Towards a Plant Experimental Assay Ontology

Nuno D. Mendes[1,2], Pedro T. Monteiro[1], Cátia Vaz[1,3], and Inês Chaves[1,4]

[1] INESC-ID - Instituto de Engenharia de Sistemas e Computadores, Lisboa, PT
{ndm,ptgm}@kdbio.inesc-id.pt,
[2] IBET - Instituto de Biologia Experimental Tecnológica, Oeiras, PT
[3] ISEL - Instituto Superior de Engenharia de Lisboa, Lisboa, PT
cvaz@cc.isel.ipl.pt,
[4] ITQB - Instituto de Tecnologia Química e Biológica, Oeiras, PT
ichaves@itqb.unl.pt

Abstract. The Plant domain has been the subject of several attempts to structure and formally define terms and corresponding relations, such as their anatomical features, developmental stages, and the application of particular experimental procedures to a biological problem. However, a focus on experimental assays in order to describe the whole experimental procedure, to the best of our knowledge, has only been attempted in the context of a very general description based on classical views of the scientific method [1]. In this study, we focus on the development and proposal of an ontology dedicated to the description of these experimental procedures, regardless of the scientific questions that prompted the assays. This ontology includes entities from three distinct realms (biological, physical and data), which include both experimental products, their relations and the protocols describing their manipulation. The final outcome is a useful and comprehensive ontology in the plant domain, to be used as a log book by experimentalists, providing a formal relation between entities.

Keywords: Ontology, Modeling, Biology, Experimental assays

Introduction

Ontologies originated from the need to formally specify a controlled set of terms and their relationships in the context of a knowledge domain. The advantages of this type of approach include the ability to share structured information between different users and software tools, to reuse the established vocabulary, and, not less importantly, to make domain assumptions explicit [2].

Following their inception as a formal knowledge representation technique, several tools have since surfaced which elicit their practical use, including sophisticated ontology editors [3], semantic databases and query languages.

The wealth of data generated by contemporary biological experimental studies presents several challenges, not only due to its magnitude, but also to its heterogeneity and interdependence. Additionally, the diversity of protocols, tools and data formats, as well as the several different context-specific parameters used at different steps render the fundamental requirement of experimental reproducibility much more difficult to accomplish. This calls for an attempt to meaningfully represent the experimental procedures as well as the data they produce.

Here we focus our attention on experimental assays designed for studying the plant domain. Studies with other types of subjects (e.g animals or bacteria) may be significantly different, which justifies our choice to restrict the scope of our efforts.

Problem and Designed Solution

Several ontologies describing developmental and anatomical characteristics of plants, their environment and even the types of stress they can be subject to have already been proposed (e.g. Plant Ontology (PO) [4], Plant Trait Ontology (TO) [5], Plant Infectious Diseases (IDOPlant) [6]). Although ontologies specifically dedicated to the description of experimental procedures in general do exist [1], their foremost concern is the description of experimental design, hypothesis testing and the ultimate goal of the experiments. The ontology proposed here, on the other hand, is mainly dedicated to the description of the pipeline of manipulations performed from specimens to data.

This focus on the actual experimental procedure rather than the teleological and epistemological foundations of the assay is justified by the observation that the data produced by experimental assays may be, and is, generally used in several subsequent studies with, presumably, different objectives and that there is much data produced outside the scope of a clear experimental design or a particular biological question (e.g. genome sequencing). There are, however, several opportunities to reuse preexisting ontologies, namely in the description of the growth environment [7], designation of the fractioned samples [4], or the modelling of the notion of time when referring to operations in general [8].

Our ontology, which is currently in active development (summarised in Figure 1), considers essentially three types of entities: biological entities, physical entities and data entities. An additional term - material entity - is used to group two top-level concepts from the biological and physical realms, for convenience. An ancillary ontology was also created to describe iterative processing pipelines (see Figure 2).

Biological entities refer to biological material or to manipulations thereof. A central concept is that of Biological Subject, which can be a specimen or a specimen pool (a single plant, or a group of plants, which can be several individual clones, individuals of different species, etc), or a Biological Sample, which is a transformation of an original Biological Subject. These transformations can be either experimental manipulations (e.g. biotic or abiotic stress) or the result of fractioning (e.g. the isolation of a particular tissue or another anatomical feature). Successive Biological Samples can be obtained by applying a sequence of transformations. Physical entities refer to non-living material, usually the result of an extraction procedure (e.g. nucleic acids extraction, protein extraction, etc), or the product of an analysis protocol (e.g. a gel obtained from electrophoresis), or the manipulations thereof. Data entities, likewise, refer to informational concepts and their manipulation. Here Data Subjects are organized in subclasses related to a particular *DataAcquisitionProtocol*. Thus, *ImageData* is obtained using an *ImageAcquisitionProtocol*, *ProteinData* with, for instance, *MassSpectrometry*, and so forth.

Some of the above mentioned entities have object properties characterizing their relationships and effectively imposing restrictions between the instances of each entity. Examples of these properties are: a *PhysicalAggregate* which contains one or more *PhysicalSubject*, or a *Specimen* which has a *Taxonomy* and it contains one or more *SpecimenPool*.

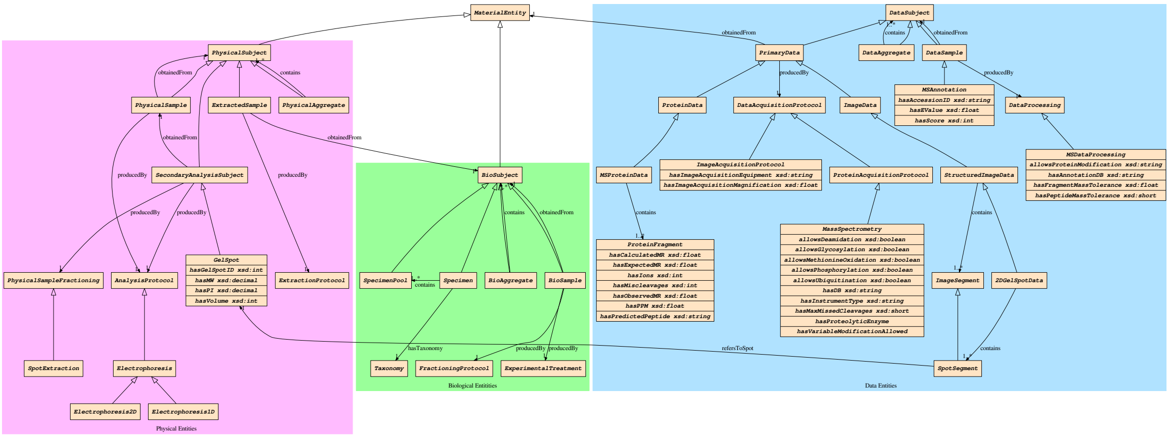

Fig. 1. Ontology representing experimental products, their relations and protocols divided into three main categories of entities: biological, physical and data. A full description of the proposed ontology (generated by Protégé's OWLDoc plugin) can be found at http://kdbio.inesc-id.pt/PEAO/.

The distinctions we here present about the nature of these entities (biological, physical or data) are not explicitly distinguished by the ontology, but their interdependence is constrained by the domain and range of the object properties it defines.

Ongoing Work and Ontology Implementation

As an example illustrating the usefulness of our proposition, we considered the data collected from the interactions between coffee and coffee leaf rust. The data used as test-case is the description of the experimental design and stress imposition, the cytological results, protein gel electrophoresis images, mass spectrometry results and protein assignment results with annotation and extraction and analysis protocols. These data is continually being produced by the Centro de Investigação das Ferrugens do Cafeeiro of Instituto de Investigação Científica Tropical (IICT), and it has been analysed by Instituto

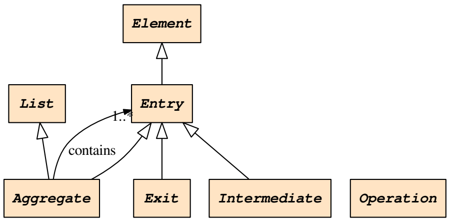

Fig. 2. Auxiliary ontology describing iterative process pipelines.

de Tecnologia Qumica e Biológica/Instituto de Biologia Experimental Tecnológica (iBET/ITQB) in collaboration with Instituto de Engenharia de Sistemas e Computadores (INESC-ID).

The plant experimental assay ontology and the pipeline patterns ontology, are being developed using the modelling development tool Protégé [3]. This tool is widely used for modelling a wide range of domains, providing tools for visualization, ontology merging, refactoring or embedded reasoners for the logical verification of the ontology.

The parameters of the several types of manipulations of biological, physical and data entities are given as data properties of those manipulations. For instance, a *SpotSegment*, a special case of *Data Entity*, refers to a *GelSpot*, a special case of *Physical Entity*. The latter contains four data properties: *hasGelSpotID*, *hasMW*, *hasPI* and *hasVolume*, describing the results of an electrophoresis. Each electrophoresis, which is a particular type of an analysis protocol that can be performed to a *Physical Entity*, will then contain multiple *GelSpot*. On the other hand, a *SpotSegment* is a special case of an *ImageSegment* which is part of a *StructuredImage*, corresponding to an image of the actual gel produced by electrophoresis.

Conclusions

This work is a first attempt at enabling the creation of a repository of data produced by plant experimental assays. The representation of the data using an ontology elicits the preservation of the semantic relationships between the entities represented therein, which facilitates the interpretation of the results and the integration of data produced by different experiments. Additionally, the ontology is easily extensible to incorporate new types of data and experimental manipulations and can, therefore, accommodate a growing number of different experimental procedures.

Acknowledgments: This work was supported by national funds through FCT (Fundação para a Ciência e a Tecnologia), under projects PEst-OE/EEI/LA0021/2013 and DataStorm EXCL/EEI-ESS/0257/2012.

References

1. L.N. Soldatova, R.D. King: An Ontology of Scientific Experiments. J. R. Soc. Interface. 3(11), 795–803 (2006)
2. Michael Hammer, Dennis McLeod: Database Description with SDM: A Semantic Database Model. TODS 6(3), 351–386 (1981)
3. J.H. Gennari, M.A. Musen, R.W. Fergerson, W.E. Grosso, M. Crubezy, H. Eriksson, N.F. Noy, S.W. Tu: The Evolution of Protégé: an Environment for Knowledge-based Systems Development. Int. J. Hum. Comput. Stud. 58, 89–123 (2003)
4. Ilic K., E. A. Kellogg, P. Jaiswal, F. Zapata, P. F. Stevens, L. Vincent, S. Avraham, *et al.*: The Plant Structure Ontology, a Unified Vocabulary of Anatomy and Morphology of a Flowering Plant. Plant Physiology 143:587–599 (2007)
5. Jaiswal P.: Gramene Database: A Hub for Comparative Plant Genomics. Methods Mol. Biol. 678:247–275 (2011)
6. Walls R. L., B. Smith, J. Elser, A. Goldfain, D. W. Stevenson, and P. Jaiswal: A Plant Disease Extension of the Infectious Disease Ontology. Proc. 3rd Intl. Conf. Biomedical Ontology. (2012)
7. http://www.ontobee.org/browser/index.php?o=EO. Accessed January 28, 2014
8. http://www.w3.org/TR/owl-time/. Accessed May 29, 2014