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 ¹, Tanguy Lallemand ¹, Mickaël Ivanoff ¹, Rayan Eid ¹, Fabrice Dupuis ¹, Sylvain Gaillard ¹, Florian Blessing ¹, Sandra Pelletier ¹ and Julie Bourbeillon ¹

Institut de Recherche en Horticulure 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.

Why a specific terminology management system?

Controlled vocabulary

Retrieve all leaf images (including leaflets)?

No structure – no links between related items

Reference Ontologies

Genericity - A lot of useless concepts in the local context
Genericity – No representation of species specific sets of terms
A computer sciences notion – A theoretical framework leading to complex correct use

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

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 ans Terms are translated (i18n) in several Languages.

Interfaces

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.

Conclusion

References


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

Database management System
PostgreSQL

Web services / Libraries
Python 3
Qooxdoo javascript framework

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

Why a specific terminologies management system?

• 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

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

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 ans Terms are translated (i18n) in several Languages.

Interfaces

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.

Conclusion

References