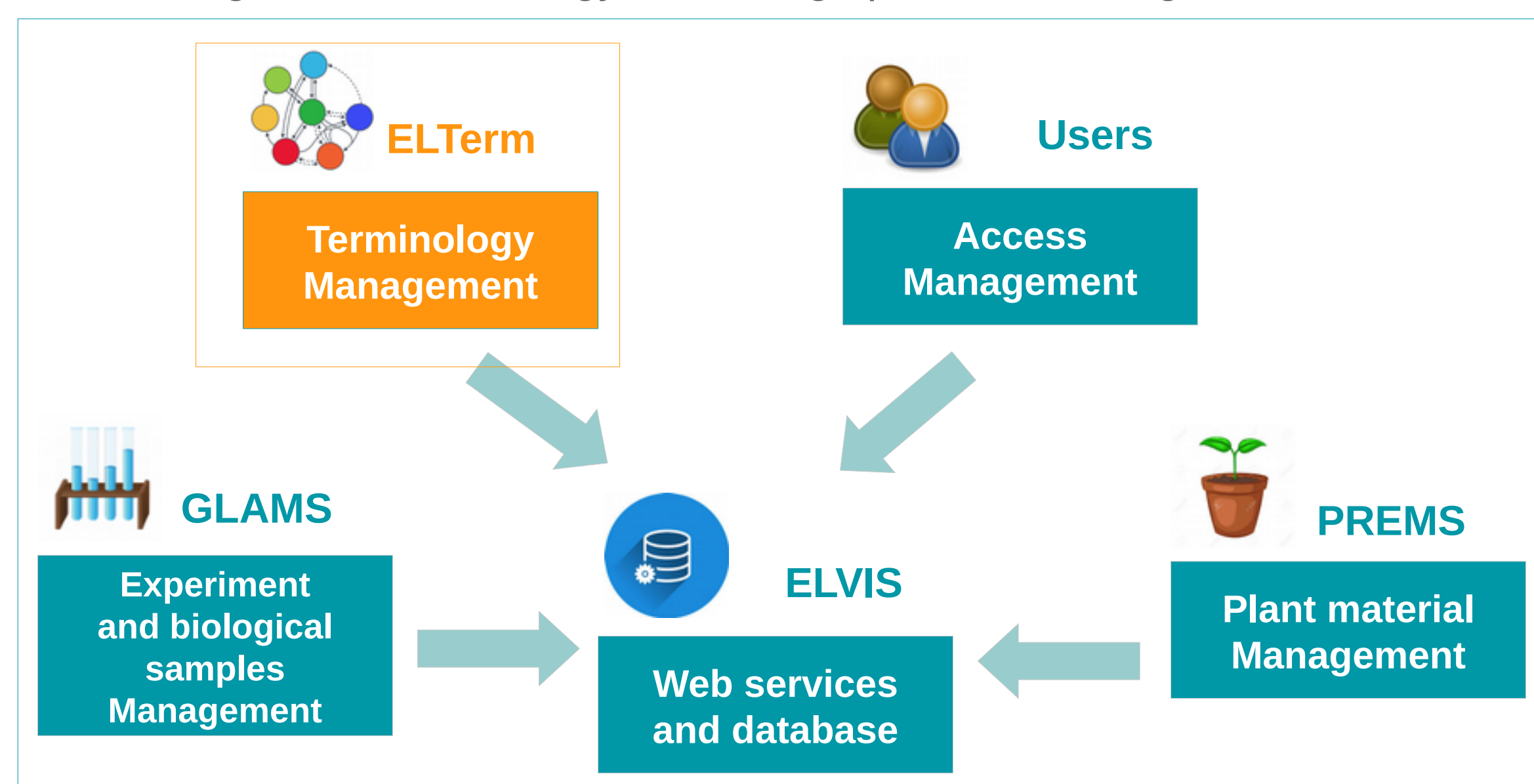
High-throughput meta-analyses of -omics or phenotypic data require a standardized collection of data associated with the experiments. It is a sine qua non condition to exploit this great amount of data. We have developed the ELTerm tool to manage the metadata associated with the experiments that we carry out on perennial or annual plants in our institute. This tool is a companion of the ELVIS database[1] so we named it ELTerm for ELvis TERMinology module. We briefly explain below the ontology and the organization of our data management system.



## Plant experiments management system

### Architecture

- ELVIS: core database and web services with restricted access according user status
- PREMS: manages informations relating to plants (species, varieties, offspring, origin, lots, etc.) and associated phenotypic notations
- GLAMS: manages laboratory samples related to scientific projects
- **ELTerm**: manages the terminology according specifics ontologies



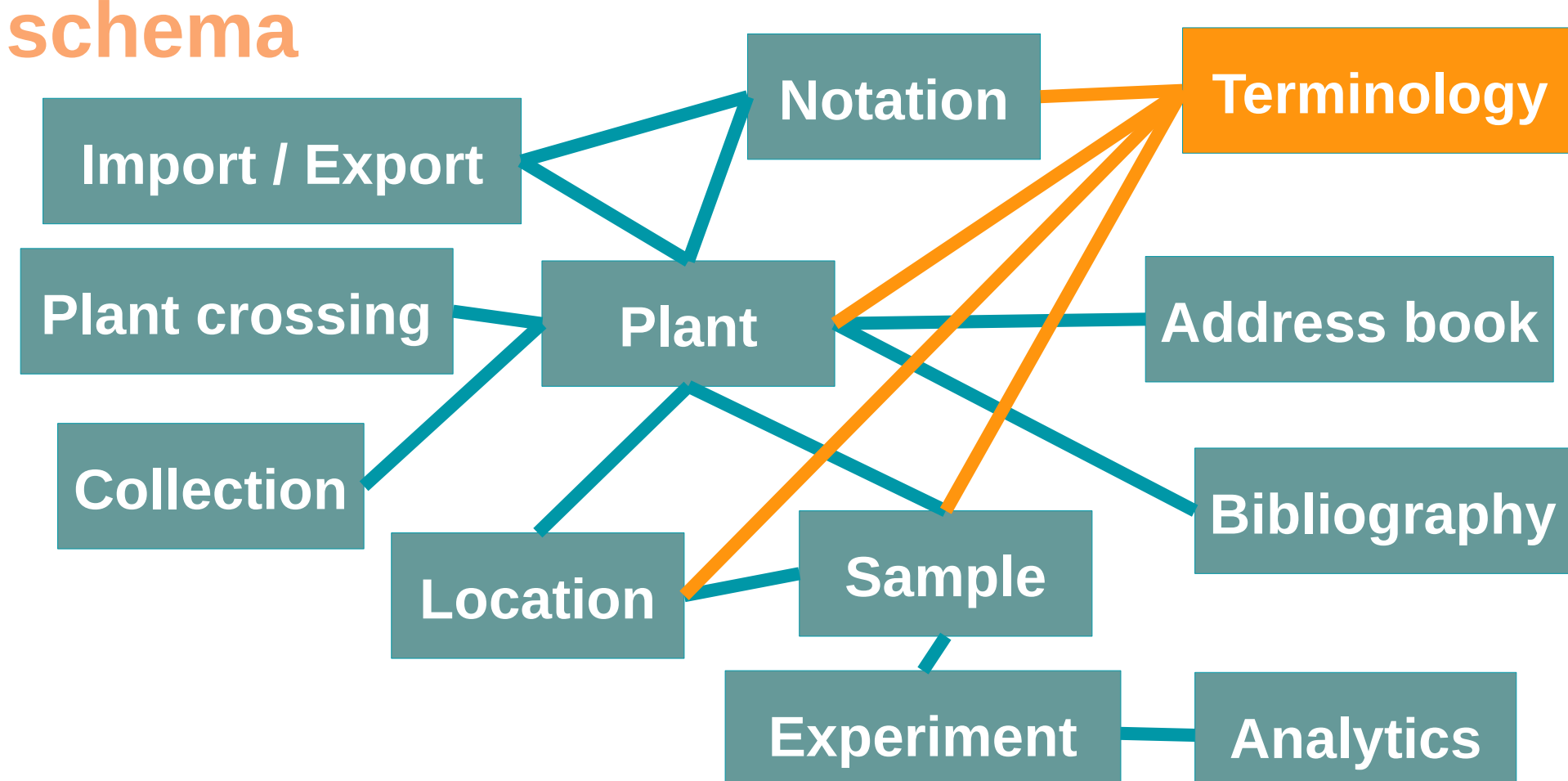
### Technologies



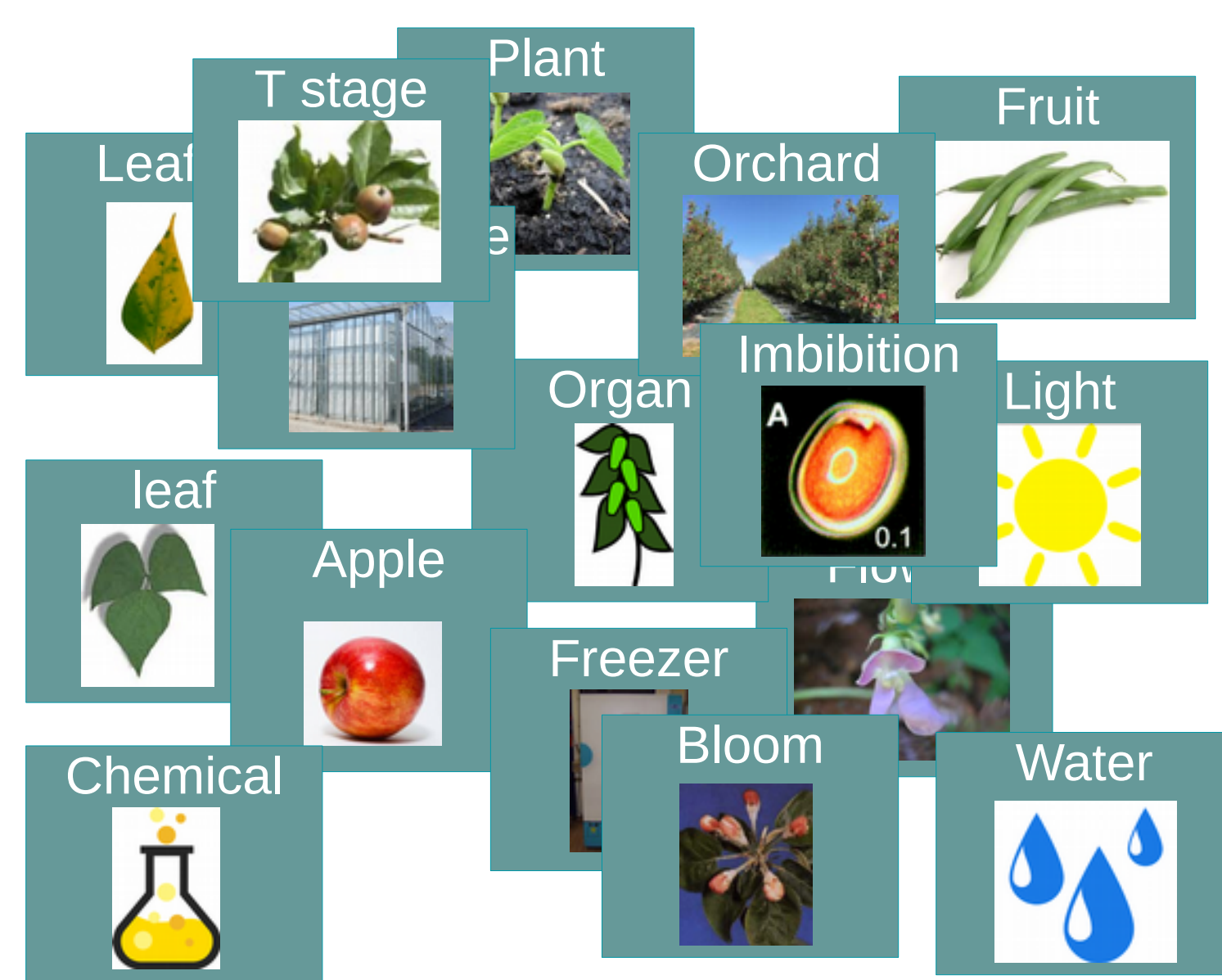
### Simplified database schema

The ELVIS database schema is organised in several modules to manage plants, experiments and associated informations.

The terminology module is used to store metadata which are used to annotate plants, notations, locations and samples.



### Needed metadata representations



Metadata associated with the data stored in ELVIS suggest the need for knowledge representation regarding:

- **Plant anatomy**
  - Generic representation
  - Species specific representations
- **Experimental conditions**
- **Development stage**
  - For the whole plant
  - For specific organs (seed, flower, fruit, leaf)
- **Locations**
  - For plant growth
  - For sample storage

## References

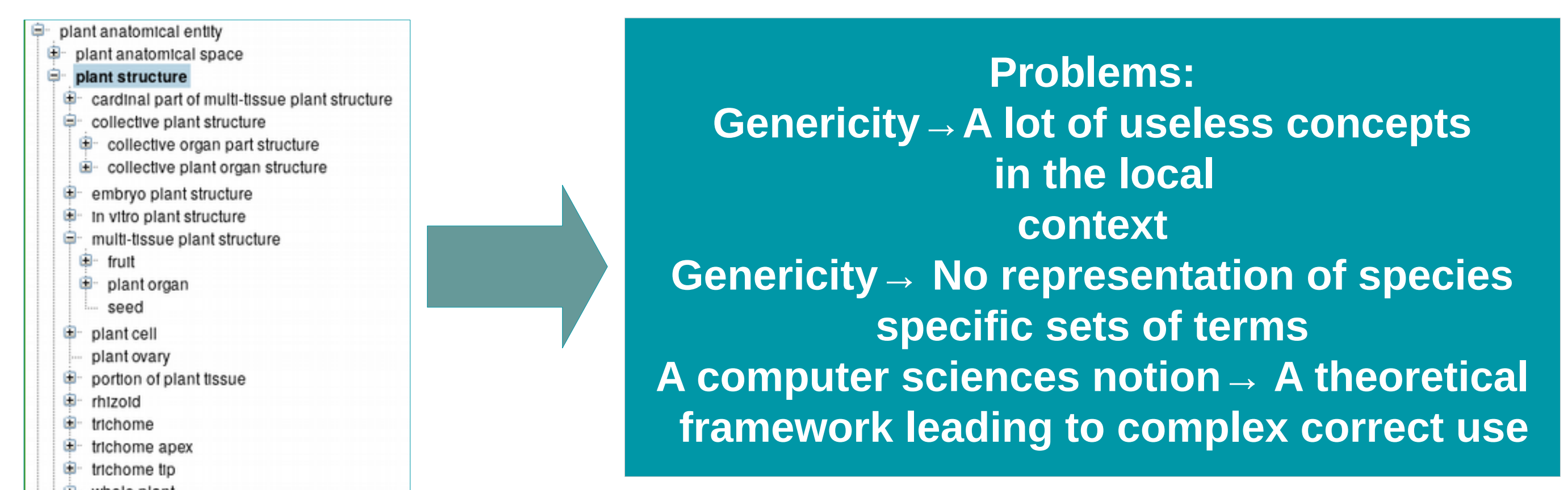
- [1] Dupuis F, Lelièvre A, Pelletier S, Thouroude T, Bourbeillon J, Gaillard S. PREMS/ELVIS : a local plant biological resource management system. 3-7 juillet 2017, Journées Ouvertes Biologie Informatique et Mathématiques (JOBIM'2017) Lille, France.
- [2] Pelletier S, Gaillard S, Aubourg S, Martin Magniette M-L, Brunaud V, Tamby J-P, Pereira H, Höfte H, Renou J-P. CORGI : Co-Regulated Gene Investigator. 28-30 juin 2016, Journées Ouvertes Biologie Informatique et Mathématiques (JOBIM'2016) Lyon, France.
- [3] Lallemand T, Gaillard S, Pelletier S, Landès C, Aubourg S, Bourbeillon J. Visualizing metadata change in gene networks and clusters. 2-5 juillet 2019, Journées Ouvertes Biologie Informatique et Mathématiques (JOBIM'2019) Nantes, France.

## Why a specific terminologies management system?

### Controlled vocabulary



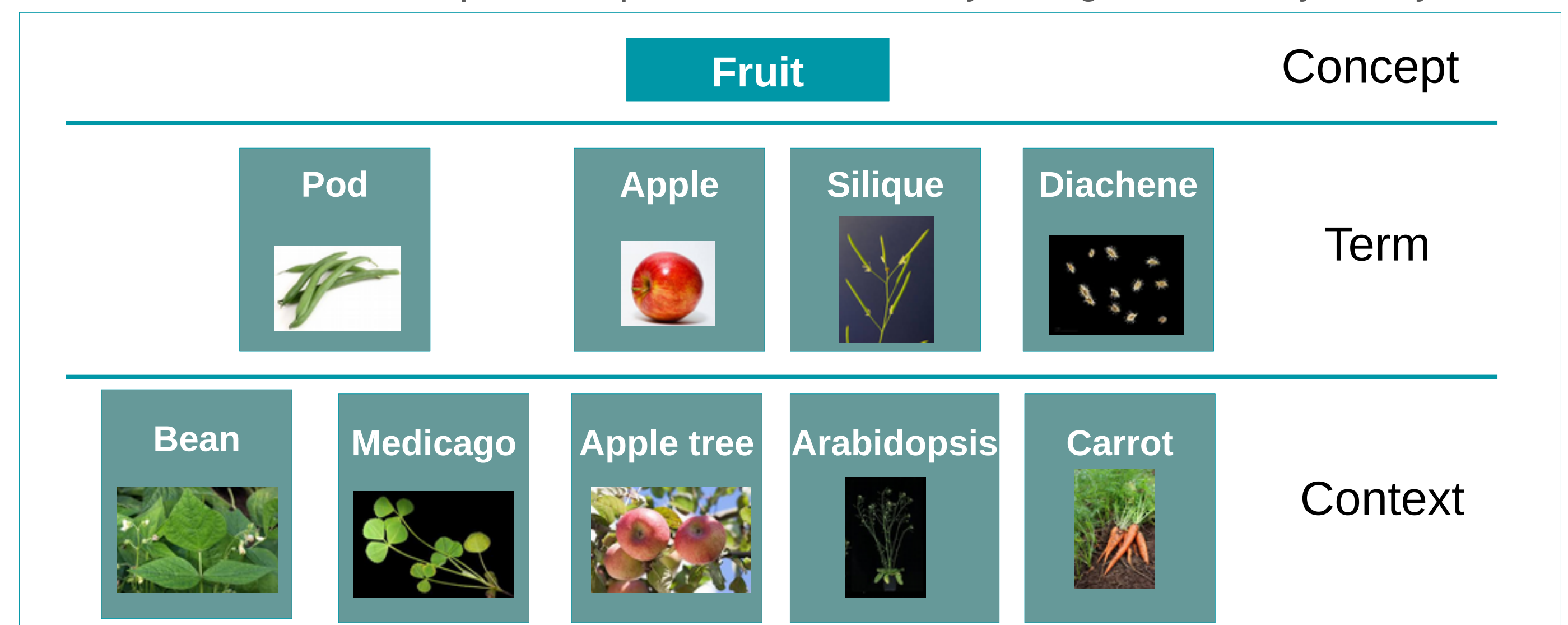
### Reference Ontologies



## Terminology management system

### Principle

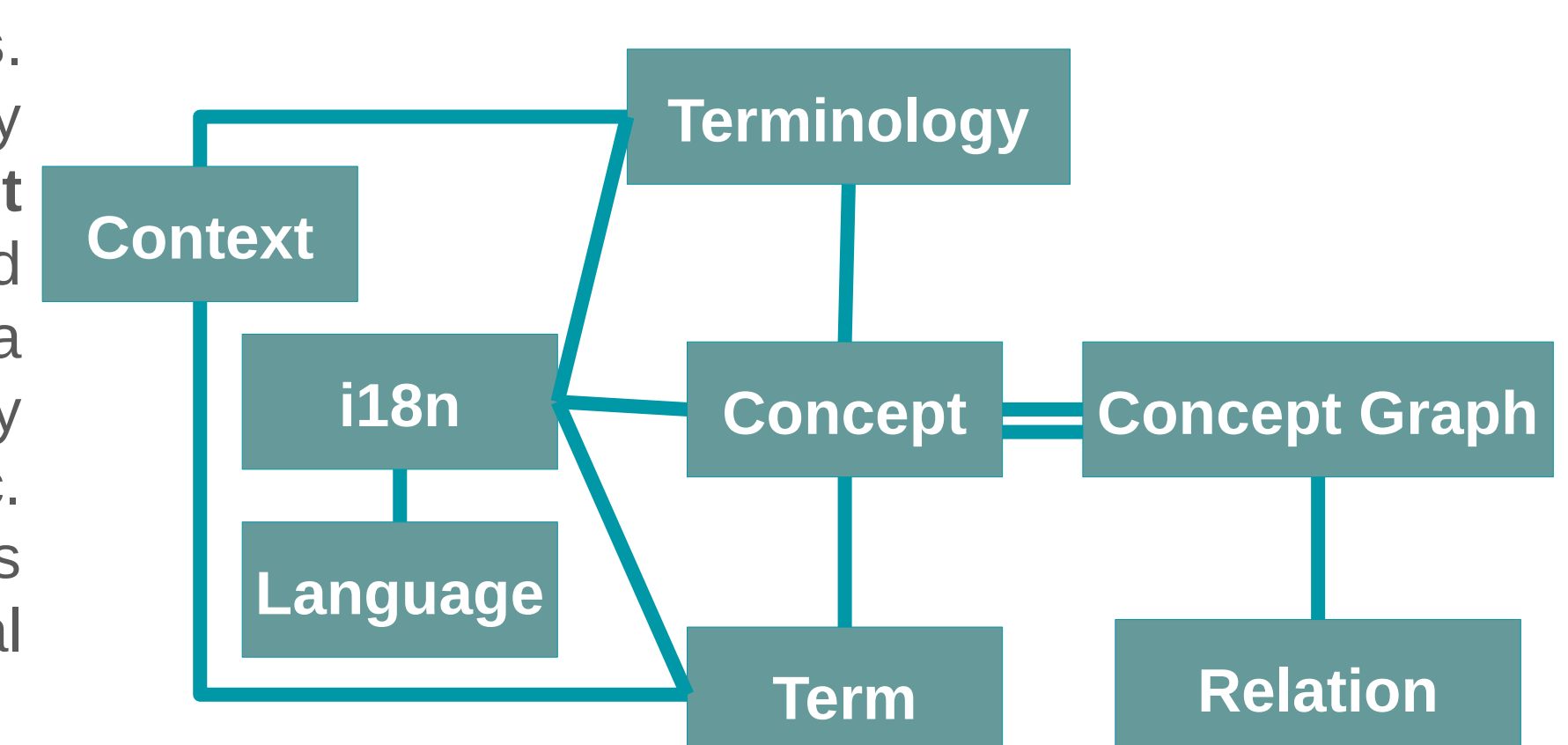
- A terminology as a direct acyclic graph:
  - Concepts as nodes
  - Relationships as edges
- A terminology as a generic representation of the world:
  - For instance « Fruit »
- A « context » notion to represent specific terms used by biologists on a day to day basis



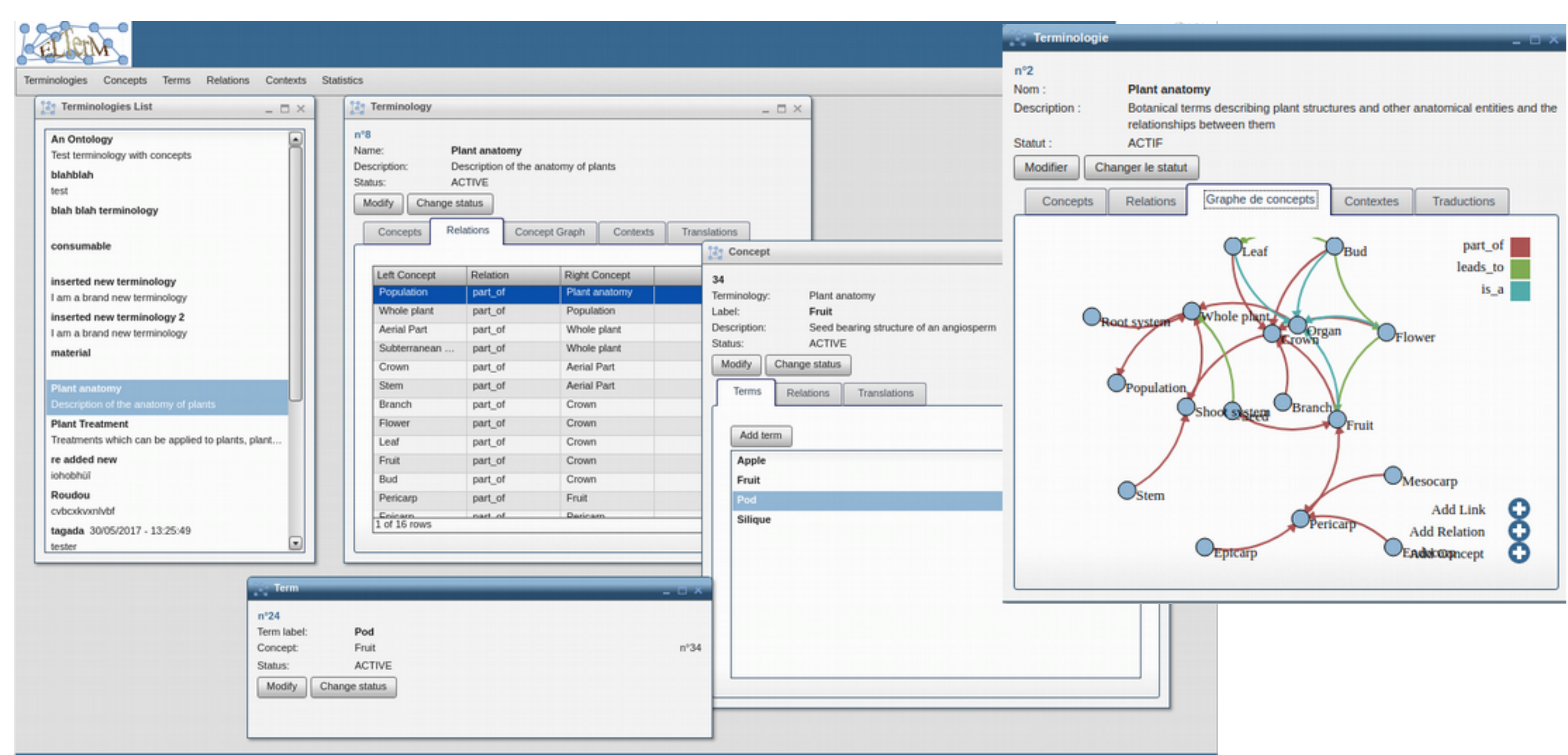
- An inclusion in the database schema and a similar graphical user interface

### Simplified terminology database schema

A **Terminology** regroups **Concepts**. Concept pairs are linked by **Relations** to form a **Concept Graph**. Concepts are associated with **Terms** which are relevant to a specific **Context**. Terminologies may also be Context specific. Terminologies, Concepts and Terms are translated (**i18n**) in several **Languages**.



### Interfaces



## Conclusion

Our plant experiment management system includes a functional terminology management module which is inspired by the ontology notion but largely simplifies it for ease of use by biologists in our local context. It introduces a « context » notion to manage synonyms or equivalence between terms corresponding to the same concept in various species. This allows use to perform meta-analyses, in particular multi-species studies, or to regroup data by exploiting relations between concepts, for instance subsumption.