

ELTerm: a terminology module for a plant data management system

Lysiane HAUGUEL¹, Tanguy LALLEMAND¹, Rayan EID¹, Fabrice DUPUIS¹, Sylvain GAILLARD¹,
Florian BLESSING¹, Sandra PELLETIER¹ and Julie BOURBEILLON¹
IRHS, Agrocampus-Ouest, INRAE, Université d'Angers, SFR 4207 QuaSaV, 49071, Beaucouzé, France

Corresponding author: julie.Bourbeillon@agrocampus-ouest.fr

1 Introduction

IRHS is a French plant research laboratory. Its BIDEfl (for Bioinformatics plant Defense Investigations) team develops methods and tools to help other teams manage and analyse their datasets:

- Development of a data management system which links research projects, experiments and the biological material being used (along with its characteristics) and produced results,
- Development of various data analysis tools and in particular tools aiming at meta-analysis

In order to perform meta-analyses, it is mandatory to use a unified vocabulary across experiments to describe the datasets, that is to say use shared terminologies to fill in meta-data fields.

2 Terminology management in ELVIS/PREMS

The data management system consists of a data management layer, named ELVIS (for Experiment and Laboratory on Vegetal Information System), and graphical interfaces, called PREMS (for Plant Ressource Management System) [1]. Both layers include a terminology module which contents are used to describe data in other modules. This module allows to store several terminologies related to various topics: organism taxonomy, experimental conditions (in particular stresses), plant development stages, plant anatomy, etc. The general underlying principal is similar to standard terminologies representations such as TermBase Exchange [2]. The objective is to both keep the representation simple without going into more complex representation such as those used in ontologies by knowledge management engineers while still being expressive enough to suit our needs.

A terminology regroups a set of concepts in a direct acyclic graph where nodes are concepts and edges are named and represent relationships between concepts. Each concept is associated to a set of terms which support it. A terminology is ideally generic: for instance "Plant Anatomy". Therefore concepts are represented by generic terms: for instance "Fruit". However biologists usually use specific words: for instance people working on *Malus domestica* refer to "Apple" and not "Fruit". In our representation such specific words are terms associated to the relevant concept. Moreover, in order to keep track of the relevant context of use of a given word, we introduced a context notion. For instance the term "Apple" is associated with the "Malus" context. In PREMS, it will only be presented when people are inputting "Malus" related data. Therefore this context notion allows us to represent both specific and generic information along with the equivalence between the two. It also allows us to map our local ontologies, designed to be close to the day to day use of the biologist teams, with reference ontologies such as the Plant Ontology [3].

Acknowledgements

This work was supported by the region Pays de Loire as part of the RFI "Objectif Végétal" program.

References

- [1] Lysiane Hauguel, Fabrice Dupuis, Sylvain Gaillard, Julie Bourbeillon, Claudine Landes, and Sandra Pelletier. Mise en place d'un lms enrichi par une organisation harmonisée des métadonnées. In *Journées Ouvertes de Biologie, Informatique et mathématiques*, Nantes, France, 2-5 July 2019.
- [2] ISO/TC37/SC3. Iso 30042:2019 management of terminology resources — termbase exchange (tbx). Technical report, International Organization for Standardization, 2019.
- [3] Shulamit Avraham, Chih-Wei Tung, Katica Ilic, Pankaj Jaiswal, Elizabeth A. Kellogg, Susan McCouch, Anuradha Pujar, Leonore Reiser, Seung Y Rhee, Martin M Sachs, Mary Schaeffer, Lincoln Stein, Peter Stevens, Leszek Vincent, Felipe Zapata, and Doreen Ware. The plant ontology database a community resource for plant structure and developmental stages controlled vocabulary and annotations. *Nucleic Acids Research*, 36(suppl 1):D449–D454, 01 2008.