Computers and Electronics in Agriculture 144 (2018) 126–143

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

AgroPortal: A vocabulary and ontology repository for agronomy

Clément Jonquet[a,b,f,*], Anne Toulet[a,b], Elizabeth Arnaud[c], Sophie Aubin[d], Esther Dzalé Yeumo[d], Vincent Emonet[a], John Graybeal[f], Marie-Angélique Laporte[c], Mark A. Musen[f], Valeria Pesce[g], Pierre Larmande[b,e]

[a] Laboratory of Informatics, Robotics and Microelectronics of Montpellier (LIRMM), University of Montpellier & CNRS, France
[b] Computational Biology Institute (IBC) of Montpellier, France
[c] Bioversity International, Montpellier, France
[d] INRA Versailles, France
[e] UMR DIADE, IRD Montpellier, France
[f] Center for BioMedical Informatics Research (BMIR), Stanford University, USA
[g] Global Forum on Agricultural Research (GFAR), Food and Agriculture Organization (FAO) of the United Nations, Rome, Italy

ARTICLE INFO

Keywords:
Ontologies
Controlled vocabularies
Knowledge organization systems or artifacts
Ontology repository
Metadata
Mapping
Recommendation
Semantic annotation
Agronomy
Food
Plant sciences
Biodiversity

ABSTRACT

Many vocabularies and ontologies are produced to represent and annotate agronomic data. However, those ontologies are spread out, in different formats, of different size, with different structures and from overlapping domains. Therefore, there is need for a common platform to receive and host them, align them, and enabling their use in agro-informatics applications. By reusing the National Center for Biomedical Ontologies (NCBO) BioPortal technology, we have designed AgroPortal, an ontology repository for the agronomy domain. The AgroPortal project re-uses the biomedical domain's semantic tools and insights to serve agronomy, but also food, plant, and biodiversity sciences. We offer a portal that features ontology hosting, search, versioning, visualization, comment, and recommendation; enables semantic annotation; stores and exploits ontology alignments; and enables interoperation with the semantic web. The AgroPortal specifically satisfies requirements of the agronomy community in terms of ontology formats (e.g., SKOS vocabularies and trait dictionaries) and supported features (offering detailed metadata and advanced annotation capabilities). In this paper, we present our platform's content and features, including the additions to the original technology, as well as preliminary outputs of five driving agronomic use cases that participated in the design and orientation of the project to anchor it in the community. By building on the experience and existing technology acquired from the biomedical domain, we can present in AgroPortal a robust and feature-rich repository of great value for the agronomic domain.

1. Introduction

Agronomy, food, plant sciences, and biodiversity are complementary scientific disciplines that benefit from integrating the data they generate into meaningful information and interoperable knowledge. Undeniably, data integration and semantic interoperability enable new scientific discoveries through merging diverse datasets (Goble and Stevens, 2008). A key aspect in addressing semantic interoperability is the use of ontologies as a common and shared means to describe data, make them interoperable, and annotate them to build structured and formalized knowledge. Biomedicine has always been a leading domain encouraging semantic interoperability (Rubin et al., 2008). The domain has seen success stories such as the Gene Ontology (Ashburner et al., 2000), widely used to annotate genes and their products. And other disciplines have followed, developing among others the Plant Ontology (Cooper et al., 2012), Crop Ontology (Shrestha et al., 2010), Environment Ontology (Buttigieg et al., 2013), and more recently, the Agronomy Ontology (Devare et al., 2016), TOP Thesaurus (Garnier et al., 2017), Food Ontology (Griffiths et al., 2016), the IC-FOODS initiative's ontologies (Musker et al., 2016), and the animal traits ontology (Hughes et al., 2014). Ontologies have opened the space to various types of semantic applications (Meng, 2012; Walls et al., 2014), to data integration (Wang et al., 2015), and to decision support (Lousteau-Cazalet et al., 2016). Semantic interoperability has been identified as a key issue for agronomy, and the use of ontologies declared a way to address it (Lehmanna et al., 2012).

Communities engaged in agronomic research often need to access specific sets of ontologies for data annotation and integration. For instance, plant genomics produces a large quantity of data (annotated genomes), and ontologies are used to build databases to facilitate cross-

* Corresponding author at: 161 Rue Ada, 34090 Montpellier, France.
E-mail address: jonquet@lirmm.fr (C. Jonquet).

https://doi.org/10.1016/j.compag.2017.10.012
Received 27 October 2016; Received in revised form 12 May 2017; Accepted 13 October 2017

species comparisons (Jaiswal, 2011). More recently, the focus of many scientific challenges in plant breeding has switched from genetics to phenotyping, and standard traits/phenotypes vocabularies have become necessary to facilitate breeders' data integration and comparison. In parallel with very specific crop dictionaries (Shrestha et al., 2010), important organizations have produced large reference vocabularies such as Agrovoc (Food and Agriculture Organization) (Sachit Rajbhandari, 2012), the NAL Thesaurus (National Agricultural Library), and the CAB Thesaurus (Centre for Agricultural Bioscience International).[1] These thesauri are primarily used to index information resources and databases. As more vocabularies and ontologies[2] are produced in the domain, the greater the need to discover them, evaluate them, and manage their alignments (d'Aquin and Noy, 2012).

However, while great efforts have taken place in the biomedical domain to harmonize content (e.g., the *Unified Medical Language System* (UMLS), mostly for medical terminologies) (Bodenreider, 2004) and ontology design principles (e.g., the OBO Foundry, containing mostly biological and biomedical ontologies) (Smith et al., 2007), ontologies in agriculture are spread out around the web (or even unshared), in many different formats and artifact types, and with different structures. Agronomy (and its related domains such as food, plant sciences, and biodiversity) needs an one-stop shop, allowing users to identify and select ontologies for specific tasks, as well as offering generic services to exploit them in search, annotation or other scientific data management processes. The need is also for a community-oriented platform that will enable ontology developers and users to meet and discuss their respective opinions and wishes. This need was clearly expressed by stakeholders in various roles (developers, database maintainers, and researchers) across many community meetings, such as: 1st International Workshop for Semantics for Biodiversity in 2013 (http://semantic-biodiversity.mpl.ird.fr) (Larmande et al., 2013); the "Improving Semantics in Agriculture" workshop in 2015 (Baker et al., 2015); or several meetings of the Agricultural Data Interest Group (IGAD) of the Research Data Alliance.

These motivations prompted us to build a vocabulary and ontology repository to address these needs. In this paper, we present the AgroPortal project, a community effort started by the Montpellier scientific community to build an ontology repository for the agronomy domain. Our goal is to facilitate the adoption of metadata and semantics to facilitate open science in agronomy. By enabling straightforward use of agronomical ontologies, we let data managers and researchers focus on their tasks, without requiring them to deal with the complex engineering work needed for ontology management. AgroPortal offers a robust and reliable service to the community that provides ontology hosting, search, versioning, visualization, comment, and recommendation; enables semantic annotation; stores and exploits ontology alignments; and enables interoperation with the semantic web. Our vision is to facilitate the integrated use of all vocabularies and ontologies related to agriculture, regardless of their source, format, or content type.

In order to capitalize on what is already available in other communities, we have reused the openly available NCBO BioPortal technology (http://bioportal.bioontology.org) (Noy et al., 2009; Whetzel et al., 2011) to build our ontology repository and services platform. BioPortal was originally dedicated to health, biology and medicine and has some content related to agriculture, but the portal does covers few of the facets of agronomy, food, plant sciences and biodiversity, let alone environment and animal sciences. Therefore, many in the agronomy community do not see themselves as users targeted by BioPortal. For instance, the Crop Ontology is listed on the NCBO BioPortal (along with other top-level plant-related ontologies), but is not currently fully accessible and described through this portal; none of the crop specific ontologies are available. In addition to its core repository of ontology mission, the NCBO technology also offers many applicable tools, including a mapping repository, an annotator, an ontology recommender, community support features, and an index of annotated data. All these services are reused and customized within AgroPortal to benefit its target user community.[3] Furthermore, our vision was to adopt, as the NCBO did, an open and generic approach where users can easily participate to the platform, upload content, and comment on others' content (ontologies, concepts, mappings, and projects). As explained below, we determined that the NCBO technology (Whetzel and Team, 2013) implemented the greatest number of our required features, while recognizing the technical challenges of adopting such a various and complex software.

In the following sections, we offer extensive descriptions of AgroPortal's features. We will focus on how they address community requirements expressed within five agronomic driving use cases involving important research organizations in agriculture such as Bioversity International (CGIAR), French INRA, and United Nations FAO. The rest of the paper is organized as follows: In Section 2, we review related work in ontology repositories in relation to our domain of interest. Section 3 describes the requirements of AgroPortal's initial five driving agronomic use cases. Section 4 presents our platform by extensively describing its content, as well as its features (both inherited from the NCBO BioPortal, and added by us). Section 5 analyzes how our initial five driving use case results benefit from AgroPortal. Finally, Section 6 provides a discussion of the contributions of AgroPortal, and Section 7 presents our conclusions.

2. Background and related work

With the growing number of developed ontologies, ontology libraries and repositories have been of interest in the semantic web community. Ding and Fensel (2001) presented in 2001 a review of ontology libraries that introduced the notion of "library." Then Hartman et al. Baclawski and Schneider (2009) introduced the concept of ontology repository, with advanced features such as search, metadata management, visualization, personalization, and mappings. By the end of the 2000's, the Open Ontology Repository Initiative (Baclawski and Schneider, 2009) was a collaborative effort to develop a federated infrastructure of ontology repositories.[4] d'Aquin and Noy (2012) provided the latest review of ontology repositories in 2012.

In the biomedical or agronomic domains there are several standards or knowledge organization systems libraries (or registries) such as FAIRSharing (http://fairsharing.org) Sansone et al., 2012, the FAO's VEST Registry (http://aims.fao.org/vest-registry), and the agINFRA linked data vocabularies (vocabularies.aginfra.eu) (Pesce et al., 2013). They usually register ontologies and provide a few metadata attributes about them. However, because they are registries not focused on vocabularies and ontologies, they do not support the level of features that an ontology repository offers. In the biomedical domain, the OBO Foundry (Smith et al., 2007) is a reference community effort to help the

[1] http://aims.fao.org/agrovoc, https://agclass.nal.usda.gov and http://www.cabi.org/cabthesaurus

[2] In this paper, we often use the word "ontologies" or "vocabularies and ontologies" to include ontologies, vocabularies, terminologies, taxonomies and dictionaries. We acknowledge the differences (not discussed here) in all these types of Knowledge Organization Systems (KOS) or knowledge artifacts. The reader may refer to McGuinness's discussion (McGuinness, 2003). While being an "ontology repository", AgroPortal handles all these artifact types, if they are compatibly formatted. While AgroPortal thereby enables horizontal use of these artifact types with common user interface and application programming interface, it does not leverage the full power of ontologies (e.g., reasoning), instead map all the imported artifact types to a "common simplified model."

[3] Except the "NCBO Resource Index" component, a database of 50+ biomedical resources indexed with ontology concepts (Jonquet et al., 2011) that we have not reused in AgroPortal because we work with the AgroLD use case to fulfill the mission of interconnecting ontologies and data.

[4] At that time, the effort already reused the NCBO technology that was open source, but not yet packaged in an appliance as it is today.