

Testing the Bioportal Technology with the Crop Ontology

Project description – Montpellier, March 2015.

Collaborators:

Clement Jonquet – jonquet@lirmm.fr (University of Montpellier)

Pierre Larmande – pierre.larmande@ird.fr (Institute of Research for Developing Countries and IBC project)

Elizabeth Arnaud – e.arnaud@cgiar.org (Bioversity-France)

Patrick Valduriez – patrick.valduriez@inria.fr (INRIA, ZENITH Team)

Pascal Neveu - pascal.neveu@supagro.inra.fr (French National Institute of Agronomy)



Research Context (scientific projects)

This collaboration will take advantage of the expertise engaged in two existing projects which are:

1. Semantic Indexing of French Biomedical Data Resources project (SIFR - www.lirmm.fr/sifr)

The SIFR project, led by Clement Jonquet, which investigates the scientific and technical challenges of developing ontology-based services to leverage the use of biomedical ontologies and terminologies for indexing, mining and retrieval of French biomedical data in collaboration with the National Center for Biomedical Ontology project which develops the NCBO BioPortal (<http://bioportal.bioontology.org>). The project reuses the NCBO technology to develop a French version of the ontology-based annotation workflow. The SIFR project is mainly funded by the French National Research Agency and actively collaborates with two other projects:

- The Institute of Computational Biology (IBC - <http://www.abc-montpellier.fr>) which aims to develop methods to aid data integration and knowledge management within the domain of Agronomic sciences to improve information accessibility and interoperability.
- The NUMEV Labex (<http://www.lirmm.fr/numev>) which is a funding infrastructure aiming to support trans-disciplinary projects between ICT and life sciences.

2. The Crop Ontology project of CGIAR (www.cropontology.org)

Since 2008, Elizabeth Arnaud, Bioversity International, Montpellier have led the Crop Ontology project, as part of the Generation Challenge Programme and the Integrated breeding Platform. The two primary objectives for the Crop Ontology project are: (1) To publish online fully documented lists of breeding traits used by the Breeding Management System (BMS) for producing standard fieldbooks, and (2) To support data analysis and integration of genetic and phenotypic data through harmonized breeders' data annotation. The project also offers a forum for scientists to discuss their variables, methods and scales of measurement, and fieldbooks.

Project Objectives

This project proposes to test the Bioportal technology on the Crop Ontology website to enable the creation of a thematic plant ontology repository. The project aims at reusing the outcomes of the biomedical domain in the context of plant and crop sciences for two key reasons: (i) to avoid re-developing technologies and tools that have already been designed and tested; (ii) to offer the same tools, services and formats in both domains to facilitate the interface and interaction between the domains e.g., to enable a user to query the NCBO BioPortal or the BioPortal Plant group without changing a line of code. The application of the Bioportal technology to the Crop Ontology must respect the requirements of the agricultural community and attention must be given to the

specificities of the crop domain (compared to the biomedical domain). The Crop Ontology community focuses on research for development, with immediate results, therefore needs access to tools that are rapid and fit for purpose. This project will allow us to combine complementary expertise from the Montpellier Agropolis scientific community with that of international teams to result in a practical platform that we anticipate will be highly valued by plant scientists.

Project description

Enhancing the Crop Ontology web site

The current website (www.croponotology.org) will remain the primary point of access to the Crop Ontology. This is an important consideration since the Crop Ontology is a service of the Integrated Breeding Platform (IBP), provides validated Crop concepts to Breeding database and fieldbooks, and is a key element of the CGIAR Open Data Strategy. The domain belongs to the CGIAR.

Communities engaged in plant research need to access specific sets of ontologies for plant data annotation and integration. The Crop Ontology website is a crop domain-focused website that publishes online sets of ontologies required for describing crop germplasm, traits and evaluation trials. Today, the Crop Ontology Curation Tool portal contains 18 species-specific ontologies in addition to ontologies related to the crop germplasm domain. Partners like the US Department of Agriculture (USDA), the Institut National de Recherche Agronomique (INRA), and the Polish Genomic Network called Polapgene have uploaded their ontologies. The International Wheat Initiative and data interoperability group have approached the Crop Ontology team to discuss integrating CO best practices. In addition to its role as a repository, the Crop Ontology portal offers community-oriented features such as an Excel template; easy upload of Excel files in various languages; creation of online ontologies, and filtering terms. Efforts have been made to structure and formalize the ontologies following semantic web standards as well as offering collaborative ontology enrichment and annotation features. The Crop Ontology is used by 3rd party web sites and databases (Agtrials, IBP, EU-SOL, etc) through the Application Programme Interface (API). The current website was designed for the community development of ontologies and necessitates improvements for the versioning of the ontologies and curation of the linguistic versions.

The CO web site facilitates the complete ontology-engineering life cycle starting with collaborative construction, publishing, versioning use and modification. However, there is a need to enhance the interface for the Crop Ontology term submission. For instance, the CO portal does not currently support versioning for ontologies but enables quick updates of the ontologies. It is important to support the alignment (or mapping) of terms within and across different ontologies: both within the CO itself (in different crop branch) or with other global ontologies commonly used in plant biology (Plant Trait Ontology, Plant Ontology, Pato, ENVO).

The Crop Ontology is referenced on the NCBO Bioportal aside other plant-related ontologies but is not currently fully accessible and described through this portal.

Customizing an instance of the Bioportal behind the Crop Ontology web site

The NCBO Bioportal is a well-known repository for biomedical ontologies. Its functionalities have been progressively extended, with a new release in February 2015 that aims at improving the user experience when accessing registered ontologies and using it for annotations. An instance of the Bioportal is currently being customized at the University of Montpellier, in the context of the SIFR and IBC projects (in collaboration with Pierre Larmande).

This project is essentially a feasibility study for maintaining and supporting a local Bioportal instance for plants, and reusing the recent technologies and services of the NCBO BioPortal, such as hosting, search, mappings, versioning, annotation, REST endpoint, visualization. The main objective is to enable straightforward use of plant related ontologies, freeing researchers to deal with complex knowledge engineering issues and to concentrate on biological challenges and conducting field trips.

A quick prototype of the BioPortal Plant Group

NCBO technology is domain-independent and open source. The NCBO Bioportal virtual appliance is available for the community allowing to set up a local ontology repository and eventually customize it: http://www.bioontology.org/wiki/index.php/Category:NCBO_Virtual_Appliance

We have already set-up a local instance of BioPortal on a LIRMM server. It is a prototype that currently hosts French terminologies (for the SIFR project) and English plant related ontologies (around 15) aggregated for the IBC project and testing for INRA. As part of the Wheat Data Interoperability Working Group activities, a group of users will be identified (e.g. IRD, INRA, CIRAD, Bioversity) to use and test the prototype with the Wheat terms and user needs: <http://bioportal.lirmm.fr>.

As an illustration of what could be done, we have developed an ontology group for Crop-related ontologies, provisionally called the CROP group.

| ONTOLOGY NAME | VISIBILITY | CLASSES | NOTES | REVIEWS | PROJECTS | UPLOADED | CONTACT |
|--|------------|---------|-------|---------|----------|------------|-----------------|
| Banana Ontology CO-BANANA | Public | 185 | 0 | 0 | 0 | 01/08/2015 | Clement Jonquet |
| Crop Research Ontology CO-CRO | Public | 256 | 0 | 0 | 0 | 01/08/2015 | Clement Jonquet |
| Germplasm Ontology CO-GO | Public | 386 | 0 | 0 | 0 | 04/16/2014 | Clement Jonquet |
| ICIS Germplasm Methods Ontology CO-GMO | Public | 166 | 0 | 0 | 0 | 01/08/2015 | Clement Jonquet |
| Molecular Markers Ontology CO-MMO | Public | 173 | 0 | 0 | 0 | 04/16/2014 | Clement Jonquet |
| Multicrop Passport Ontology CO-MCPO | Public | 87 | 0 | 0 | 0 | 01/08/2015 | Clement Jonquet |
| Plant Ontology PO | Public | 1,691 | 0 | 0 | 0 | 03/19/2014 | Clement Jonquet |
| Plant Trait Ontology PTO | Public | 1,326 | 0 | 0 | 0 | 04/24/2014 | Clement Jonquet |
| Rice Trait Ontology CO-RTO | Public | 0 | 0 | 0 | 0 | 07/08/2014 | Clement Jonquet |
| Wheat Trait Ontology CO-WTO | Public | 640 | 0 | 0 | 0 | 01/08/2015 | Clement Jonquet |

The example above shows some crop-specific ontologies uploaded by Clément Jonquet as a test. We need to investigate the display options further as there are some anomalies. This is possibly due to the creation in Crop Ontology of the relationships 'methods_of' and 'scale_of'. The links proposed hereunder illustrate what could be done on this instance of Bioportal using the Crop Ontology terms. Some are similar to what we can perform on the Crop Ontology website and others are additional. For example:

- BioPortal Search function searches within all the crop ontologies; for instance, a query with 'plant': <http://bioportal.lirmm.fr:8082/search?q=plant&apikey=aa5b0e2c-5a2e-4a01-b3a2-32cf024f3f27&ontologies=CO-BANANA,CO-CRO,CO-GO,CO-GMO,CO-MMO,CO-MCPO,CO-RTO,CO-WTO>
- Annotates a piece of text with all the crop ontologies: <http://bit.ly/1yJP4tT>.
- Formalizes, stores and renders mappings between ontologies of the portal (or between ontologies of the portal and the main NCBO BioPortal). For instance (copy and paste the link below in your web browser): http://bioportal.lirmm.fr:8082/ontologies/CO-WTO/classes/http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FCO_321%23_0000020/mappings

This is a mapping (of type skos:exactMatch), uploaded to the portal through the API between Plant height in CO-WTO and plant height in PO.

- The mapping feature could be useful for aligning Crop Ontology terms to the Planteome ontologies.

All other features from BioPortal will be generically available for the plant group: ontology versioning, UI widget, ontology metrics, ontology recommender service, projects listing, community feedback (comment, subscription to ontology changes), users' management (and public or private access to ontologies). In addition, two endpoints allow automatically querying the content of the portal:

- REST web service API: <http://bioportal.lirmm.fr:8082/documentation>
- SPARQL endpoint: <http://bioportal.lirmm.fr:8081/test/>

Indeed, the portal has fully embraced semantic web technologies and the ontologies will be stored in an RDF triple store, which is not the currently applied for the Crop Ontology.

Budget for collaboration

The Labex NUMEV offers the opportunity to obtain the **funding for a 1-year postdoc**, starting in May 2015 for joining Clément Jonquet's lab and testing NCBO technology with the Crop ontology project. The postdoc profile needs to be urgently submitted in order to start in May.

Role of Collaborators

Bioversity International : Bioversity's team will provide expertise on the Crop ontology infrastructure (concepts, features) and user needs. The recent NSF-awarded project called 'Planteome' enables a postdoc position at Bioversity Montpellier for the contribution of the Crop Ontology to the reference ontologies. Time permitting, this postdoc scientist could liaise with the NUMEV postdoc scientist to guide the BioPortal prototype development and assure the alignment with Planteome technology.

SIFR-LIRMM : Within the SIFR project, LIRMM collaborates with the NCBO to reuse BioPortal in order to develop a French version of the ontology based annotation workflow using French medical terminologies. In the SIFR project, we also investigate the multilingual issues in order to make NCBO technology multilingual.

IBC: Within the IBC project, we also reuse the NCBO technology to deploy a specific annotation web service for plants and we plan to extend it to crop ontology. The IBC project, with participation of INRIA and IRD also brings the expertise respectively on big scientific data management and on plant domain (genomics, trait, etc.), especially rice. Data from project like 3k genomes or IRIGIN (International Rice Genomics Initiative) might be used for testbeds.

Possible interaction with Planteome

This project is a prospective project and must take into consideration the newly NSF-awarded Planteome project. The interactions with Planteome can be assessed along the Bioportal project lifetime. We give hereunder some examples of such collaboration but we propose to hold a videoconference with the Planteome Principal Investigators sometime in March or early April to present our project idea and see if it would be feasible to collaborate in developing technical solutions.

- In the context of Planteome, the Crop Ontology will align the crop concepts validated by the agricultural community to the plant reference ontologies. We propose to test how the Bioportal technology could support this work?
- Planteome will enhance the web tool developed by Chris Mungall, Berkeley, called TermGenie, which uses patterns and reasoning to create new terms for an ontology. The Bioportal Plant group project could test the tool and see if it could be integrated into a workflow with the Bioportal local instance.
- An annotation tool already exists within the Bioportal. Planteome will also develop an annotation pipeline so this Bioportal project could contribute to the test of the Planteome pipeline to identify the best technical solutions.
- The Wheat Data Interoperability Group has already made suggestions for testing the Bioportal instance and they are also interested by the ontologies of Planteome and using the Crop Ontology concepts.

Key people involved

Dr. Clement Jonquet is assistant professor at [University of Montpellier](#), France. He is a researcher at the [Laboratory of Informatics, Robotics, and Microelectronics of Montpellier \(LIRMM\)](#), on (biomedical)ontologies, semantic data indexing and annotation, semantic Web, text mining, knowledge representation. Dr. Jonquet obtained PhD in Informatics from the same university in 2006 (about multi-agent systems, grid and service-oriented computing), then he served as a postdoc for 3 years at the [Stanford Center for Biomedical Informatics Research \(BMIR\)](#) within Pr. Mark A. Musen's group where he was working on semantic annotations of biomedical data using biomedical ontologies in the context of the [National Center for Biomedical Ontology \(NCBO\)](#) project. He contributed actively to the design, evolution and development of the [NCBO BioPortal](#) and won the 1st prize at ISWC [Semantic Web Challenge](#) 2010. Since 2013, Dr Jonquet is the PI of the [SIFR project](#) (Semantic Indexing of French Biomedical Data Resources) mainly funded by the [ANR Young Researcher](#) program. Dr. Jonquet is the (co)author of 42 publications, cumulating [1080+ citations](#), including 10 international journals. He is a member of several workshop and conference program committees related to life sciences and informatics. He was local chair of [ESWC 2013](#), program co-chair of [S4BIODIV 2013](#) and [TICE 2014](#).

Dr Pierre Larmande is scientist at IRD, France. Dr Larmande obtained PhD in Informatics from the university of Montpellier in 2007 (about federated plant data integration approaches applied to rice data). He coordinated the Oryza Tag Line phenotypic rice mutant database and the OrygenesDB T-DNA rice database. Since 2013, he is leading the data integration group at the Institute of Computational Biology part of the LIRMM, (plant)-ontologies, semantic data indexing and annotation, semantic web, Knowledge management, genomic Big Data management. He was the chair of the Semantics for Biodiversity workshop [S4BIODIV 2013](#) at the [ESWC 2013](#).

Elizabeth Arnaud, is a scientist at Bioversity International, Montpellier and, since 2008, is the Principal Investigator on the Integrated Breeding Platform's Crop Ontology Project. Elizabeth is co-Principal Investigator of the new NSF-awarded Planteome project. When she joined Bioversity, she rapidly become involved in the challenge of data integration, use of data standards, controlled vocabularies and metadata. She coordinated the Musa Germplasm Information System (MGIS) and later the CGIAR System-wide Information System on Genetic Resources (SINGER), as well as the development of a Bioversity geospatial database for collected crop samples. She was the chair of the 2009 conference of the Biodiversity Informatics Standards (TDWG), Montpellier, and Chair of the Semantics for Biodiversity workshop at the [ESWC 2013](#). Elizabeth is a member of the GBIF Scientific

Committee and is the head of Bioversity's delegation on the Global Biodiversity Information Facility (GBIF) Governing Body.

References

Laurel Cooper, et. al., (2012). *The Plant Ontology As A Tool For Comparative Plant Anatomy And Genomic Analyses*. Plant and Cell Physiology. 12/2012; DOI:10.1093/pcp/pcs163

Arnaud E. et. al., (2012). *Towards a Reference Plant Trait Ontology for Modeling Knowledge of Plant Traits and Phenotypes*. In Proceedings of the International Conference on Knowledge Engineering and Ontology Development, pages220-225, SciTePress. DOI: 10.5220/0004138302200225

Shrestha R. et. al., (2012) *Bridging the phenotypic and genetic data useful for integrated breeding through a data annotation using the Crop Ontology developed by the crop communities of practice*. Frontiers in Plant Physiology v. 3 Article 326: doi: 10.3389/fphys.2012.00326 , ISSN: 1664-042X

Noy, Natalya F., et al. "BioPortal: ontologies and integrated data resources at the click of a mouse." *Nucleic acids research* (2009): gkp440.

Whetzel, Patricia L., et al. "BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications." *Nucleic acids research* 39.suppl 2 (2011): W541-W545.

Jonquet, Clement, Nigam H. Shah, and Mark A. Musen. "The open biomedical annotator." *Summit on translational bioinformatics 2009* (2009): 56.

Wollbrett J, Larmande P. et. al., (2013). Clever creation of rich SPARQL queries from annotated relational schema: application to Semantic Web Service creation. BMC Bioinformatics. doi: 10.1186/1471-2105-14-126