

Crop Ontology: Vocabulary For Crop-related Concepts

Luca Matteis¹, Pierre-Yves Chibon², Herlin Espinosa³, Milko Skofic¹, Richard Finkers², Richard Bruskiewich¹, Glenn Hyman³, and Elizabeth Arnaud¹

¹ Bioersity International

Via dei Tre Denari 472/a

00057 Maccarese (Fiomicino) Rome, Italy

{l.matteis,m.skofic,r.bruskiewich,e.arnaud}@cgiar.org

<http://www.bioersityinternational.org/>

² Wageningen UR Plant Breeding

Wageningen University and Research Centre,

PO Box 386, 6700 AJ, Netherlands

py@chibon.fr, richard.finkers@wur.nl

³ CIAT, International Center for Tropical Agriculture

Km 17, Recta Cali-Palmira Apartado Areo 6713

Cali, Colombia

{h.r.espinosa,g.hyman}@cgiar.org

Abstract. A recurrent issue for data integration is the lack of a common and structured vocabulary used by different parties to describe their data sets. The Crop Ontology (www.croponontology.org) project aims to provide a central place where the crop community can gather to generate such standardized vocabularies and structure them into ontologies. Having standardized ontologies opens the world of the Semantic Web to data integration between different data providers. Crop Ontology is a community-based project, providing a central place for the creation of crop-related ontologies, but it can also be integrated into third-party tools through its Application Programming Interface, providing retrieval of specific terms or a more generic search functionality for all terms. The ontologies are available in RDF format, described using the OWL and RDFS standards, allowing them to be consumed by popular semantic reasoners. We believe that Crop Ontology will lead to better description of crop-related data, improving collaboration between partners and should serve as an example for other scientific fields.

Keywords: vocabularies, ontologies, Semantic Web, Linked Data, agricultural biodiversity, crops

1 Introduction

Over the last decade there has been a large increase in the number of online vocabularies and ontologies [1]. Search engines such as Google, Yahoo! and Bing, have agreed on a common vocabulary that describes entries in their databases.

This vocabulary is hosted on <http://schema.org>, allowing search engines to be consistent on the meaning of specific concepts. Many other vocabularies exist across the internet, and services such as <http://vocab.cc> allow searching them.

The Linked Data [2] initiative tries to link information across the web using the Semantic Web RDF⁴ technology as a basis. This framework enforces the use of URIs⁵ for uniquely identifying terms inside a vocabulary or ontology. This initiative has allowed the linkage of data across the web, leading to the construction of a major cloud of information [3].

This cloud however lacks crop-related data. One of the reasons for this, is the lack of standardized vocabularies, which would allow various data providers to describe their data in a consistent manner. Searching for crop terminology on popular ontology search engines⁶ websites, shows that very few standards exist in this field.

To build a standardized vocabulary that can be used by different data providers, data providers need to work together. Therefore the Crop Ontology has been built as a community-based project, allowing each member of the community to participate in the building of a vocabulary that matches their needs.

The website was developed as part of a formal Integrated Breeding Platform⁷ project of the Generation Challenge Programme⁸, to specify global semantic standards for germplasm information management.

2 What is Crop Ontology?

Crop Ontology (www.cropontology.org) allows browsing and searching a large database of crop-related terminology, structured per phenotype, breeding, germplasm and trait categories [4-6]. All of this information is freely accessible and downloadable directly through the website. Users can take part in enriching the Crop Ontology database: they can create an account and modify information through a wiki-like system that enables collaboration.

The key feature of this system is that it stores concepts in the form of ontologies. One of the most interesting aspects of building ontologies, instead of simply being a list of descriptors, is that they define relationships between concepts within a specific domain. As useful as this may sound to humans, it becomes even more important for computers. Because it is computers that are capable of understanding what these relationships mean, and can therefore help find information through semantic reasoners [7].

⁴ The Resource Description Framework (RDF) is a family of World Wide Web Consortium (W3C) specifications originally designed as a metadata data model. <http://www.w3.org/RDF/>

⁵ In computing, a uniform resource identifier (URI) is a string of characters used to identify a name or a web resource. <http://www.w3.org/Addressing/>

⁶ Ontology search-engine: <http://lov.okfn.org/dataset/lov/> or <http://swoogle.umbc.edu/>

⁷ IBP; <https://www.integratedbreeding.net/>

⁸ GCP; <http://www.generationcp.org/>

The ontologies are moderated by semantic experts who help model them, so that they can be consumed by popular semantic reasoners. Moderators of the system make sure everything is done correctly, using good semantic practices. It is important to use standard terminology to build ontologies. OWL [8] and RDFS [9] provide the foundation for these rules, and Crop Ontology uses them extensively.

The simple and easy-to-use interface allows users to browse these concepts through a collapsible tree interface, and search for specific terminology using a powerful free-text search engine. Users can then find concepts and provide feedback when needed. These features allow the direct participation of users in the building process of the ontologies.

3 Features

Crop Ontology aims to create a community of contributors interested in building standard ontologies for crop-related topics. In order to build this community and allow it to perform its goal, a number of features have been implemented: an ontology browser; the possibility to create, extend, and model an ontology; to modify and delete terms; to insert comments; and to programmatically access data through an RDF web service.

3.1 The ontology browser

Browsing is an essential feature of the Crop Ontology website. Users can easily explore the various vocabularies, read descriptions of their terms, and download an RDF version of them. It is simple to find their way through the different types of ontologies, and see the crops available, directly from within the homepage as shown in Figure 1.

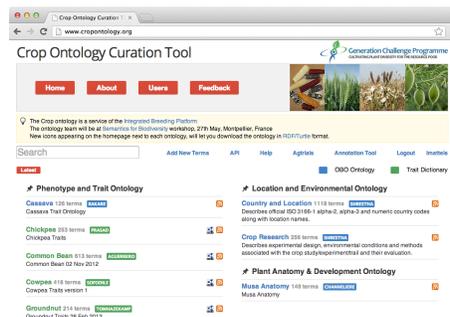


Fig. 1. Homepage of the Crop Ontology website

3.2 Create, extend and model

From the “Create an Ontology” page, as shown in Figure 2, users can immediately start experimenting with a basic interface for building ontologies. Users can create terms directly from within the website, through a dynamic collapsible tree structure. They can insert the name of concepts, and assign basic relationships to each of them, essentially allowing anyone to build a graph through a basic browser-side interface.

Upload an OBO File or **Create an Ontology** or Upload Excel Trait Template

Category: 300-499 Phenotype and Trait Ontology

Rice RiceOntology

An ontology used to describe Rice. +

Rice rdfs:subClassOf + -

StressTrait rdfs:subClassOf + -

SeedWeight rdfs:subClassOf + -

Save

Fig. 2. Web interface for creating ontologies

3.3 Modify and delete

Through the same minimalist interface the system allows also users to modify properties of specific terms. They can insert text in various languages, and upload images that allow them to better describe a concept. Crop Ontology provides simple interface components to allow anybody to modify and extend vocabularies. Figure 3 shows how “action buttons” appear at the right side of each property section, allowing users to quickly identify the action needed to modify or delete a term.

3.4 Leaving comments

Communication is one of the most important parts of community building, so in order for Crop Ontology to build its community of experts, some means of communication between the members is necessary. Under each term, a “comments” section allows users to provide feedback (Fig. 4) directly to the ontology maintainer.

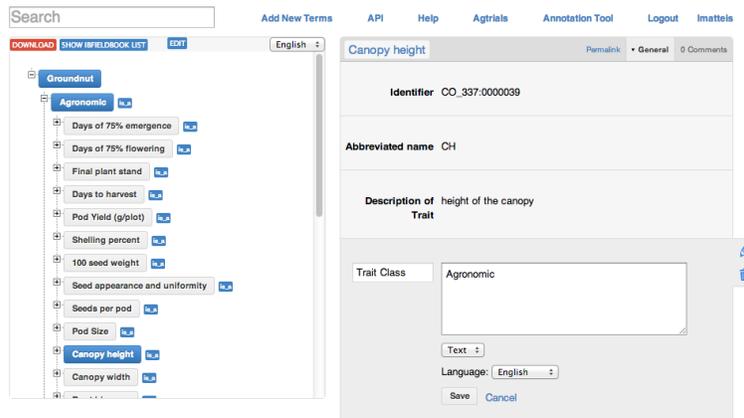


Fig. 3. Edit a term directly from the web-interface

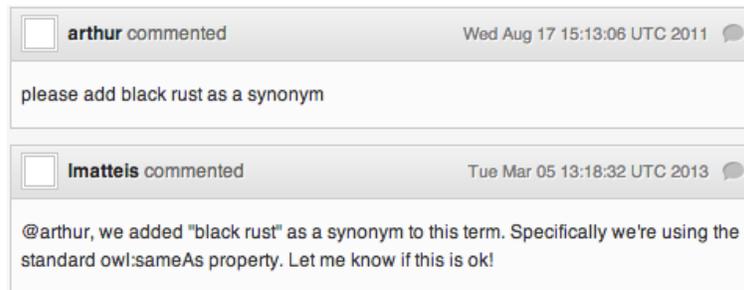


Fig. 4. Comments from the Crop Ontology website

3.5 Graph visualization

Visualizing an ontology as a graph can help the understanding of the relationship between different concepts and how these concepts are structured within their specific domain. An example of an ontology graph is shown in Figure 5.

3.6 RDF support

Crop Ontology decided to adopt the RDF framework. RDF relies on the idea that any piece of information can be described in the form of subject-predicate-object expressions, known as triples. The interesting aspect of the triples is that they are capable of universally storing and linking data: resources are described using URIs, which allows data to be identified and linked in a standard common way, using referenceable resources.

RDFS and OWL are used within the Crop Ontology as they provide standard vocabularies for defining, relating and giving meaning to concepts. By making crop-related data compliant to these standards, they can feed into other data that also use this format, and benefit from them in ways it couldn't otherwise.

<https://github.com/bioversity/Crop-Ontology>.

Anybody can use and improve this system, making it a piece of software that others can model to fit their needs.

The ontologies can be downloaded in the popular RDF/Turtle¹² format. This format is well supported by many semantic reasoners such as Apache Jena¹³, and it is possible to convert it into other RDF serialization formats if needed.

Google App Engine¹⁴ is also a major component of the Crop Ontology stack. Hosting the application on Google's cloud relieves concerns about the underlying hardware of the computers that are running the software. This gives us more time to concentrate on the development of the product itself, without concerns regarding system administration tasks.

The cloud also provides greater scalability. Many servers are instantiated based on the request load. This essentially makes the system resilient to high-traffic demand, and more resistant against brute-force attacks.

5 Conclusions

Linked Data, and all the technology behind it, is clearly the foundation for data integration of various different information resources. Providing a simple user-interface, such as Crop Ontology, to novice users who are not familiar with all the technologies involved, has proved to be a useful exercise. It has given users the capacity to transform their databases, that were hidden behind personalized schemas, into sharable and linkable resources.

Crop communities are going to continue being involved in the creation of crop-related vocabularies. There are huge numbers of crops that have not been described, and a great deal of information that has not been annotated. The work of bringing more species and more groups into the picture is critical for the continued success of the Crop Ontology. Apart from the Integrated Breeding Platform, many other crop data providers have expressed their interest in us-

¹² Turtle (Terse RDF Triple Language) is a format for expressing data in the Resource Description Framework (RDF) data model. <http://www.w3.org/TeamSubmission/turtle/>

¹³ Jena provides a collection semantic tools and Java libraries. <http://jena.apache.org/>

¹⁴ Google App Engine is a platform as a service (PaaS) cloud computing platform for developing and hosting web applications in Google-managed data centers. <https://developers.google.com/appengine/>

ing the Crop Ontology: AgTrials¹⁵, GENESYS¹⁶ and GRIN-Global¹⁷ are in the process of making their data available as RDF resources, with proper linkages to Crop Ontology, allowing it to be linked and discoverable within the Semantic Web.

The system will continue growing with new features also thanks to the open-source community behind it, which constantly feeds the project with fixes and improvements. The future roadmap for the project development includes better integration with richer OWL sublanguages such as OWL DL¹⁸, which allows for greater expressiveness and more complex relationships of the ontologies.

Finally we think that Crop Ontology not only is a useful software system capable of modeling generic ontologies, but in the context of agricultural biodiversity it also provides a meeting ground for various crop communities to discuss and build the next generation of standard crop vocabularies, which are an essential component for the future of biodiversity data management and discoverability.

6 Acknowledgments

The authors would like to thank the data providers who have contributed to submitting data to the Crop Ontology: Peter Kulakow, Bakare Moshood, Sam Ofodile, Ousmane Boukare, Antonio Lopez Montes (IITA); Trushar Shah, Prasad Peteti, Praveen R Reddy, Ibrahim Sissoko, Eva Weltzien, Isabel Vales, Suyah Patil (ICRISAT); Reinhard Simon (CIP); Inge van den Bergh, Stephanie Channeliere (Bioversity International); Mauleon Ramil, Nikki Borgia, Ruairaidh Sackville-Hamilton (IRRI); Alberto Fabio Guerero, Steve Beebe, Roland Chirwa (CIAT). We would also like to thank Rosemary Shrestha and Thomas Hazekamp for providing technical expertise in the field of ontology development, and Arwen Bailey (Bioversity International) for her editorial support. Finally we thank Generation Challenge Programme (GCP) for providing the fund for this collaborative Crop Ontology development and implementation project work.

References

1. Vatan, B., Vandenbussche, P.: <http://lov.okfn.org/dataset/lov/stats/> (2013)
- ¹⁵ AgTrials is an information portal developed by the CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS). <http://www.agtrials.org/>
- ¹⁶ GENESYS is an important and very rich source of information on plant genetic resources diversity of seeds conserved in Genebanks worldwide, crops and crop wild relative material. <http://www.genesys-pgr.org/>
- ¹⁷ GRIN-Global provides the worlds crop genebanks with a powerful, flexible, easy-to-use global plant genetic resource information management system. <http://www.grin-global.org/>
- ¹⁸ OWL DL supports those users who want the maximum expressiveness while retaining computational completeness. <http://www.w3.org/TR/2004/REC-owl-features-20040210/#s2.2>

2. Berners-Lee, T.: <http://www.w3.org/DesignIssues/LinkedData.html> (2006)
3. Heath, T.: <http://lod-cloud.net/> (2013)
4. Richard Bruskiewich, Guy Davenport, Tom Hazekamp, Thomas Metz, Manuel Ruiz, Reinhard Simon, Masaru Takeya, Jennifer Lee, Martin Senger, Graham McLaren, and Theo Van Hintum. 2006. The Generation Challenge Programme (GCP)-Standards for Crop Data. *OMICS* 10(2):215-219
5. Rosemary Shrestha, Elizabeth Arnaud, Ramil Mauleon, Martin Senger, Guy F. Davenport, David Hancock, Norman Morrison, Richard Bruskiewich and Graham McLaren. 2010. Multifunctional crop trait ontology for breeders' data: field book, annotation, data discovery and semantic enrichment of the literature. *AoB PLANTS* 2010 doi: 10.1093/aobpla/plq008
6. Rosemary Shrestha, Luca Matteis, Milko Skofic, Arlet Portugal, Graham McLaren, Glenn Hyman and Elizabeth Arnaud. 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. *Front. Physiol*— doi: 10.3389/fphys.2012.00326
7. R. Mishra and S. Kumar. Semantic web reasoners and languages. *Artificial Intelligence Review*, 2010. DOI 10.1007/s10462-010-9197-3
8. Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D.L., PatelSchneider, P.F., Stein, L.A.: *OWL Web Ontology Language Reference*. Technical Report <http://www.w3.org/TR/2004/REC-owl-ref-20040210/>, W3C (2004)
9. Brickley, D., R.V. Guha: <http://www.w3.org/TR/rdf-schema/> (2004)
10. Gardler, R.: <http://news.slashdot.org/story/13/01/29/2252237/how-open-source-could-benefit-academic-research> (2013)