HAL
open science

Harnessing the power of unified metadata in an ontology repository: the case of AgroPortal

Clement Jonquet, Anne Toulet, Biswanath Dutta, Vincent Emonet

To cite this version:

Clement Jonquet, Anne Toulet, Biswanath Dutta, Vincent Emonet. Harnessing the power of unified metadata in an ontology repository: the case of AgroPortal. Journal on Data Semantics, 2018, pp.1-31. 10.1007/s13740-018-0091-5 . lirmm-01852080v2

HAL Id: lirmm-01852080

https://hal-lirmm.ccsd.cnrs.fr/lirmm-01852080v2

Submitted on 31 Aug 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Journal on Data Semantics
https://doi.org/10.1007/s13740-018-0091-5

ORIGINAL ARTICLE

Harnessing the Power of Unified Metadata in an Ontology Repository: The Case of AgroPortal

Clement Jonquet[1,2] · Anne Toulet[1] · Biswanath Dutta[3] · Vincent Emonet[1]

Received: 20 July 2017 / Revised: 20 July 2018 / Accepted: 2 August 2018
© The Author(s) 2018

Abstract

As any resources, ontologies, thesaurus, vocabularies and terminologies need to be described with relevant metadata to facilitate their identification, selection and reuse. For ontologies to be FAIR, there is a need for metadata authoring guidelines and for harmonization of existing metadata vocabularies—taken independently none of them can completely describe an ontology. Ontology libraries and repositories also have to play an important role. Indeed, some metadata properties are intrinsic to the ontology (name, license, description); other information, such as community feedbacks or relations to other ontologies are typically information that an ontology library shall capture, populate and consolidate to facilitate the processes of identifying and selecting the right ontology(ies) to use. We have studied ontology metadata practices by: (1) analyzing metadata annotations of 805 ontologies; (2) reviewing the most standard and relevant vocabularies (23 totals) currently available to describe metadata for ontologies (such as Dublin Core, Ontology Metadata Vocabulary, VoID, etc.); (3) comparing different metadata implementation in multiple ontology libraries or repositories. We have then built a new metadata model for our AgroPortal vocabulary and ontology repository, a platform dedicated to agronomy based on the NCBO BioPortal technology. AgroPortal now recognizes 346 properties from existing metadata vocabularies that could be used to describe different aspects of ontologies: intrinsic descriptions, people, date, relations, content, metrics, community, administration, and access. We use them to populate an internal model of 127 properties implemented in the portal and harmonized for all the ontologies. We—and AgroPortal's users—have spent a significant amount of time to edit and curate the metadata of the ontologies to offer a better synthetized and harmonized information and enable new ontology identification features. Our goal was also to facilitate the comprehension of the agronomical ontology landscape by displaying diagrams and charts about all the ontologies on the portal. We have evaluated our work with a user appreciation survey which confirms the new features are indeed relevant and helpful to ease the processes of identification and selection of ontologies. This paper presents how to harness the potential of a complete and unified metadata model with dedicated features in an ontology repository; however, the new AgroPortal's model is not a new vocabulary as it relies on preexisting ones. A generalization of this work is studied in a community-driven standardization effort in the context of the *RDA Vocabulary and Semantic Services Interest Group*.

Keywords Ontology metadata vocabulary · Semantic description · Ontology repository · Ontology selection · Ontology relation · BioPortal · AgroPortal

✉ Clement Jonquet
jonquet@lirmm.fr

Anne Toulet
toulet@lirmm.fr

Biswanath Dutta
bisu@drtc.isibang.ac.in

Vincent Emonet
emonet@lirmm.fr

[1] Laboratory of Informatics, Robotics and Microelectronics of Montpellier (LIRMM), CNRS and University of Montpellier, Montpellier, France

[2] Center for Biomedical Informatics Research (BMIR), Stanford University School of Medicine, Stanford, CA, USA

[3] Documentation Research and Training Centre (DRTC), Indian Statistical Institute, Bangalore, India

Published online: 29 August 2018

1 Introduction

In 2007, Swoogle's homepage [1] announced searching over 10.000 ontologies. Today, a simple Google Search for "filetype:owl" returns around 34 K results. How much ontologies are available online now? The big data deluge and the adoption of the semantic web to semantically describe and link these data [2] have made the number of ontologies grow to numbers for which machines are mandatory to index, search and select them. It has become cumbersome for domain experts to identify the ontologies to use so that automatic recommender systems have been designed to help them with this task, as for instance in the biomedical domain [3]. However, machines need metadata to facilitate the exploitation of any data, including ontologies. It is established that metadata is often too much neglected by data providers [4] even if it is now identified as a requirement to make the data FAIR [5]. But as any other data, ontologies have themselves to be *Findable, Accessible, Interoperable, and Re-usable*. Although there are multiple dimensions to make ontologies FAIR, one will agree developing open ontology repositories, and libraries is one of them. Such libraries are the best environment in which the metadata about ontologies can be described and valued. However, can we say that ontology developers describe their ontologies with relevant metadata properties that will facilitate manual or automatic search, identification and selection of ontologies? There exists a significant number of metadata vocabularies that could be used for ontologies but none of the existing ones can completely meet this need if taken independently. Therefore, how can we make ontologies more FAIR?

When someone is interested in an ontology, he/she may like to know: Who edited or contributed? When? What methodology or tool was used? Which natural language is used? Which formats are available? What is the metrics? Is it free of use or licensed? Who is using it? In addition, when someone is interested about ontologies of a domain, he/she may like to know: How ontologies can be grouped together? Which are most used? What are the relations between them? What are the common practices? Who are the key contributors of the domain? Or the most important organizations? All this information can be represented by metadata properties. Capturing that information is both a technical challenge—we need models, tools and automated population—and a data curation challenge. Indeed, the information or metadata about an ontology is often dispatched within web sites, scientific articles, documentation or sometimes not existing at all except in the brain of the original ontology developers. There is a need for metadata authoring guidelines and for harmonization of existing metadata vocabularies to simplify their use and enlarge their adoption. For instance, the recent *Minimum Information for Reporting of an Ontology* initiative (https://github.com/owlcs/miro) [6] proposes the MIRO guidelines to ontology developers when reporting an ontology, e.g., in a scientific article.

In this paper, we adopt the perspective of designers of an ontology repository and report on our effort to develop a unified ontology metadata model for this repository. We measure its impact on facilitating ontology descriptions, identification and selection. In the following, we will review the current practices related to describing ontologies and using ontology metadata vocabularies. We have observed some limitations, lack of harmonization and confusions in the practices. This is not surprising when considering the efforts needed to just identify the potentially relevant vocabularies that could be used to describe ontologies.[1] Indeed, a few of these vocabularies are dedicated to ontologies and vocabularies (e.g., OMV, DOOR, VOAF), or datasets (e.g., VOID, DCAT, SCHEMA) and others capture more general metadata (e.g., DC, DCT, PROV, DOAP).[2] They are often not maintained anymore, sometimes very specific or too general and of course, they are rarely aligned one another despite their significant overlaps. Furthermore, there have been several ontology repository projects that did not also take the problem seriously enough to support the description of their ontologies with standard vocabularies [7, 8]. With the exception of the Linked Open Vocabularies registry [9, 10], the MMI Ontology Registry and Repository [11], and to some extent, the NCBO BioPortal [12], the question of harmonization and standardization of ontology descriptions have not really been a central matter, although this is changing now (e.g., the OBO Foundry community metadata effort). The Linked Open Vocabularies is a good counter example; it has developed and adopted VOAF as a unified model to describe metadata and relations between vocabularies. Now, even if the metadata vocabulary is limited (16 properties), the platform has more than 600 resources described with the same model.

In the rest of the paper, we will adopt a definition of metadata including anything that can be said to describe an ontology, structured data or free descriptions: how and why it is built, used, changed, accessed and how it relates to other ontologies and datasets. That will include properties going from (1) intrinsic properties, e.g., name, URI, creation date; (2) relation to other ontologies, e.g., imports,

[1] In this paper, we will consider the terms ontologies, terminologies, thesaurus and vocabularies as the type of knowledge organization systems [42] or knowledge artifacts [41]. Those are the subjects we are interested in describing. However, to facilitate the reading, we will use the word *ontology* to identify the subject that is described by metadata (e.g., Movie Ontology, Human Disease Ontology, MeSH thesaurus, etc.) and the word *vocabulary* to identify the semantic resources used to described ontologies (e.g., OMV, DC, DCAT, etc.).

[2] Please refer to column 'prefix' of Table 3 all along the paper for acronyms definitions of metadata vocabularies. We will consistently use upper case acronyms corresponding to the vocabulary namespace throughout the paper to refer vocabularies.

is mapped to, disagrees with; (3) community contributions, e.g., notes, project using, endorsements; (4) content-based properties, e.g., SPARQL endpoint, bulk RDF download, search endpoint. As discussed in the paper, such information when available and properly harmonized facilitates the ontology identification and selection processes, which has been assessed as crucial to enable ontology reuse [6, 13–15].[3] In addition, good and harmonized metadata provides information about the ontology landscape, especially when looking at a specific domain. For instance, when looking at the OBO Foundry ontologies [16], one may ask himself (1) if *OBO Edit* is actually the most used tool to develop ontologies stored in the foundry? (2) Who are the key persons in this community to talk to when starting a new ontology? (3) Which are the most involved organizations? (4) Which are the most active ontologies?

In this paper, we have made a systematic review of metadata vocabularies and their properties in order to build a list of metadata properties that can be used to describe ontologies inside our own ontology repository. The objective of this work is not to propose another "vocabulary" for ontology metadata, i.e., a SKOS or OWL resource that we would promote as a new standard to reuse in any ontology description. Indeed, our list relies completely on preexisting vocabularies (cf. discussion in Sect. 7.1). Our objective was to address the need of a common metadata model inside an ontology repository, i.e., implementing a way to compare ontologies side by side and describe the global landscape of all the ontologies in a library or repository.

The list proposed has been built following an analysis of current ontology metadata practices:

- We have reviewed the most standard and relevant vocabularies (23 totals, e.g., Dublin Core, VOID, Ontology Metadata Vocabulary, Data Catalog Vocabulary, etc.) to describe metadata for ontologies. For each of these vocabularies, we have selected the significant properties to describe objects that an ontology could be considered a certain type of, e.g., dataset, an asset, a project or a document. For instance, an ontology may be seen as a prov:Entity object and then the property prov:wasGeneratedBy may then be used to describe its provenance.
- We have reviewed the current use of metadata vocabularies by sampling 805 ontologies and measuring which vocabularies (and which properties in those vocabularies) are actually used by ontology developers.
- We have studied some of the most common ontology repositories available in the semantic web community, and especially the NCBO BioPortal (which is the reference platform to host and retrieve biomedical ontologies worldwide) to capture in our list, the properties that were actually implemented by the repositories but that would represent an information not specific to the portal. We have considered the features/properties implemented by the portal as "another vocabulary" (later called BioPortal Metadata) incorporated into our list.

As the result, we obtained a list of 346 relevant properties to describe different aspects of ontologies that we have categorized for better understanding. Someone developing an ontology will of course not have to fill them all but can consider them as a list of candidate properties to use. We then grouped those properties into a unified and simplified model of 127 properties that includes the 46 properties originally offered by the NCBO BioPortal and reuses properties of the reviewed metadata vocabularies for the rest [17]. We have implemented this new ontology metadata model within AgroPortal [18], an ontology repository, based on the NCBO technology. AgroPortal hosts ontologies and offers ontology-based services for agronomy, food, plant sciences and biodiversity domains. AgroPortal's new metadata model supports much more metadata properties than the original NCBO one, enabling very precise description of ontologies. For instance, the model captures which kind of knowledge organization system the file uploaded to the portal is (e.g., thesaurus, ontology, taxonomy, terminology, etc.). We also have properties to capture information such as licenses, ontology editor used, syntax, etc. We can also capture how ontologies are related to other resources (web site, publication, wiki, datasets, etc.) and other ontologies. Most metadata are automatically extracted from the original ontology file, if present, when the ontology is uploaded to the portal. Or it can be in some cases automatically generated by the portal. We have completely refactored the AgroPortal ontology metadata edition page to facilitate the job to ontology developers when uploading an ontology to the portal and manually editing metadata.

With a new edition interface and a common model available for all the ontologies in the portal, we have then spent a significant amount of time to edit and curate ourselves ontology descriptions, and we have asked the ontology developers to validate our edits and complete them. This has resulted in our capability to automatically aggregate information about ontologies and vocabularies to facilitate the comprehension of the whole agronomical ontology landscape by displaying diagrams, charts and networks about all the ontologies on the portal (grouping, types of ontologies, average metrics, most frequent licenses, languages or formats, leading contributors and organizations, most active ontologies, etc.). We have

[3] In this paper, we define *identification and selection* of an ontology as the processes of choosing the right ontology for a given task when searching for ontologies on an ontology library or repository. It can be based on the content of the ontology, its type, community or level of adoption in a community. Sometime this process may be semi-automatized with tools such as the NCBO Recommender (also available in AgroPortal) [3].

added several new features to AgroPortal's ontology description and browsing pages and have now a specific page dedicated to visualizing the "landscape" of ontologies (http://agroportal.lirmm.fr/landscape) that displays synthetized information, using diagrams, charts and figures, about the ontologies developed in agronomy with the goal of facilitating ontology identification, selection and get a better comprehension of the landscape of ontologies. Of course, these new functionalities rely on the quality of the metadata extracted from the ontologies or edited on the portal. Such visualizations are also meant to motivate the ontology developers to document and describe more their ontologies. An evaluation survey conducted with AgroPortal's users shows evidence of the influence of ontology metadata on ontology identification and selection and reports on the very positive evaluation of the new functionalities by AgroPortal's users.

The rest of the paper is organized as follows: Sect. 2 presents a few motivating use cases from our work on ontology repositories; Sect. 3 discusses related work in metadata vocabularies and ontology libraries. In Sect. 4, we report on our analysis of current ontology metadata practices that have driven our methodology, described Sect. 5, to select a large list of properties and to implement a restricted and unified new ontology metadata model in AgroPortal. Section 6 presents the results obtained by implementing the new model in AgroPortal, populating the metadata and designing new interfaces to facilitate the comprehension of the ontology landscape. The section also reports about evaluating the new features with AgroPortal's user community. Sections 7 and 8, respectively, discuss the perspectives and issues in ontology metadata and concludes the paper.

2 Motivating Use Cases

Our work on ontology metadata is related to our research and development on ontology repositories. Indeed, LIRMM develops and maintains two ontology repositories which are based on the NCBO technology [19]. One, the SIFR BioPortal (http://bioportal.lirmm.fr) is developed within the context of the *Semantic Indexing of French biomedical Resources* project and focus on French biomedical ontologies and terminologies. The main goal of the SIFR project is to develop a French Annotator [20] similar to what exists within the NCBO BioPortal [21]. The second ontology repository, AgroPortal (http://agroportal.lirmm.fr) [18], targets the agricultural community (not restricted to any language but using English as default) and the project has for primary mission to host and describe vocabularies and ontologies. In the paper, we will only describe the use cases and implementation done within the AgroPortal project; however, it is important to note that this work is generic and has also been implemented in the SIFR BioPortal.

Data integration and semantic interoperability in agronomy—and related domains—have become a crucial scientific challenge. Recently, the research community as adopted 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 [22, 23]. The FAIR principles also reinforced that vision [5]. AgroPortal's main objective is to be a reference ontology repository for agronomy, plant sciences, biodiversity, and nutrition. We reused the openly available NCBO BioPortal technology (http://bioportal.bioontology.org) [12] to build our first ontology repository and services platform. We have now an advanced prototype, and the latest version (v1.4) was released in July 2017. It currently hosts 100 public semantic resources, with more than 2/3 of them not present in any similar ontology repository (like NCBO BioPortal) and 8 privates. Today, AgroPortal offers a robust and reliable service to the community that features ontology hosting, search, versioning, visualization, comment, services for semantically annotating data with the ontologies, as well as storing and exploiting ontology alignments and data annotations.

Among the first feedbacks and requirements of new users were the ability to describe ontology metadata with additional fields that what BioPortal originally provided. For instance, the RDA Wheat Data Interoperability (WDI) working group (http://ist.blogs.inra.fr/wdi) recommendations [24] pointed to AgroPortal to find standard wheat-related ontologies, but they needed licensing and access rights information to be more explicit and consistent. The group also required that the endorsement of the WDI for certain ontologies shall be made explicit on AgroPortal, in order to encourage the reuse of some specific ontologies. The LovInra initiative (http://lovinra.inra.fr) at the French National Institute for Agricultural Research (INRA) adopted AgroPortal to publish vocabularies produced or co-produced by INRA scientists and foster their reuse beyond the original researchers. They needed to classify knowledge artifacts by types, formats, syntax, and formality.

Besides the "simple addition" of new metadata fields to the original model, the needs expressed by the early AgroPortal adopters were also related to the relations between ontologies and how would the repository help figuring out which ontologies to use. We may cite two concrete examples:

- Several ontologies are developed in parallel to capture wheat (or soy) phenotypes.[4] It became important for AgroPortal to capture the maximum information about the

[4] The Wheat Phenotype ontology [69] and IBP Wheat Trait Ontology developed within the Crop Ontology project [70]. Similarly, the Soy Ontology developed by the curator of the SoyBase database (www.soybase.org) and Soybean ontology also developed in the Crop Ontology project.

ontologies to make explicit to the community which ontology to use depending on their situation. New information such as the organization endorsing or supporting an ontology or the relation between the ontologies are useful metadata in that case.
- Ontologies are never developed isolated. Sometimes capturing the relations between the ontologies is quite cumbersome. For instance, the Planteome project [25] develops reference ontologies for plants such as the Plant Ontology and Plant Trait Ontology. The latter is connected to the specific crop trait ontologies developed within the Crop Ontology project [26]. In addition, they all use Gene Ontology [27] and Phenotype And Trait Ontology [28] to annotate gene products and qualify their phenotypes.

We will show throughout the paper how our new ontology metadata model and realization within AgroPortal help to answer these needs.

3 Related Work in Ontology Metadata Description

Metadata is generally described as the data about the data. The topic of ontology or vocabulary metadata is a subset of metadata research in general [4, 29]. In Sect. 4.1, we list metadata vocabularies reviewed from the literature; in the following, we only focus on general papers and references on the subject.

According to Obrst et al. [30], a metadata vocabulary must include a wider range of metadata features. For instance, metadata from a development perspective consists of information such as competency questions, ontological commitments, and design decisions; metadata from an implementation perspective consists of information for reasoning support, languages, rules, conformance to external standards and so forth. Properly defined ontology metadata has been a motivation of several applications of ontologies such as design of ontology repositories and libraries [12, 16, 31–33], ontology selection [34] automatic production of documentation [35], ontology sharing [36].

Capturing the metadata about "electronic objects" has been the original motivation of the DCMI [37] and multiple standardization bodies.[5] The *Dublin Core* (DC) and *DCMI Metadata Terms* (DCT) are the results of these initiatives. Today, semantically rich metadata is identified as one of requirements to produce FAIR data [5] and it becomes the core mission of research projects such as the Center for Expanded Data Annotation and Retrieval [38] which tackles the challenge of authoring and predicting biomedical datasets metadata.

An important effort has been made in the recent years to define vocabularies for datasets. The Semantic Web Health Care and Life Sciences (HCLS) working group of the W3C have produced a community profile which reviews many of them and proposes a set of recommendations when describing datasets [39]. The FAIRsharing.org action also builds a database of "data and metadata standards, inter-related to databases and data policies" [40] to which AgroPortal's content is now automatically pushed. More recently, the BioSchemas initiative (http://bioschemas.org) has also started a community effort to extend Schema.org with metadata properties that would be relevant for life sciences data. Although we do believe ontologies can somehow be seen as "datasets"—often the closest objects in vocabularies—they have some particularities that require more specific metadata vocabularies as we will see Sect. 5.2.

Ontologies are some kind of knowledge artifacts [41] or knowledge organization systems [42]. Efforts have been made to develop metadata vocabularies or application profiles adapted to such systems, for example, the Networked Knowledge Organization Systems (NKOS) working group [43] or the Ontology Metadata Vocabulary working group [44] which results will be further commented later. The Open Ontology Repository Initiative [32] was a collaborative effort to develop a federated infrastructure of ontology repositories and was also interested in the subject. In 2016, a survey was made to the wide ontology developer community with the goal to capture the *Minimum Information for Reporting of an Ontology* and lead to guidelines, recently published [6], on what should be reported about an ontology and its development, in the context of ontology description papers. Although, the intention is slightly different from our work, we believe most information that can be expressed in a scientific article presenting an ontology—including narrative sections such as motivation, knowledge acquisition or change management—can also be captured as appropriate metadata in the ontology itself; we have included in our ontology metadata model some properties to do so. Recently, a new task group (partially lead by the authors) on "ontology-metadata" has been attached to the Research Data Alliance *Vocabulary and Semantic Services Interest Group*.

Finally, the work on ontology metadata is closely related to the one on ontology libraries and repositories. Indeed, with the growing number of ontologies, ontology libraries and repositories have been of interest in the semantic web community. Ding and Fensel [45] presented in 2001 a review of ontology libraries that introduced the notion of "library." Then Hartmann et al. [46] introduced the concept of ontology repository, with advanced features such as search, metadata management, visualization, personalization, and mappings. Most ontology libraries are always capturing some metadata

[5] ISO: http://www.niso.org/schemas/iso25964/ or ISO/IEC: http://metadata-standards.org/11179/#A3 or ISO/IEC 19763-3:2010.

as described Sect. 4.3. D'Aquin and Noy [47] provided the latest review of ontology libraries in 2012. Naskar and Dutta [8] reviews how some ontology libraries use ontology metadata vocabularies.

4 Analysis of Current Ontology Metadata Practices

This analysis was made following three approaches: (1) we have reviewed the most standard and relevant metadata vocabularies available (23 totals) to select properties to describe ontologies; (2) we have reviewed how are these vocabularies used within 805 selected ontologies from known ontology libraries; (3) we have studied some of the most common ontology repositories available in the semantic web community to capture how they are dealing with ontology metadata and to which extent they rely on standard vocabularies.

4.1 Analysis of Existing Metadata Vocabularies to Describe Ontologies or Other General Resources

In the following, we describe the vocabularies that to some extent have been proposed to describe metadata about ontologies. It includes first of all the W3C Recommendations available to describe semantic resources: *Resource Description Framework Schema* (RDFS), *Web Ontology Language* (OWL) and *Simple Knowledge Organization System* (SKOS). Then the *Ontology Metadata Vocabulary (*OMV) [44] produced in the context of several EU projects and published in 2005. OMV (v.2.4.1) consists of 15 classes, 33 object properties, and 29 data properties. Unfortunately, the initiative stopped in 2007. Under the latest OMV version (2.4.1), two physically separated modules are proposed: OMV Core (provide the relevant metadata to support the ontology reuse settings) and OMV Extensions (to allow ontology developers and users to specify task- or application-specific ontology-related information). One limitation of OMV was not to be aligned to (or reuse) standard vocabularies at that time. This limitation has been recently partially addressed by a work published end of 2015: the *Metadata for Ontology Description* (now referred as MOD1.0) [7] which is similar to OMV (without using it). It has been designed as an ontology consisting of 15 classes (mod:Ontology + 10 others + 4 from FOAF), 18 object properties and 33 data properties among 7 of them were not included in OMV. For naming the metadata elements, it has reused existing properties from SKOS, FOAF, DC and DCT. Despite of the seven new properties, MOD1.0 still misses numerous relevant properties as we will see later. In Sect. 7.1, we describe our new join work on MOD1.2 [48] done consequently to the work presented here.

In 2005, the quite simple but relevant *Vocabulary for annotating vocabulary descriptions* (VANN) was made available and quite used since then. In 2009, the *Descriptive Ontology of Ontology Relations* (DOOR) [49] has been published but never really used outside of the NeON project. It was a very formal vocabulary that described precisely and in a logical manner 32 relations between ontologies organized in a formal hierarchy. DOOR did incorporate the ontologies relations offered by OWL. More recently, the *Vocabulary of a Friend* (VOAF) [50] was created to "describe vocabularies (RDFS vocabularies or OWL ontologies) used in the Linked Data Cloud. In particular, it provides properties expressing the different ways such vocabularies can rely on, extend, specify, annotate or otherwise link to each other. It relies itself on DC and VOID." Although VOAF was developed to capture relations between ontologies, it makes no use or reference to OWL or DOOR (with which it captures similar properties). In 2014, the NKOS working group of the Dublin Core proposed the *NKOS Application Profile* (http://nkos.slis.kent.edu/nkos-ap.html) which introduces 6 new properties and reused 22 properties from other vocabularies. [51] published a study made a few years ago to identify the relevant terminology metadata models that could form the foundation for a standard ontology profile for use by the NCI (National Cancer Institute), NCBO (National Center for Biomedical Ontology), and NCRI (National Cancer Research Institute, UK) community. This community effort on identifying the useless or ambiguous element from OMV proposed a few small changes but went no further.[6]

Ontologies share some characteristics with web datasets or data catalogs. Indeed, in the semantic web vision, ontologies are themselves sets of RDF triplets. We thus argue that some properties that have been defined to describe web datasets are relevant to ontologies also. Among the recent work to describe "datasets," there are: the *Vocabulary of Interlinked Datasets* (VOID) [52], a W3C Note proposed in 2011 which can be used "to express general metadata based on DC, access metadata, structural metadata, and links between datasets." VOID allows to describe two main objects void:Dataset and void:Linkset which are sets of links between datasets. The vocabulary also includes URIs for license or serialization formats. *Identifiers.org* (IDOT) [53] is a small vocabulary intended to "referencing of data for the scientific community, with a current focus on the Life Sciences domain." It was developed by the European Bioinformatics Institute to specify, among other things, URI

[6] Some elements are removed (e.g., omv:hasPriorVersion), some element are renamed (e.g., name to fullName, acronym to shortName), some class definitions are modified and two new elements namely, certifiedBy, mandatedBy, are added into the revised set.

patterns. The *Data Catalog Vocabulary* (DCAT), which is the most recent W3C Recommendation for metadata (and uses DCT) and its profile, *Asset Description Metadata Schema* (ADMS), used to describe semantic assets (data models, code lists, taxonomies, dictionaries, vocabularies) created by the EU's Interoperability Solutions for European Public Administrations (ISA). Finally, *Schema.org* has been proposed and adopted in 2011 by Google, Bing and Yahoo! and do include a dataset class.

To describe other kinds of resources, one will find the following vocabularies: *Friend of a Friend Vocabulary* (FOAF) or *Description of a Project* (DOAP) to describe documents and projects. The *Creative Commons Rights Expression Language* (CC) for licensed work. *SPARQL 1.1 Service Description* (SD) for describing SPARQL endpoints. And the *Provenance Ontology* (PROV) and *Provenance, Authoring and Versioning* (PAV) for describing provenance (PAV specializes terms from PROV and DCT). Finally, the *OboInOwl specification* [54] converts OBO ontology header properties to OWL. This is not a standard but some of these properties are handled by the OBO Edit ontology editor and therefore often used.

Other vocabularies recently published or under development, from which we have not selected any properties in our ontology repository metadata model include *Extension to the VOID* [55], which is an extension of VOID mainly for partitions and statistical descriptions. *Citation Typing Ontology* (CiTO) describes citations between entities (one property only is actually relevant for us). The *Protocol for Web Description Resources* (POWDER) provides a mechanism to describe and discover web resources. The *DDI-RDF Discovery Vocabulary* (DISCO) which is a vocabulary to describe studies. The *Information Artifact Ontology* (IAO) [56], which was defined for representation of types of information content entities such as documents, databases, and digital images. The *Semanticscience Integrated Ontology* (SIO) [57] which describes many different types of informational entities and relations between them. [58] have proposed a metadata vocabulary for the Lemon model [59] called *LInguistic Metadata* (LIME) for describing linguistic resources and linguistically enriched datasets. Finally, we must also mention the document ISO/IEC 19763-3 (Metamodel framework for interoperability (MFI)—Part 3: metamodel for ontology registration) which latest version is from 2010 and is not public.

Table 1 summarizes and compares these vocabularies. This review of existing metadata vocabularies (and our work presented in Sect. 5.2) clearly shows no existing vocabularies really cover enough aspects of ontologies to be used solely and despite a few exceptions, metadata vocabularies do not rely on one another. Plus, there is a strong overlap in all the vocabularies studied which redefine things that have already been described several times before (such as dates for which 25 properties are available). When dealing with harmonized metadata in the context of, for instance, an ontology repository, there exists an obvious technical and semantics challenge: being able to process ontologies that could have been described with one or several of those metadata vocabularies. Plus, many of the vocabularies do not support dereferenceability making impossible for the machine to automatically access the semantic description of the properties (e.g., domain, range) defined within the vocabulary. The fact of having multiple vocabularies for describing ontologies (or any other thing) should not be an issue: redundancies on one side enables specificity on the other side. However, in the semantic web vision, we would expect vocabularies to match and rely on one another more. To address our need of properly defining ontologies in an ontology repository, this review gave us a list of candidate metadata properties. In Sect. 5, we will present how we have built a list of properties for AgroPortal's new metadata model based on the studied vocabularies. In Sect. 7.1, we will discuss the need for metadata authoring guidelines and for harmonization of existing metadata vocabularies beyond the AgroPortal project.

4.2 Analysis of Current Use of Ontology Metadata Vocabularies

To get a sense of the quantity and origin of existing metadata vocabularies actually used by ontology developers, we downloaded and semi-automatically analyzed 1107 OWL ontologies taken from different sources: 594 from NCBO BioPortal, 53 from AgroPortal, 260 from MMI Ontology Registry and Repository, 97 from the OBO Foundry, 82 from DERI Vocabularies, and 21 from ProtégéWiki.[7] Once ontology duplicates removed—by matching name or base URIs—we obtained a corpus of 805 ontologies. Because of the sources of the ontologies, this corpus is slightly influenced by certain domains (biomedicine, biology, agronomy, environment); although it might bias the results, we are still confident they are quite representative, especially in these domains. We provide here the result of the analyzed ontologies.

We found 128 ontologies (16%) without any description or annotation. For rest of the 677 ontologies (84%), the number of properties used in describing the ontologies is ranging from 1 to 32. For instance, out of the 53 ontologies retrieved from AgroPortal, there are two ontologies having only one metadata. Overall, there are 354 ontologies (44%) for which ten or more properties (and maximum 32) are observed. For rest of the 323 ontologies (40%), the number of metadata per ontology is below 10.

[7] It is important to understand that we have looked at the metadata in the original ontology file, not the metadata captured by BioPortal or AgroPortal in their internal model.

Table 1 Comparison of reviewed metadata vocabularies

Prefix	Name	Year (version)	Rely on other vocabularies	D	R	Comments
adms	Asset Description Metadata Schema	2013	dc, dcat, foaf, schema + vCard	Y	N	Profile of DCAT. Created by EU's ISA body to help standards publishers
cc	Creative Commons Rights Expression Language	2008		Y		Used to describe copyright licenses in RDF
dc	Dublin Core Elements	2012	–	Y	R	The "original" Dublin Core set of 15 classic metadata terms
dcat	Data Catalog Vocabulary	2014	dc, foaf, vcard	Y	R	W3C Recommendation for data catalog
dct	DCMI Metadata Terms	2012	–	Y	R	An up-to-date specification of all metadata terms maintained by the DCMI
doap	Description of a Project	2012	foaf	Y		Vocabulary to describe software projects
door	Descriptive Ontology of Ontology Relations	2009	–	N		Very formal ontology relation ontology
foaf	Friend of a Friend Vocabulary	2014 (v0.99)	–	Y	N	Linking people and information on the Web. Used as a reference by multiple vocabularies
idot	Identifiers.org	2018	–	Y		Provides stable and perennial identifiers for data records used in the Life Sciences
mod	Metadata for Ontology Description & Publication 1.0	2017 (v1.2)	owl, rdfs, dct, foaf, skos, omv, vann, pav, prov, sd, doap	N		Ontology designed specially to describe ontologies, extension of OMV mainly, but relies on many other metadata vocabularies. Work inspired by our work on AgroPortal
nkos	Networked Knowledge Organization Systems Application Profile	2015 (v0.2)	dc, adms, dcat, prov + frbrer, frsad, wdrs	Y		NKOS is a Dublin Core Application Profile for describing knowledge organization systems
oboInOwl	OboInOwl Mappings	2011 (v1.2)	–	N		A namespace created when transforming OBO ontologies to OWL
omv	Ontology Metadata Vocabulary	2009 (v2.4.1)	–	N		Ontology especially created to describe ontologies. Partially adopted by ontology libraries
owl	OWL 2 Web Ontology Language	2012 (v2)	–	Y	R	W3C Recommendation to create ontologies. Offer a few properties to describe them also
pav	Provenance, Authoring and Versioning	2015 (v 2.3.1)	dc, prov	Y		Lightweight ontology specializing prov to describe provenance
prov	Provenance Ontology	2013	–	Y	R	W3C Recommendation for describing provenance metadata
rdfs	RDF Schema	2014 (v1.1)	–	Y	R	W3C Recommendation for describing any RDF resource
schema	Schema.org	2017 (v3.3)	–	Y		Google, Yahoo!, Bing agreed metadata standard for Web objects
sd	SPARQL 1.1 Service Description	2013	–	Y	R	W3C Recommendation for describing SPARQL services
skos	Simple Knowledge Organization System	2009	–	Y	R	W3C Recommendation for describing thesauri, terminologies, vocabularies
vann	Vocabulary for annotating vocabulary descriptions	2005	–	Y		Lightweight vocabulary for annotating descriptions of vocabularies
voaf	Vocabulary of a Friend	2013 (v2.3)	dc, void	Y		Vocabulary to describe vocabularies and their relations
void	Vocabulary of Interlinked Datasets	2011	dc, foaf	Y	N	Widely adopted vocabulary to describe datasets and their relations

D column states if property URIs are dereferenceable (Y or N); R column states if it is a W3C or Dublin Core Recommendation (R), note (N), or none of the two (blank)

We have also observed in total 30 metadata vocabularies that are being used to describe the ontologies. The 19 most frequently used ones are exemplified in Table 2. Notice that among these, around 1/3 of them are W3C or Dublin Core recommended vocabularies. The rest of vocabularies forms the long tail of the curve of the used metadata vocabularies with a couple of uses or mostly only one. They include recommended standards (e.g., Schema.org), community standards (e.g., CITO, ADMS, DOAP) or very specific vocabularies (e.g., PRISM, EFO, IRON). Some other findings of this study are:

- Most of all these 30 vocabularies are general in purpose. Some metadata vocabularies, which were specially proposed with the purpose of annotating/describing ontologies (e.g., VOID, VOAF, DOOR), are mostly absent or barely used, with the exception of OMV which is not surprisingly among the most used vocabulary.
- However, the presence of OMV—and omvmmi complement to OMV—is mostly explained by the important number of ontologies taken from the MMI Ontology Registry and Repository that has adopted and enforced OMV in the ontologies hosted on their repository. In a previous similar study on 222 ontologies [48], which does not include MMI ontologies but included 61 ontologies randomly selected via Google, OMV was completely absent. This clearly illustrates the impact of harmonized community practices (or repository enforcement) on ontology metadata.
- Two vocabularies among the most used (oboInOwl and protege) are present because they are automatically included in ontologies by ontology development software.[8] Similarly, from Table 2 we can see that rdfs:comment, owl:versionInfo and owl:imports are among the most frequently used metadata elements. We think the reason for their frequent use is because of their ready availability in the ontology editors. For instance, a selected set of metadata elements from rdfs and owl are made readily available in Protégé annotation tab. We may assume most ontology developers find it handy when annotation properties are readily available in the ontology editor's annotation tab, rather than referring a vocabulary available on the Web but not in the editor. The case of owl:imports is slightly different. It is required for functional reasons to import ontologies.
- Multiple properties express the same information. For instance, in providing the name of the ontology, some have used dc:title while some other have used dct:title. Similarly, some people have used dct:license to provide the licensing information, while some others have used cc:license.
- There is a confusion between the use of DC and DCT as the latter includes and refines the 15 primary properties from the former. Some developers prefer to refer DC and some prefer DC Terms for the similar element. The reason could be the unavailability of a precise guideline on how and when to use the DC core and DCT elements. In the context of semantic web applications, although using DC is not incorrect, DCMI recommends using DCT that provides domain and range information for properties.[9]
- Some metadata elements are used in an improper way. For instance, skos:definition shall only be used to supply a complete explanation of the *intended meaning of a* (*SKOS*) *concept* as the other SKOS "documentation properties" and is not supposed to be used to described ontologies (unless an ontology is considered a concept).
- Generic properties such as rdfs:comment or dc:date are used instead of more specific ones such as respectively dc:description or dc:created/modified.
- The study also revealed 12 custom properties used to describe metadata (not reported in Table 2) declared in the main namespace of the ontology, e.g., primary_author_and_curator, wasRevisionOf, contributing_author. This may illustrate a not so good practice which consists in creating a new local property when in need.

We previously conducted a similar smaller study [48] and came to similar outcomes. Another one was conducted by Tejo-Alonso et al. [35]: Their study consisted of total 23 RDFS/OWL metadata vocabularies (the "most popular from prefix.cc"): They were especially interested in how much the metadata vocabularies are themselves described with proper metadata properties. The authors arrived at similar conclusions than us with our larger study: (1) rdfs/owl popularity; (2) dc/dct confusion; (3) frequency of auto-generated properties; (4) generic property over specific ones; (5) different properties for similar information.

Concerning the description of knowledge resources with metadata, we also like to mention an exceptional example found in the context of the AgroPortal project: Agrovoc, which is the reference multilingual thesaurus in agriculture developed by FAO, is explicitly and extensively defined by a so-called "VOID profile"[10] which lives aside from the main

[8] The *oboInOwl* namespace is used by the OBO2OWL converter when converting Open Biomedical Ontology format to OWL. The high frequency of this vocabulary is explained because half of our ontologies were selected from the NCBO BioPortal that contains many ontologies originally developed in OBO (often with the OBOEdit software). The *protege* name space was used in previous (~v3) of Protégé mostly to customize the user interface when displaying the ontology. It was not to describe the ontology.

[9] http://wiki.dublincore.org/index.php/FAQ/DC_and_DCTERMS_Namespaces.

[10] http://aims.fao.org/aos/agrovoc/void.ttl.

Table 2 Most frequent used vocabularies over a corpus of 805 ontologies

Prefix	Number	Properties used (number of times)
omv	2169	acronym (251), creationDate (251), description (251), hasCreator (251), name (251), uri (251), usedontologyengineeringtool (157), version (148), keywords (126), hasContributor (109), documentation (74), reference (49)
omvmmi	1697	creditRequired (251), origMaintainerCode (251), hasContentCreator (193), hasResourceType (186), shortNameuri (151), temporarymmirole (108), origvocManager (107), contactRole (106), contact (99), origvocuri (60), origvocDocumentationuri (40), creditCitation (38), origvocDescriptiveName (36), origvocSyntaxFormat (30), origvocKeywords (23), origvocVersionid (16), origvocLastModified (1)
dc	1599	creator (456), description (309), date (307), contributor (183), source (77), title (102), subject (47), format (31), license (28), publisher (21), rights (17), language (8), identifier (6), modified (3), coverage (2), issued (1), type (1)
dct	652	modified (86), title (85), created (84), partOf (81), status (81), type (81), description (62), publisher (60), creator (9), license (6), issued (3), subject (2), contributor (3), isreferencedby (3), identifier (1), isrequiredby (1), language (1), date (1), source (1), format (1)
owl	498	versionInfo (183), imports (210), versionIRI (74), priorVersion (22), ontology (4), incompatibleWith (3), backwardCompatibleWith (1), deprecated (1)
oboInOwl	283	default-namespace (54), hasOboFormatVersion (53), savedBy (49), date (47), auto-generated-by (40), namespaceIdRule (7), treat-xrefs-as-equivalent (5), hassubset (4), remark (4), treat-xrefs-as-is_a (4), treat-xrefs-as-genus-differentia (3), format-version (2), pairwise-disjoint (2), treat-xrefs-as-has-subclass (2), treat-xrefs-as-reverse-genus-differentia (2), comment (1), data-version (1), default-relationship-id-prefix (1), next-id (1), property-value (1)
rdfs	265	comment (174), label (68), seeAlso (16), isDefinedBy (7)
vann	166	preferredNamespacePrefix (83), preferredNamespaceUri (83)
foaf	102	homepage (91), mbox (6), page (4), isPrimaryTopicOf (2)
obo	33	iao_0000116 (10), idspace (4), date (3), default-relationship-id-prefix (3), format-version (2), remark (2), comment (1), iao_0000117 (1), iao_0000412 (1), definition (1), editorialNote (1), historyNote (1), imports (1), is_metadata_tag (1), license (1)
skos	19	altLabel (6), prefLabel (6), definition (5), changeNote (1)
protégé	19	defaultLanguage (19)
nemo_annot	10	created_date (2), curator (2), modified_date (2), pref_label (2), synonym (2)
vaem	9	dateCreated (1), hasAspectsCope (1), hasCatalogEntry (1), hasDisciplineScope (1), hasDomainScope (1), hasRole (1), lastUpdated (1), revisionNumber (1), usesNonImportedResource (1)
cc	4	license (4)
dcat	3	landingPage (2), downloadURL (1)
asthma	2	creator (1), defaultLanguage (1)
pav	2	version (2)
void	2	dataBrowse (1), dataDump (1)

For namespaces either see Table 1 or in some cases on https://prefix.cc

thesaurus file and uses 7 metadata vocabularies to describe Agrovoc with RDF statements.

This review helped us to decide which vocabulary and/or property shall be "prioritized" when selecting properties for our unique model in an ontology repository. The final step was then to look at how other ontology libraries were dealing with metadata.

4.3 Analysis of Metadata Representation Within Ontology Libraries

We have studied some of the most common ontology libraries and repositories available in the semantic web community, and especially the NCBO BioPortal, to analyze: (1) how they are dealing with ontology metadata; (2) to which extent they rely on previously analyzed metadata vocabularies. We have only been interested in the metadata that are "nonspecific" to the repository, i.e., specific fields required for implementation purposes were ignored.

We consider under the term libraries any kind of web tool (repository, registry or portal) that somehow focus on ontologies and/or vocabularies [45]. In particular, we have explicitly reviewed:

1. Repository or portals including the NCBO BioPortal [12], Ontobee [60], EBI Ontology Lookup Service [10], MMI Ontology Registry and Repository [11], the ESIP portal (based on NCBO technology), and AberOWL [61];
2. Registries or catalogs including the OKFN Linked Open Vocabularies [9], OBO Foundry [16], WebProtégé (http://webprotege.stanford.edu), Agrisemantics Map of Data Standards (http://vest.agrisemantics.org) [62], FAIR-Sharing (https://fairsharing.org) [40];
3. Web indexes such as Watson [63], Swoogle [1] (or Sindice.com, not reviewed because not accessible anymore).

We have reviewed the metadata properties used by all these libraries and considered them for our listing to be implemented in our portal. As later explained, we have used BioPortal as baseline. Each of the reviewed libraries uses, to some extent, some metadata fields but do not always use standard metadata vocabularies:

- NCBO BioPortal repository [12] uses 66 metadata properties that serves as the basis for our listing.[11] These properties are defined in an in-house vocabulary (here called BioPortal Metadata and identified with the namespace bpm) that is not formally described outside of BioPortal but because of the portal adoption of JSON-LD, can be formally used.[12] For 10 properties, BioPortal reuses OMV names but redefines them in its own namespace (e.g., bpm:omvacronym). Other than the 10 OMV property names, BioPortal does not use any other metadata vocabulary. Over the 66 properties used by BioPortal, we have classified 46 (36 locally defined +10 from OMV) as nonspecific to the portal. BioPortal user interface (and web services) allows to edit most of the properties and some of them are automatically generated (e.g., metrics). Because they originally use the same source code, the situation is the same for ESIP portal and AgroPortal before our work.
- MMI Open Ontology Repository, which was originally also based on BioPortal code, did later embrace OMV more and added a few other metadata properties (omvmmi extension). The repository administrators do edit the ontology metadata of the files hosted on the portal to harmonize them.
- Linked Open Vocabulary registry [9] explicitly uses VOID and VOAF; the latter was actually created for this purpose. The LOV is a very good example of good use of harmonized metadata that has inspired us a lot. More than 600 vocabularies (as of May 2017) are described with common metadata fields facilitating manual and automatic search. In addition, LOV is not limited to VOAF and recommends the use of other standard vocabularies.[13] It is important to note that the metadata is either entered by the developer submitting the vocabulary then curated by the registry administrators. Some are also automatically generated and, in both cases, LOV always relies on standard vocabularies to store the information.
- OBO Foundry [16] refers metadata from around 20 vocabularies including DC, FOAF, IDOT, VOID, DOAP, DISCO, etc.[14] The OBO Foundry community effort is important, and they encourage the ontology developers to edit the metadata, aside from the main ontology file, in a specific document (in MD or YAML format) hosted on GitHub aside of the ontology files and parsed by the OBO Foundry application to display ontology descriptions.[15] OBO Foundry administrators manually curate/edit ontology metadata in complement of ontology developers.
- Ontobee [60] offers a few (6–7) common metadata (e.g., IRI, home, contact) and then display any other metadata

[11] http://data.bioontology.org/documentation#OntologySubmission and #Ontology.

[12] Originally, the NCBO developed the BioPortal Metadata Ontology (http://purl.bioontology.org/ontology/BP-METADATA) which imports OMV. But the current implementation is not completely in sync with this vocabulary anymore.

[13] http://lov.okfn.org/Recommendations_Vocabulary_Design.pdf.

[14] http://obofoundry.github.io/registry/context.jsonld.

[15] For instance: https://github.com/OBOFoundry/OBOFoundry.github.io/blob/master/ontology/envo.md.

properties originally included in the ontology as "annotation properties." The portal also counts a few metrics.

- Similarly, AberOWL [61] and OLS [10], have a few common properties and then display the rest (included in the ontology file) as annotation properties. By comparison to OBO Foundry, the common properties are not described with standard vocabularies.

For a recent review of ontology libraries and their metadata, the reader might refer to [8], briefly summarized in [7]. In these papers, authors showed that ontology metadata vocabularies are rarely used by ontology libraries: 4[16] ontology libraries over the 13 studied have partially used the OMV.

5 Building a List of Properties to Describe Ontologies

5.1 Method to Select Properties from Existing Vocabularies

Enlightened by the analysis presented in the previous section, we have accomplished a systematic review (as methodologically described by [64]) of the vocabularies previously identified with the following research question in mind: *Which existing properties could be used to describe ontologies*? The previously listed vocabularies have been identified from: (1) the semantic web literature; (2) investigating ontology libraries; (3) related similar studies such as the one for dataset by the HCLS working group. Vocabularies were selected based on their degree of standardization, relevance for ontologies and current usage by ontology developers. The final list of the 23 reviewed vocabularies and the numbers of property reused are available in Table 3, plus the NCBO BioPortal metadata model that we used as baseline and listed as a vocabulary with the prefix "bpm."

We now describe selection criteria for properties to be used by our ontology portal. The goal of this list was to delimit the set of properties that our ontology repository will "parse," i.e., the ones that will be automatically recognized and used to populate the unified ontology metadata model. Indeed, our motivation was to improve metadata management within AgroPortal, a portal based on the NCBO technology. For other important reasons in the AgroPortal project (maintenance, collaboration, support, interoperability), keeping our ontology repository backward compatible with NCBO was mandatory. Therefore, each time a property was already captured by the BioPortal model, we would add it to the list and not change it to another property that the analysis Sect. 4 would have shown more relevant. The criteria for inclusion were the following, considered by order of importance:

1. Relevance for describing an ontology—the property may have a sense if used to describe an ontology.
2. Being not "specific" to a library—even if the ontology library helps to populate or predict the property, the property would capture an information that belongs to the ontology. For instance, properties such as credentials on the portal or maintenance information, or local parsing status are considered "specific."
3. Semantic consistency—there must not be any conflict (e.g., disjoint classes) if someone would describe an ontology with all the listed properties. For instance, an ontology may be an instance of omv:Ontology, void:Dataset and cc:Work at the same time.
4. Being a W3C or Dublin Core Recommendations.
5. The frequency of use in the study presented in Sect. 4.2.
6. Priority to vocabularies specific for ontologies rather than to the ones specialized for more general object (cc:Work, dcat:DataSet, sd:Service, etc.).

Although we agree dereferenceability is an important criterion for a vocabulary, we have not excluded properties that are not dereferenceable, even it means a machine would hardly understand the semantics of the property. We will mention this as a requirement for a future ontology metadata vocabulary in Sect. 7.1.

5.2 Properties Selected from Existing Metadata Vocabularies

For each of these vocabularies, we have selected the significant properties to describe objects that an ontology could be considered a certain type of, e.g., a dataset, an asset, a project or a document. For instance, an ontology may be seen as a prov:Entity object and then the property prov:wasGeneratedBy may then be used to describe its provenance. We illustrate with examples as often as possible.

The first things to look at are the properties available in the W3C standard vocabularies, such as RDFS, OWL, and SKOS. Indeed, they include some annotation properties that we can use to describe ontologies if we consider them instances of rdfs:Resource, owl:Ontology or skos:conceptScheme.

rdfs:label, rdfs:seeAlso, rdfs:comment, owl:versionInfo, owl:versionIRI, owl:imports, owl:priorVersion, owl:backwardCompatibleWith, owl:incompatibleWith, owl:deprecated, skos:prefLabel, skos:altLabel, skos:hiddenLabel, skos:hasTopConcept, skos:notation

SKOS label properties can be used to denote the alternative or non-conventional names of an ontology. For instance, the Phenotype And Trait Ontology is also known as "PATO,"

[16] The study reported 3 only, but the case of MMI was a mistake.

Table 3 Vocabularies studied in this review + BioPortal

Prefix	Namespace	Resource	#T	#S	#U
adms	http://www.w3.org/ns/adms#	adms:Asset	13	11	0
cc	http://creativecommons.org/ns#	cc:Work	5	5	2
dc	http://purl.org/dc/elements/1.1/	NA	15	15	0
dcat	http://www.w3.org/ns/dcat#	dcat:Dataset	5	4	0
dct	http://purl.org/dc/terms/	dcmi:Dataset, dcmi:Collection	55	38	13
doap	http://usefulinc.com/ns/doap#	doap:Project	25	18	3
door	http://kannel.open.ac.uk/ontology#	owl:Ontology	32	11	6
foaf	http://xmlns.com/foaf/0.1/	foaf:Document	11	10	4
idot	http://identifiers.org/idot/	dct:Dataset	9	6	1
mod	http://www.isibang.ac.in/ns/mod#	mod:Ontology	27	26	1
nkos	http://w3id.org/nkos#	rdfs:Resource	6	4	0
oboInOwl	http://www.geneontology.org/formats/oboInOwl#	owl:Ontology	13	9	0
omv	http://omv.ontoware.org/2005/05/ontology#	omv:Ontology	37	37	35
owl	http://www.w3.org/2002/07/owl#	owl:Ontology	11	7	2
pav	http://purl.org/pav/	prov:Entity	30	16	2
prov	http://www.w3.org/ns/prov#	prov:Entity	22	10	2
rdfs	http://www.w3.org/2000/01/rdf-schema#	rdfs:Resource	7	3	0
schema	http://schema.org/	schema:Dataset	90	41	7
sd	http://www.w3.org/ns/sparql-service-description#	sd:Service	13	1	1
skos	http://www.w3.org/2004/02/skos/core#	skos:conceptScheme	14	5	1
vann	http://purl.org/vocab/vann/	rdfs:Resource	6	5	3
voaf	http://purl.org/vocommons/voaf#	voaf:Vocabulary	16	12	5
void	http://rdfs.org/ns/void#	void:Dataset	24	16	5
bpm	http://data.bioontology.org/metadata	bpm:Ontology bpm:OntologySubmission	36	36	34
Total			522	346	127

Column #T is the total number of properties provided by the vocabulary for column Resource type (or rdfs:Resource). Column #S is the number of properties *selected* in the list from this vocabulary (only vocabularies within the same namespace). Column #U is the number of properties *used* as default property in the implementation of the new ontology repository model. For instance, for foaf:Document, we have reviewed a total of 11 properties and considered 10 of them were relevant to describe ontologies and are now parsed by AgroPortal, but only 4 have been explicitly used as "default" property in the new model

"Phenotypic Quality Ontology," or "Ontology of phenotypic qualities."

Then the Dublin Core Metadata Initiative standards are available. Dublin Core does not always specify the domain of its properties. We have assumed that all of them accept rdfs:Resource as domain. We have included the 15 DC properties and 38 DCT properties that are relevant for describing ontologies (only DCT is listed hereafter):

dct:title, dct:accessRights, dct:isPartOf, dct:hasVersion, dct:bibliographicCitation, dct:language, dct:dateSubmitted, dct:description, dct:created, dct:date, dct:issued, dct:rightsHolder, dct:modified, dct:conformsTo, dct:contributor, dct:creator, dct:subject, dct:rights, dct:license, dct:format, dct:type, dct:requires, dct:isVersionOf, dct:relation, dct:coverage, dct:publisher, dct:identifier, dct:source, dct:abstract, dct:alternative, dct:hasPart, dct:isFormatOf, dct:hasFormat, dct:audience, dct:valid, dct:accrualMethod, dct:accrualPeriodicity, dct:accrualPolicy

DCT's accrual properties can be used for instance to describe the process by which an ontology is updated and

new concepts are added or removed. This has been established as an important aspect by the *Minimum Information for Reporting of an Ontology* guidelines.

Among the vocabularies available for ontologies we have taken all the properties from OMV and MOD[17] considering an ontology an instance of omv:Ontology and mod:Ontology. We only list the ones in OMV namespace (when they are named the same in MOD):

omv:acronym, omv:name, omv:hasOntologyLanguage, omv:reference, omv:URI, omv:naturalLanguage, omv:documentation, omv:version, omv:creationDate, omv:description, omv:status, omv:resourceLocator, omv:numberOfClasses, omv:numberOfIndividuals, omv:numberOfProperties, omv:modificationDate, omv:numberOfAxioms, omv:keyClasses, omv:keywords, omv:knownUsage, omv:conformsToKnowledgeRepresentationParadigm, omv:hasContributor, omv:hasCreator, omv:designedForOntologyTask, omv:endorsedBy, omv:hasDomain, omv:hasFormalityLevel, omv:hasLicense, omv:hasOntologySyntax, omv:isOfType, omv:usedOntologyEngineeringMethodology, omv:notes, omv:usedOntologyEngineeringTool, omv:useImports, omv:hasPriorVersion, omv:isBackwardCompatibleWith, omv:isIncompatibleWith, mod:accessibility, mod:module, mod:ontologyInUse, mod:sponsoredBy, mod:competencyQuestion, mod:vocabularyUsed, mod:homepage

OMV properties (and individuals) are particularly relevant as they have been explicitly created to describe ontologies. They are the only ones in our study enabling to capture information such as the methodology applied to create the ontology or the task/role for which an ontology has been designed. For instance, the Medical Subject Headings (MeSH) terminology has been designed for indexing scientific medical publications (omv:IndexingTask), which is different from the Gene Ontology that has been developed to annotate gene products (omv:AnnotationTask). Among the new properties from MOD, mod:competencyQuestion corresponds to properties suggested for instance by [65]

There exist two specific vocabularies for representing relations. From DOOR, that is very detailed and formal, we have selected 11 of the most significant, in addition to the 4 from OWL. We had to draw the line, and we considered 15 formal relations from these two vocabularies were enough in most cases to describe ontology relations. VOAF properties (applied to a voaf:Vocabulary) were almost completely included, except 4 statistical properties (that are relevant only for a specific repository):

door:semanticallyIncludedIn, door:imports, door:priorVersion, door:backwardCompatibleWith, door:owlIncompatibleWith, door:ontologyRelatedTo, door:similarTo, door:comesFromTheSameDomain, door:isAlignedTo, door:explanationEvolution, door:hasDisparateModelling, voaf:classNumber, voaf:propertyNumber, voaf:extends, voaf:reliesOn, voaf:similar, voaf:hasEquivalencesWith, voaf:specializes, voaf:usedBy, voaf:metadataVoc, voaf:generalizes, voaf:hasDisjunctionsWith, voaf:toDoList

The property door:explanationEvolution or voaf:specializes can be used to say that an ontology is a latter version that is semantically equivalent to another ontology and specializes it. For instance, International Classification of Diseases, 10th revision (ICD-10) has for prior version ICD-9 and for specialization ICD-10-CM (Clinical Modification made by US National Center for Health Statistics).

From NKOS Application Profile, we have selected 4 properties among the 6 new ones defined in the namespace and have in that case considered the properties would be applied to rdfs:Resource. Two have been excluded because we already have more precise properties in other vocabularies (nkos:serviceOffered and nkos:sizeNote).

nkos:alignedWith, nkos:basedOn, nkos:updateFrequency, nkos:usedBy

Among the metadata vocabularies to describe datasets, we have reviewed VOID, a W3C Note proposed in 2011 to describe RDF datasets. It allows describing two main objects void:Dataset and void:Linkset which are set of links between datasets. The vocabulary also includes URIs for license or serialization formats. void:Dataset can be described with 24 properties including a few metrics plus some from DCT. From VOID, we picked-up 16 relevant properties.

void:subset, void:classPartition, void:propertyPartition, void:rootResource, void:classes, void:properties, void:triples, void:entities, void:exampleResource, void:vocabulary, void:sparqlEndpoint, void:dataDump, void:openSearchDescription, void:uriLookupEndpoint, void:uriRegexPattern, void:uriSpace

For instance, void:uriRegexPattern may be used to explain the pattern that some ontologies use when building their URIs and concept identifiers, e.g., (ICD-10)'s codes respect a structure that keeps track of the chapter, and hierarchy (K70.3 code for "Alcoholic cirrhosis of liver" is the 3rd of "Alcoholic liver disease" (K70) which are all in the "Diseases of the digestive system" Chapter (K)).

[17] Except mod:size that was new in MOD and ambiguous ("the size of an ontology").

A few of the properties from Indentifiers.org (IDOT) (6) shall be relevant to describe ontologies also:

idot:state, idot:obsolete, idot:alternatePrefix, idot:identifierPattern, idot:preferredPrefix, idot:exampleIdentifier

VANN is a small vocabulary created to describe vocabularies, which includes:

vann:preferredNamespacePrefix, vann:preferredNamespaceUri, vann:usageNote, vann:example, vann:changes

The property idot:preferredPrefix or vann:preferredNamespacePrefix can be used to store the preferred prefix when using the ontologies. See for example, http://prefix.cc for all possible prefix values.

DCAT is the W3C Recommendation since January 2014 to describe data catalogs; it offers a dcat:Dataset class relevant for ontologies. DCAT uses DCT and also offers properties with domain dcat:Distribution, but we have not taken those ones to restrict our selection to the dcat:Dataset class (among the 4 missed properties, 3 finds equivalent in other vocabularies). Then from ADMS, which is a profile of DCAT used to describe semantic assets (data models, code lists, taxonomies, dictionaries, vocabularies), we took 19 properties for class adms:Asset (or no domain) but only 11 specifically defined in the adms namespace, because ADMS used several other vocabularies treated in this study:

dcat:landingPage, dcat:contactPoint, dcat:keyword, dcat:theme, adms:sample, adms:status, adms:versionNotes, adms:representationTechnique, adms:prev, adms:last, adms:next, adms:includedAsset, adms:identifier, adms:supportedSchema, adms:translation

In the SIFR BioPortal project [20], we are interested to formally represent that some ontologies are the translated version of other ones (usually stored in the NCBO BioPortal). For instance, the French Medical Dictionary for Regulatory Activities Terminology is translated from the English version. The adms:translation can be used for this.

Schema.org (SCHEMA) can describe multiple types of resources. We have identified the schema:Dataset type as the closest one to describe ontologies. Schema.org is very rich to describe schema:Dataset (including properties inherited of schema:CreativeWork and schema:Thing), we have identified 41 relevant properties:

schema:distribution, schema:includedInDataCatalog, schema:spatial, schema:about, schema:alternativeHeadline, schema:associatedMedia, schema:audience, schema:author, schema:award, schema:comments, schema:contributor, schema:copyrightHolder, schema:creator, schema:dateCreated, schema:dateModified, schema:datePublished, schema:workExample, schema:fileFormat, schema:hasPart, schema:isPartOf, schema:inLanguage, schema:isBasedOn, schema:keywords, schema:license, schema:mainEntity, schema:publisher, schema:publishingPrinciples, schema:review, schema:schemaVersion, schema:sourceOrganization, schema:translator, schema:version, schema:alternateName, schema:description, schema:image, schema:mainEntityOfPage, schema:citation, schema:name, schema:url, schema:translationOfWork, schema:translation

For instance, the property schema:includedInDataCatalog may be used to store the fact that an ontology is hosted in different ontology libraries. This is, for instance, the cases for the OBO Foundry ontologies that are, in addition of the foundry being uploaded in NCBO BioPortal, Ontobee, OLS and AberOWL. With such a property properly populated, everyone will always know in which library to find an ontology.

If we consider an ontology as different kinds of objects, additional relevant vocabularies may be used. Thus, FOAF can be used to describe an ontology as an instance of foaf:Document, DOAP if an ontology is viewed as development project (doap:Project) and CC to see it as a cc:Work[18]:

foaf:name, foaf:homepage, foaf:isPrimaryTopicOf, foaf:page, foaf:primaryTopic, foaf:maker, foaf:topic, foaf:depiction foaf:logo, foaf:fundedBy, doap:name, doap:blog, doap:language, doap:wiki, doap:release, doap:description, doap:created, doap:download-page, doap:helper, doap:maintainer, doap:translator, doap:audience, doap:download-mirror, doap:service-endpoint, doap:screenshots, doap:repository, doap:bug-database, doap:mailing-list, cc:attributionName, cc:attributionURL, cc:license, cc:morePermissions, cc:useGuidelines

More and more ontology developers have turned to GitHub to store and release their ontologies, for example, the Environment Ontology (https://github.com/EnvironmentOntology). The DOAP properties are thus very relevant to capture the metadata about the ontology development project.

Two vocabularies for representing provenance information are included: PROV and PAV. PAV specializes terms from PROV and DCT. It contains 40 properties (including 30 specific ones) with no constraint on range or domain. When incorporating PROV and PAV, we had to focus on the main properties offered to describe prov:Entity (but potentially more maybe used):

[18] We have here an inconsistency as doap:Project are themselves subclasses of foaf:Project and because foaf:Project and foaf:Document are disjoints. We let to ontology developers the choice.

prov:generalizationOf, prov:generatedAtTime, prov:wasAttributedTo, prov:wasInfluencedBy, prov:wasDerivedFrom, prov:wasRevisionOf, prov:specializationOf, prov:invaliatedAtTime, prov:wasGeneratedBy, prov:wasInvalidatedBy, pav:hasCurrentVersion, pav:hasVersion, pav:version, pav:createdOn, pav:authoredOn, pav:contributedOn, pav:lastUpdateOn, pav:contributedBy, pav:authoredBy, pav:createdBy, pav:createdWith, pav:previousVersion, pav:hasEarlierVersion, pav:derivedFrom, pav:curatedBy, pav:curatedOn

From the OboInOwl specification, we took 9 of the 13 properties (and the alternative names, not listed, e.g., savedBy):

oboInOwl:format-version, oboInOwl:data-version, oboInOwl:date, oboInOwl:saved-by, oboInOwl:auto-generated-by, oboInOwl:import, oboInOwl:synonymtypedef, oboInOwl:default-namespace, oboInOwl:remark

Finally, we have selected sd:endpoint from SPARQL 1.1 Service Description.

5.3 Existing Properties in Ontology Repositories

In order to manage versioning, access rights and metadata, BioPortal model stores ontologies with two objects: one Ontology which is actually the shell for multiple Submissions that contains the real content of an ontology. The Ontology object contains the most usual metadata (name, acronym, administrators, viewing restriction, group and categories) that will remain over versions, whereas the Submission objects contain the detailed metadata (description, metrics, contact, etc.) and links to the actual content of that specific version. For example, the following REST service calls will return, respectively, the Ontology object and the latest Submission for the NCI Thesaurus:

http://data.bioontology.org/ontologies/NCIT?display=all
http://data.bioontology.org/ontologies/NCIT/latest_submission?display=all

We have reviewed the complete list of properties offered by those two objects (including direct properties and links returned by the API): 25 for Ontology and 41 for Submission. From them, we picked-up the ones (46) that are not specific to BioPortal. For instance, the administrator (different from contact) of an ontology in BioPortal is an information that has sense only within BioPortal and therefore does not belong to the original ontology.

For homogeneity, we use the namespace bpm in the following list, even if those properties do not actually belong to a formal vocabulary (we do not include hereafter the 10 OMV properties originally used by BioPortal):

bpm:group, bpm:viewOf, bpm:submissions, bpm:reviews, bpm:notes, bpm:projects, bpm:views, bpm:analytics, bpm:ui, bpm:properties, bpm:classes, bpm:roots, bpm:prefLabelProperty, bpm:definitionProperty, bpm:synonymProperty, bpm:authorProperty, bpm:hierarchyProperty, bpm:obsoleteProperty, bpm:obsoleteParent, bpm:homepage, bpm:publication, bpm:released, bpm:diffFilePath, bpm:pullLocation, bpm:contact, bpm:metrics.classes, bpm:metrics.individuals, bpm:metrics.properties, bpm:metrics.maxDepth, bpm:metrics.maxChildCount, bpm:metrics.averageChildCount, bpm:metrics.classesWithOneChild, bpm:metrics.classesWithMoreThan25Children, bpm:metrics.classesWithNoDefinition, bpm:downloadRdf, bpm:downloadCsv

Once a primary version of the list was created from BioPortal plus the standard metadata vocabularies, we also analyzed the other ontology repositories. We did not find other properties that were not already covered by our review so far. From the OBO Foundry, the only exceptions were the properties inside the obofmd namespace (non-dereferenceable), that seems to be the ones the OBO Foundry developers did not find in any vocabulary. Although we have matches for 4 over 5 of these properties, we did not integrate those by the lack of information about them (plus this namespace was not identified in Sect. 4.2). AberOWL contains also a property species that we did not pick up as this is specific to the biomedical domain and unsatisfiable classes which are an interesting information for the ontology evaluation, but not for ontology description. OLS contains also two properties that we do not already had (reasonerType and oboSlims) but were not included by the lack of information. Even if we have an interest in biological and agronomical ontologies, we did not include in this list, properties that are domain specific. All the properties can be used to describe ontologies from any domain.

5.4 Results: A Complete List of Properties to Describe Ontologies and a Unified Model for AgroPortal

After the two steps described in the previous section, we end up with a complete list of 346 properties that could be used to describe ontologies. These properties will, therefore, be parsed by AgroPortal when an ontology is uploaded in order to populate the values of unified model implemented for all the ontologies on the portal. With the 346 properties of this list, we cover most of the properties identified in Table 2 except the ones in namespaces that are not relevant for ontologies (e.g., nemo_annot, vaem and asthma), portal specific (e.g., omvmmi), format specific or not defined as a vocabulary (e.g., obo), or software specific (e.g., protege) or within the oboInOwl namespace but not in the OBO in OWL specification [54]. Among the 31 properties from Tejo-Alonso et al.'s study [35], we cover 25 properties. The six proper-

ties not included are 4 SKOS "documentation properties" (e.g., skos:changeNote, skos:definition), that according to the SKOS specification are intended to provide information relating to concepts although there is no domain restriction for these properties. The two others are rdfs:isDefinedBy[19] and vs:terms_status excluded for an equivalent reason. We, therefore, believe our complete list of properties that will be parsed by our ontology repository include most of the properties actually used by ontology developers.

Among those properties of the complete list, there was obvious overlap. Indeed, some properties define exactly the same thing, e.g., the version information of an ontology can be described by omv:version, owl:versionInfo, mod:version, doap:release, pav:version and schema:version. And some properties define very similar things such as for instance the homepage of an ontology project: bpm:homepage, foaf:homepage, cc:attributionURL, mod:homepage, doap:blog, and schema:mainEntityOfPage. With the purpose of simplifying our list, and implement a restricted unified model within our ontology repository, we have grouped properties of exact or similar meaning by selecting a "default" property that we would use in our ontology metadata model. The role of these equivalences (we voluntary do not use the word mapping or alignment) is not to build a unique vocabulary for describing ontologies (although this question will be discussed in Sect. 7), but to implement an unified model for describing ontologies in an ontology repository that would help us address the challenges explained in Sects. 1 and 2. When selecting the "default" property, we applied the following rules that are specific to our context:

1. Do not change the properties that were already in BioPortal. As previously explained, we had to keep AgroPortal backward compatible with BioPortal (we will further discuss this in Sect. 7). Except for 3 metric properties that we have duplicated to enable users to reset themselves the number of classes, individuals and properties, we have reused all the 34 other properties already implemented in BioPortal;
2. Pick up the OMV property if existing (to stay consistent with BioPortal's historical choice of using OMV);
3. If not available within OMV, choose property from any other vocabulary offering the best correspondence by giving preference when possible to W3C Recommendations or Notes. With this in mind, we prefer dct:publisher to schema:publisher and adms:schemaAgency. Or, foaf:fundedBy rater than mod:sponsoredBy and schema:sourceOrganization.

We came up with a list of 127 properties in the restricted unified model including the 46 original ones from BioPortal (nonspecific) and 82 new ones from metadata vocabularies. For a better comprehension, we categorized the properties as illustrated in Table 4. Among them, 17 properties from BioPortal cannot be mapped to any of the studied vocabularies, which means that they are candidates for extending one of the studied vocabularies or creating a new one (cf. Sect. 7.1). For example: bpm:group, bpm:downloadCsv, or a few metrics, and properties describing the classes.

When selecting a default property for the unified model and grouping properties by equivalences, we had to make choices (that we have tried less arbitrary possible). These were guided by our context and motivation (i.e., implementing this model in AgroPortal) and shall differ from projects with other motivations. Here are a few examples of these choices:

- We kept omv:notes over rdfs:comment, or adms:versionNotes in order to stay consistent with BioPortal's choice of partially adopting OMV. This choice was made in 2009 right after the OMV vocabulary was proposed and according to us, this was a good choice at that time. We would not necessarily encourage the use of omv:notes (or any OMV property for which a more standard vocabulary already provides something) over rdfs:comment anymore now. Indeed, this is a limitation of OMV that we have pointed out. Finally, our model includes 35 of the 37 relations of OMV. The two missing are omv:reference and omv:resourceLocator that we have not included because BioPoral already offered a property for them (but not the OMV one!) respectively bpm:publication and bpm:pullLocation.
- For a property that was not already captured by BioPortal or OMV, such as the fact that an ontology is deprecated, we give priority to established standards, e.g., owl:deprecated over idot:obsolete as the OWL property (which applies to any IRI) comes from a W3C Recommendation.
- We selected dct:publisher over schema:publisher as our analysis has shown that Dublin Core (and Elements) properties are widely used among ontology developers. This might of course change in the future considering the pace of adoption of Schema.org.[20]
- For the relation between an ontology and a view of this ontology, BioPortal defines bpm:viewOf and bpm:views that we have kept, respectively, over dct:isPartOf (or schema:isPartOf or void:subset or door:sematicallyIncluedIn) and dct:hasPart (or

[19] Although the domain of rdfs:isDefinedBy is rdfs:Resource, it is defined by the RDF specification as: "may be used to indicate an RDF vocabulary in which a resource is described.".

[20] On that example, one can regret the fact that Shema.org has not itself adopted Dublin Core or that the two organizations do not work together. Similarly, Schema.org and DCAT are particularly rich and we shall follow closely the effort of harmonizing them in the future.

Table 4 Restricted list of 127 properties ("default") implemented in AgroPortal's unified metadata model

Category	List of properties in this category
Intrinsic properties	omv:acronym, omv:name, dct:alternative, skos:hiddenLabel, omv:URI, owl:versionIRI, dct:identifier, omv:version, omv:status, owl:deprecated, omv:hasLicense, omv:hasOntologyLanguage, omv:hasFormalityLevel, omv:hasOntologySyntax, omv:naturalLanguage
Description	omv:description, bpm:publication, omv:documentation, dct:abstract, cc:morePermissions, cc:useGuidelines, schema:copyrightHolder, bpm:pullLocation, omv:notes, omv:keywords, omv:isOfType, omv:designedForOntologyTask, omv:usedOntologyEngineeringTool, omv:usedOntologyEngineeringMethodology, omv:conformsToKnowledgeRepresentationParadigm, dct:coverage, mod:competencyQuestion, foaf:depiction, foaf:logo, foaf:homepage, schema:associatedMedia, bpm:diffFilePath, vann:example, idot:exampleIdentifier, vann:preferredNamespaceUri, vann:preferredNamespacePrefix, void:uriRegexPattern, bpm:prefLabelProperty, bpm:definitionProperty, bpm:synonymProperty, bpm:authorProperty, bpm:hierarchyProperty, bpm:obsoleteProperty, bpm:obsoleteParent, schema:includedInDataCatalog
People	omv:hasCreator, omv:hasContributor, dct:publisher, pav:curatedBy, bpm:contact, schema:translator
Grouping	omv:hasDomain, bpm:group
Relation	omv:useImports, omv:hasPriorVersion, omv:isBackwardCompatibleWith, omv:isIncompatibleWith, bpm:viewOf, bpm:views, bpm:submissions, bpm:hasPart, dct:isFormatOf, dct:hasFormat, door:ontologyRelatedTo, door:similarTo, door:comesFromTheSameDomain, door:explanationEvolution, door:hasDisparateModelling, door:isAlignedTo, schema:translationOfWork, schema:workTranslation, voaf:usedBy, voaf:generalizes, voaf:hasDisjunctionsWith
Content	omv:keyClasses, bpm:ui, sd:endpoint, voaf:metadataVoc, bpm:csvDump, bpm:properties, bpm:classes, bpm:roots, void:dataDump, void:uriLookupEndpoint, void:openSearchDescription, bpm:downloadRdf, bpm:downloadCsv
Community	omv:knownUsage, omv:endorsedBy, bpm:projects, dct:audience, bpm:analytics, foaf:fundedBy, bpm:reviews, bpm:notes, voaf:toDoList, doap:repository, doap:bug-database, doap:mailing-list, schema:award
Date	omv:creationDate, bpm:released, omv:modificationDate, dct:valid, pav:curatedOn
Metrics	omv:numberOfClasses, omv:numberOfIndividuals, omv:numberOfProperties, omv:numberOfAxioms, bpm:maxDepth, bpm:maxChildCount, bpm:averageChildCount, bpm:classesWithOneChild, bpm:classesWithMoreThan25Children, bpm:classesWithNoDefinition, void:entities
Provenance	dct:source, prov:wasGeneratedBy, prov:wasInvalidatedBy, dct:accrualMethod, dct:accrualPeriodicity, dct:accrualPolicy

schema:hasPart or oboInOwl:hasSubset or adms:sample) to keep our model backward compatible.

The selection of default properties and equivalences is the more subjective part of our work. Our choices were driven by our needs and are subject to future modifications (see discussion Sect. 7.1). Somehow, they had to be made to nourish our project of demonstrating the power of harmonized metadata in an ontology repository. We shall certainly update these choices to accommodate small changes based on user feedback or experience. The latest complete list of properties and the equivalences implemented in AgroPortal are available via a web service call: http://data.agroportal.lirmm.fr/submission_metadata

6 Harnessing the Power of Unified Metadata in AgroPortal

Our goal was to implement a new metadata model into an ontology repository and give sense and valorize these metadata. We want to illustrate inside an ontology repository why ontology metadata are important and how they can be leveraged to provide new interesting insights to ontology developers and final users. We also believe that it is the role of an ontology repository to capture and give sense to metadata information interlinking ontologies together (e.g., the relation between ontologies).

6.1 Implementation Within AgroPortal

We have used the restricted list of Table 4 to implement a unified ontology metadata model within AgroPortal. We have added the 79 new properties into the original model (of 46 properties) precisely respecting the cardinalities of the properties.[21] This model is used to describe the ontologies being "hosted" within the portal, not the original ontology (to which only the original developers have authority on). Technically and formally speaking, this means that the metadata properties populated within AgroPortal apply to resources created by the portal, not the original URIs of the ontologies. For example, the National Agricultural Library Thesaurus (NALT) has for URI: http://lod.nal.usda.gov/nalt but the metadata properties, represented in JSON-LD within AgroPortal are assigned to the following resources: http://d

[21] With the objective of keeping our implementation simple, we have decided to add every new property to the Submission object. The range is generally either an URI or a String.

ata.agroportal.lirmm.fr/ontologies/NALT http://data.agroportal.lirmm.fr/ontologies/NALT/submissions/3.[22]

This gives us more flexibility when implementing a unified metadata model and facilitates the valorization and use of the metadata over all the ontologies, although it could create a confusion in terms of linked data being produced by the portal. For instance, an ontology creator may have used dc:title in the original ontology file but we will actually use the property omv:name for the metadata being stored on the portal.[23]

When an ontology is uploaded, AgroPortal extracts automatically most of the ontology metadata if they are included in the original file or populates some of them (e.g., metrics, endpoints, links, examples). Those values can manually be changed after by ontology developers or the portal administrators if they want to provide another value. We populate the 127 properties of the unified model by automatically parsing any of the 346 properties of the complete list presented in Sects. 5.2 and 5.3. When the original ontology file uses a property to capture metadata, we copy the value of this property to the default property chosen in the unified model and assign it to the resource created to represent the ontology within AgroPortal. Sometimes, the properties happen to be the same but often they are not. In the (very exceptional) case where multiple properties from the original file map to the same default property within the model, we aggregate the values or use multiple instances of the default property to keep all the original information. Then AgroPortal's REST web service will return the metadata of the hosted ontology, not the ones from the original file. Advanced users can still access the original metadata using the AgroPortal's SPARQL endpoint (http://sparql.agroportal.lirmm.fr/test) on which both URIs (hosted and original) are queryable. For example, if an ontology developer would use dc:creator for John, Alice and Tom and then pav:createdBy for NIH, WHO and NCBI, then AgroPortal' REST service API will return omv:hasCreator for John, Alice, Tom, NIH, WHO and NCBI. The SPARQL endpoint will return the original metadata.

For each ontology, available and uploaded in the portal, we collaborate with the ontology developers to extensively describe their metadata and we have spent a significant amount of time editing, curating and harmonizing the metadata. Information is generally found in other libraries (e.g., LovInra, VEST Registry, OBO Foundry, FAIRsharing) or identified in the publications, web sites, documentation, etc. found about the ontologies.

Now all the ontologies within AgroPortal are described with the same unified metadata model and we have invested a significant effort in editing metadata. This has resulted in three important new features for AgroPortal (Table 5):

- AgroPortal's ability to semantically capture and display a very large number of information about an ontology. The *Ontology Summary* page allows getting all the metadata information about a specific ontology. It helps users to know more about the ontologies they are using (or consider using); this will facilitate the ontology selection process and overall, make ontologies more FAIR. Plus, thanks to the portal architecture, all these data is formally described, with semantic web (standard) vocabularies and available as linked data (JSON-LD). In addition, we have entirely redesigned AgroPortal's ontology submission page to facilitate the edition of the metadata. Whenever possible, the user interface facilitates the selection of the metadata values, while in the backend those values are stored with standard URIs. For instance, the user interface will offer a pop-up menu to select the relevant license (CC, BSD, etc.) while the corresponding URI will be taken from the RDFLicense dataset (http://rdflicense.appspot.com). Knowledge organization systems types are taken from the NKOS Types Vocabulary of the Dublin Core initiative.[24] Natural languages are taken from the LEXVO vocabulary [66]. Ontology syntax values are provided by the W3C.[25] Some other values (the type of ontology or formality level) are taken as individuals from OMV. An example using the OntoBiotope ontology metadata page in AgroPortal is shown in Fig. 1.
- Advanced ontology search and selection thanks to AgroPortal's *Browse Ontologies* page (Fig. 2) which offers a convenient user interface with sorting, filtering, and facets that facilitate the identification of the ontology(ies) of interest. We now offer nine facets, based on the metadata, to filter ontologies including four new ones (content, natural language, formality level, type) as well as seven options to sort this list including two new ones (name, released date). These new features facilitate the process of selecting relevant ontologies.
- We have begun facilitating the comprehension of the agronomical ontology landscape by displaying diagrams, charts, and graphs about all the ontologies on the portal (average metrics, most used tools, leading contributors and organization, and more). We have created a new AgroPor-

[22] AgroPortal web service API requires a key to answer the data.agroportal calls. Users of the API will have to create an account on AgroPortal to get an APIkey (the same procedure is required with NCBO BioPortal). For the NCBO BioPortal, examples may be found here: http://data.bioontology.org/documentation#OntologySubmission.

[23] The perfect technical choice would have been the one of LOV, which only deals with a unified metadata following a specific announced vocabulary; however, we have demonstrated that only one or two standard vocabularies do not cover all the required fields for ontologies.

[24] http://wiki.dublincore.org/index.php/NKOS_Vocabularies (ANSI/NISO Z39.19-2005).

[25] https://www.w3.org/ns/formats/.

Table 5 Summary of metadata use within AgroPortal ontology repository

	Ontology Summary page	Browse Ontologies page	Landscape page
Description	Gives all the metadata information about a specific ontology	Allows to search, order and select ontologies using a facetted search approach, based on the metadata	Allows to explore the agronomical ontology landscape by automatically aggregating the metadata fields of each ontologies in explicit visualizations (charts, term cloud and graphs)
New compared to BioPortal	The whole "Additional Metadata" block which corresponds to properties from our new model. Plus the "Get my metadata back" buttons	Four additional ways to filter ontologies in the list (content, natural language, formality level, type) as well as two new options to sort this list (name, released date)	This page did not exist in the original BioPortal
Example (user interface)	http://agroportal.lirmm.fr/ontologies/ONTOBIOTOPE (see also Fig. 1)	http://agroportal.lirmm.fr/ontologies (see also Fig. 2)	http://agroportal.lirmm.fr/landscape (see also Fig. 3 to Fig. 9)
Example (API call)	http://data.agroportal.lirmm.fr/ontologies/ANAEETHES/submissions/2?display=all	http://data.agroportal.lirmm.fr/ontologies	E.g., to get omv:hasLicense property http://data.agroportal.lirmm.fr/submissions?display=hasLicense

tal *Landscape* page that displays metadata "by property" (as opposed as "by ontology" as in Fig. 1) by aggregating the metadata values (Sect. 6.2).

6.2 AgroPortal's Landscape Page

We have now a specific page dedicated to visualizing the ontology landscape in AgroPortal that facilitates analysis of the repository content. The landscape page helps to figure out what are some of the main domain of interests as well as common development practices when creating a vocabulary or ontology in agronomy. Of course, this information relies on the metadata extracted from the ontologies or edited on the portal. Such visualizations are also meant to motivate the ontology developers to document and describe more their ontologies. In the following, we present some views (figures) automatically created with the content of the repository from May 2017. Whenever possible, we also explicitly mention the metadata property used to generate the view.

Within AgroPortal (as in the original BioPortal) we organize the ontologies in relevant group and categories (Fig. 3): each time an ontology is uploaded into the portal, it is manually assigned a group and/or category. The groups allow bringing together ontologies from the same project or organization for better identification of the provenance. The categories are another way to classify ontologies in the portal by domain. The groups and categories are customizable and will be adapted in the future to reflect the evolution of the portal's content and community feedback. Another good aspect of the portal's architecture is that it provides URIs for any objects in the portal including groups and categories e.g., http://data.agroportal.lirmm.fr/categories/FARMING identifies the category "Farms and Farming Systems." External applications can now use these URIs to organize ontologies or tag them.

The most commonly adopted format is OWL (Fig. 4) which confirms the agronomy community has clearly turned to the W3C Recommendation for building ontologies. In addition, we already host six vocabularies in SKOS, which shall be a format that will grow in the future. It has been adopted, for instance, by the ANAEE Thesaurus, Agrovoc, NAL and CAB Thesaurus. Figure 4 also shows that most of the ontologies are in the range between 100 and 10 K classes (or concepts), although a few big resources have been uploaded. The metrics in AgroPortal are automatically computed by the OWL-API, but they can be overridden manually. The size of the ontology is generally the number of classes (except with the SKOS format, where it is the number of individuals).

Ontology labels are mostly in English (Fig. 5) although we have seven resources that offer French labels (mostly because of our French collaborators). Multilingual resources include Agrovoc and NAL Thesaurus. Figure 5 also shows that among the 31 ontologies that have explicitly defined licensing information, all of them are openly accessible with different licenses. Note AgroPortal can also host private ontologies or restrict download for public ones.

The type and formality level of resources are described in Fig. 6. The number of upper level ontologies (not specifically dedicated to agriculture) is maintained low and not surprisingly most of the ontologies are domain or application ontologies. Acknowledging the "ambiguity" of these information, as there are no standard definitions of the type and formality level of a knowledge organization system, we do think this information is useful and may help to select the right resources for a given task [14].

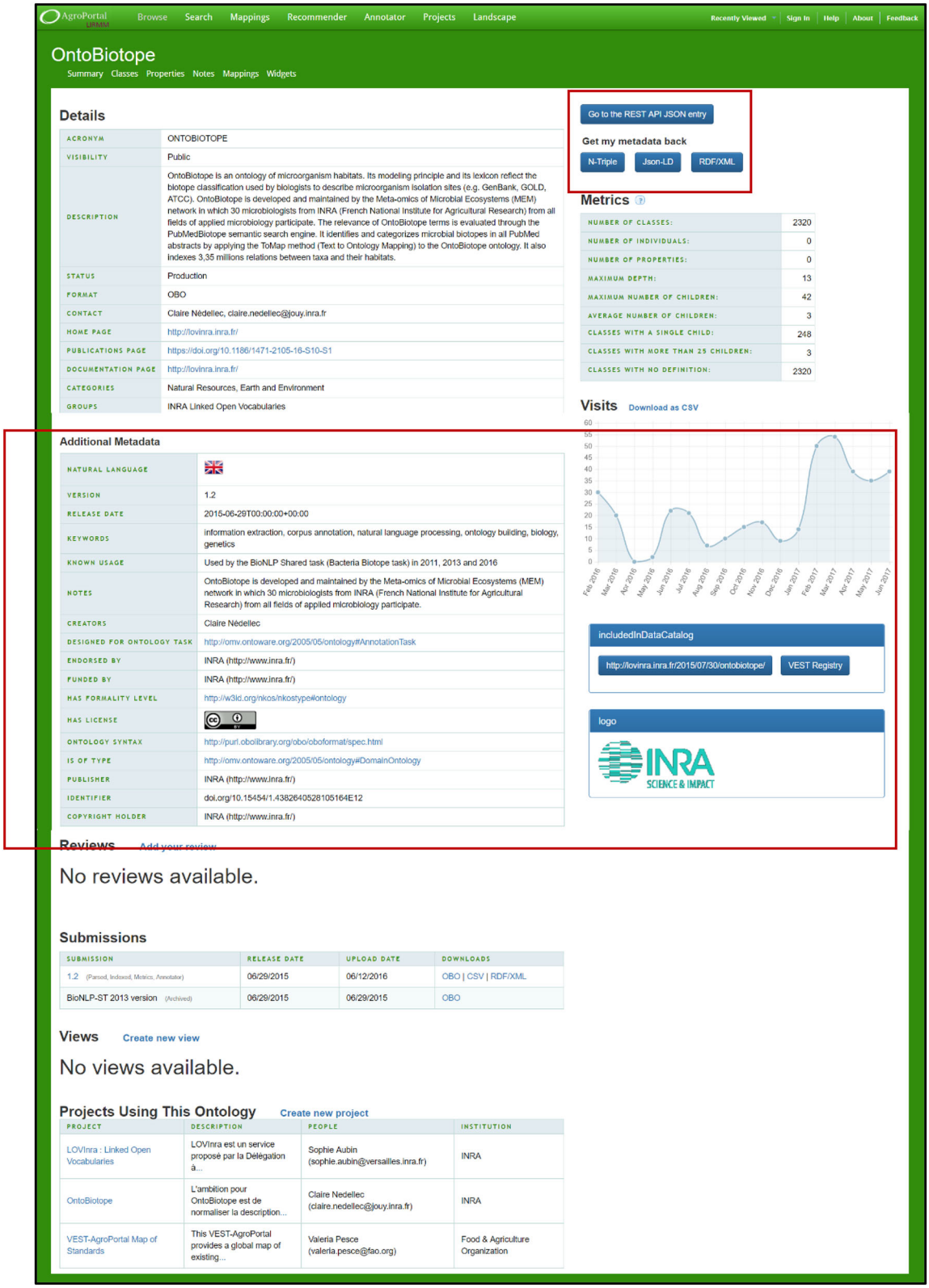

Fig. 1 Screenshot of the *Ontology Summary* page for the OntoBiotope ontology (http://agroportal.lirmm.fr/ontologies/ONTOBIOTOPE). The section "Additional Metadata" has been automatically extracted from the content of the original ontology file or edited by AgroPortal admin or the ontology owner. We have not yet implemented the change at the user interface level to display nice values rather than the raw URIs. This will be done in the next future

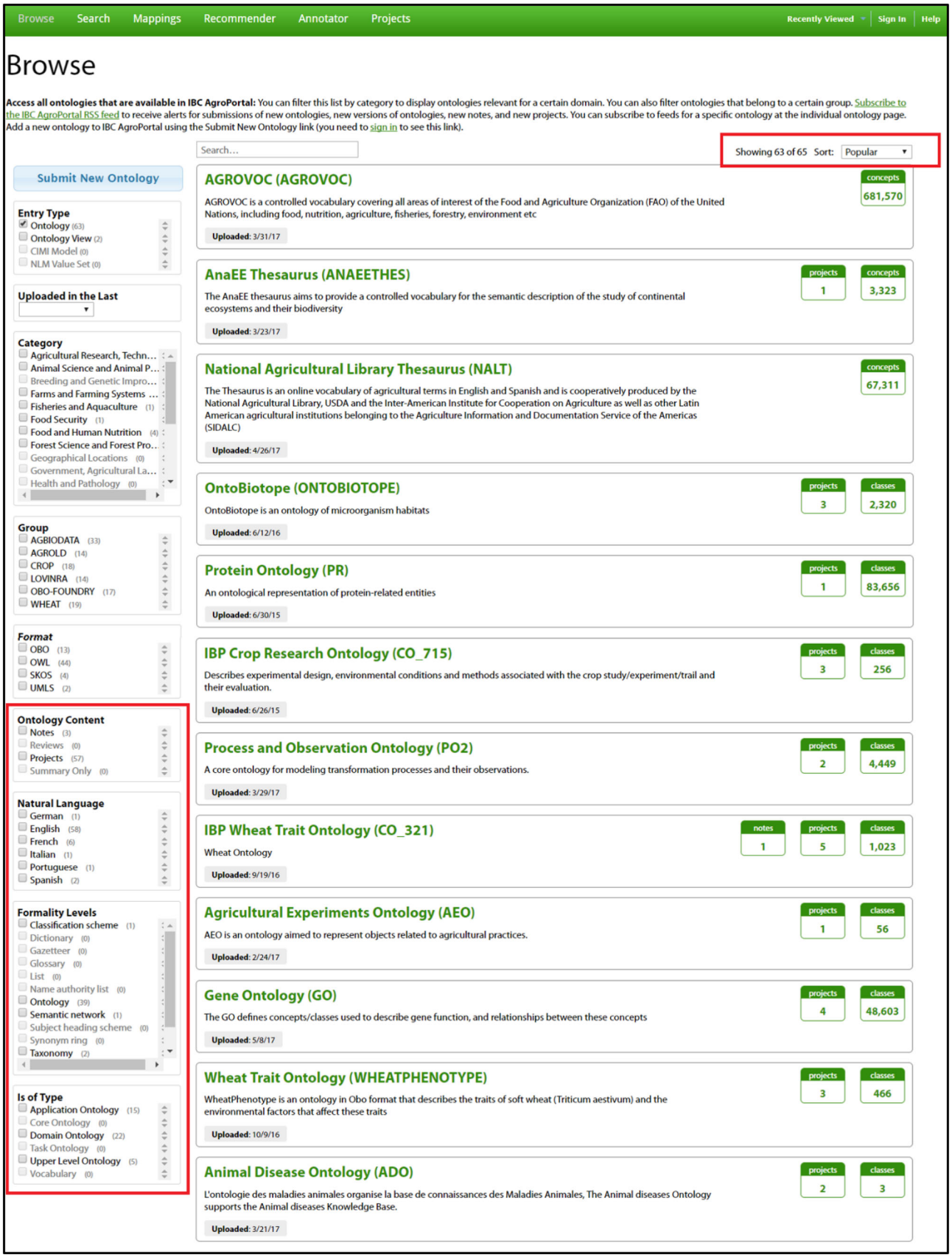

Fig. 2 Screenshot of the *Browse Ontologies* page. Facetted search (left hand side) and sorting (top right corner) offer new ways to select ontologies

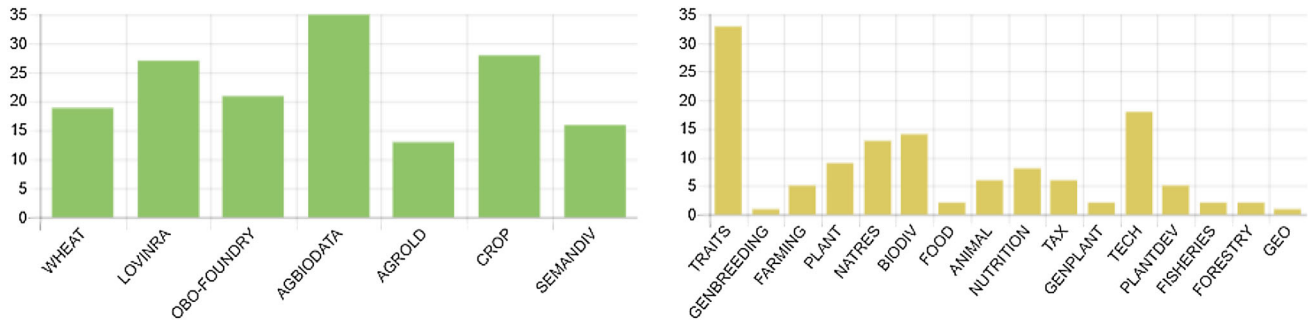

Fig. 3 Distribution of ontologies by Group (bpm:group) and Categories (bpm:hasDomain)

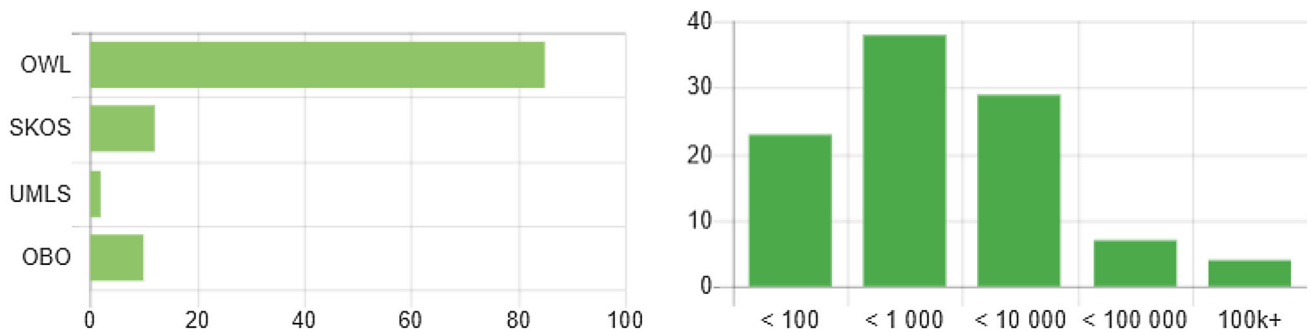

Fig. 4 Ontologies by format (omv:hasOntologyLanguage) and sizes (bpm:metrics.classes or bpm:metrics.individuals)

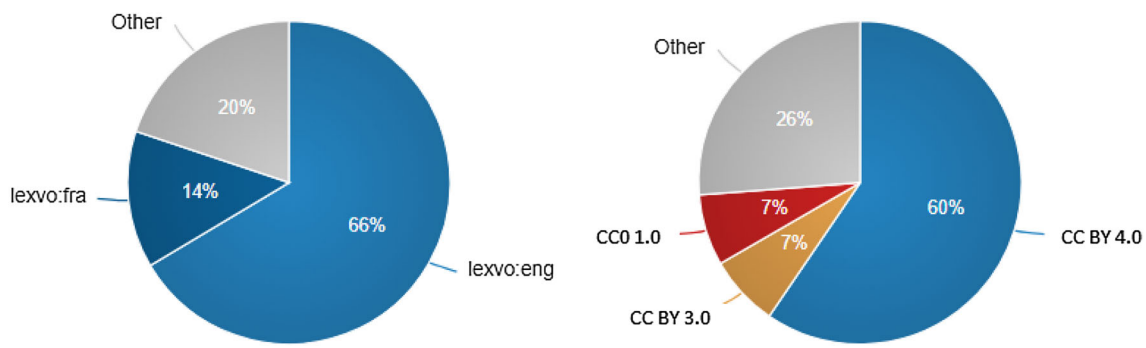

Fig. 5 Natural languages (omv:naturalLanguage) used for labels and licenses (omv:hasLicense) of ontologies

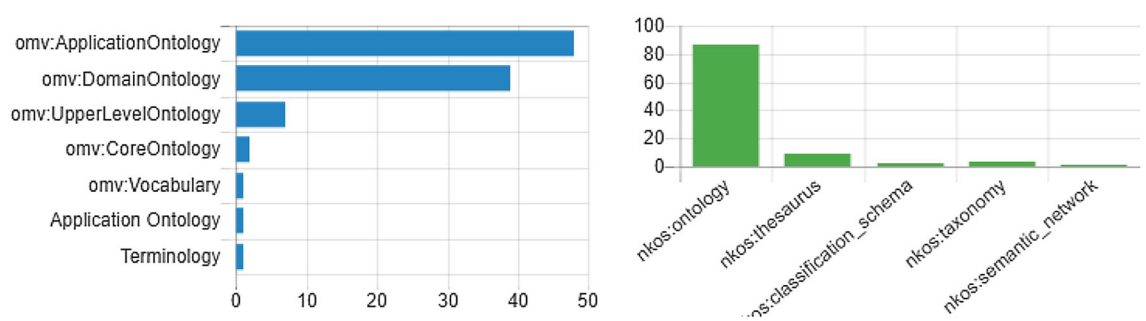

Fig. 6 Ontology types (omv:isOfType) and formality levels (omv:hasFormalityLevel)

Figure 7 is an aggregation (term cloud) of several properties that relate ontologies and organizations. Such a view is interesting to identify which organizations are the most involved in funding, adopting or endorsing ontologies. Figure 8 is a similar cloud showing which ontologies are the most actively commented, reviewed or used within research

Fig. 7 Most mentioned organizations (aggregation as a term cloud from the properties dct:publisher, foaf:fundedBy and omv:endorsedBy)

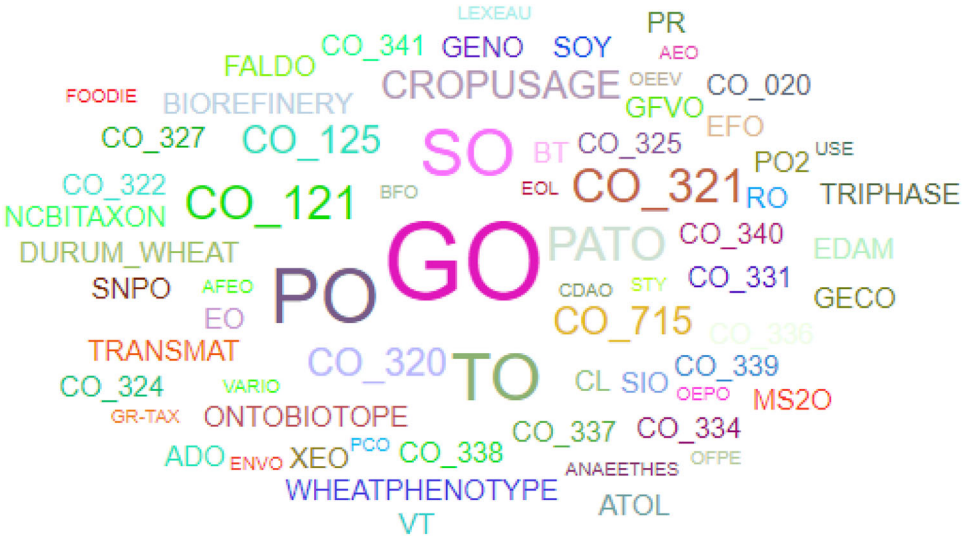

Fig. 8 Most active ontologies (count aggregation of omv:notes, bpm:reviews and bpm:projects per ontology)

projects. Indeed, AgroPortal features a few community features [67] such as: *ontology reviews* or *notes* that can be attached in a forum-like mode to a specific ontology or class, in order to discuss the ontology (its design, use, or evolution) or allow users to propose changes to a certain class. Plus, AgroPortal provides a project list edited by its users that materialize the ontology-project relation (http://agroportal.lirmm.fr/projects), i.e., which project uses which ontologies.

The new metadata model allows capturing multiple relations between ontologies or between ontologies and external resources. For instance, relations to capture that an ontology is aligned to another one, represents knowledge from the same domain, is compatible or incompatible with another one, imports or uses another one, is translated from or more generally related to another one. We have used 14 of these relations to automatically represent AgroPortal's ontologies network. Figure 9 shows the cluster of the ontologies maintained and extended within the Planteome project [25]. It captures the information that all the Crop Ontologies (CO_*) are aligned to the Trait Ontology, itself interconnected to the Plant Ontology and Plant Environment ontology. The Soy Ontology, developed outside of the Crop Ontology project, also appears as related to both TO and CO_336 (the Soybean Ontology developed within the Crop Ontology project). Figure 9 is only a subset of the network. The landscape page within AgroPortal displays the whole network and filters it per ontology relations.

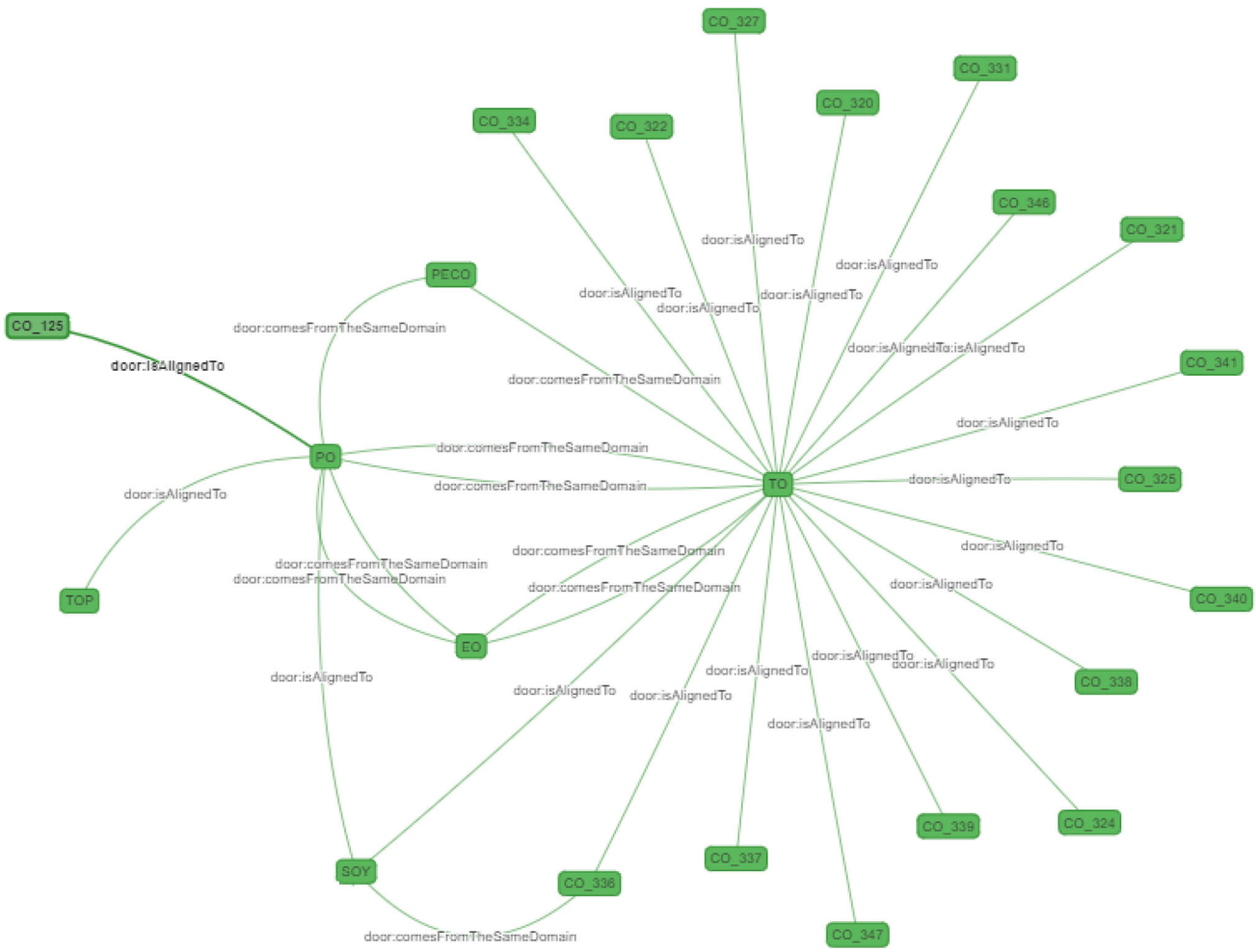

Fig. 9 Subset of the ontology network showing the relations between reference plant ontologies (here properties door:isAlignedTo and door:comeFromTheSameDomain)

6.3 User Appreciation Survey

To evaluate the impact and appreciation of the new features enabled by our changes in AgroPortal's ontology metadata model, we conducted a survey with typical five-level Likert scale questions. Each question asked for the participant's opinion about how much the new page (or new features in the page) "helped identifying and selecting" relevant ontologies (except for the ontology submission edition page, which was concerned about editing ontology metadata). With this survey, we liked to assess AgroPortal's new metadata model ability to ease ontology identification and selection. Plus, we asked open questions about each page to get users inputs in terms of how to improve ontology metadata within AgroPortal in the future. The survey was sent only to the AgroPortal users mailing list which had 131 members then. We had 32 responses that are analyzed hereafter. 2/3 of the participants were both users and administrators of one or several ontologies in AgroPortal. The last third was only regular AgroPortal users who usually search and find relevant ontologies and concepts. The questions and responses are presented in Table 6.

Globally, the helpfulness of the pages was clearly established by the survey with almost ¾ of positive responses on average for all questions. Displaying more metadata on the Summary page and being able to filter out ontologies with metadata facets on the Browse page was much appreciated. The Landscape page was ranked as a bit less "useful" than the others (with 53.2% responses explicitly positive) getting still some positive feedbacks and relevant criticisms. An additional question related to the usefulness of the page to "understand about the ecosystem of ontologies in agronomy and close related domains" obtained 75% of positive responses. The absence of response in the "Not at all" and very limited responses in the "Not so" columns show that everyone agrees about the role of metadata when identifying and selecting ontologies. Still, the exploitation of metadata to facilitate this process is improvable. Among the comments

Table 6 User appreciation survey responses (percentage)

Question/page	Extremely helpful	Very helpful	Somewhat helpful	Not so helpful	Not at all helpfull
New ontology summary page	15.6	59.4	21.9	3.1	0
New ontology browse page	31.3	56.3	9.3	3.1	0
New landscape page	12.6	40.6	37.5	9.3	0
New ontology submission edition page (optional)[a]	19	57.2	19	4.8	0
Average	19.63%	53.38%	21.93%	5.08%	0%

[a]Only 21 responses for this last question

on: (1) the Summary page, some were about improving the user interface by keeping only the relevant fields and using something else than URLs or URIs. (2) the Browse page, most were positive as facets are often appreciated to search information, although the lack of description of the facets was often reported. (3) the Landscape page, many comments were requesting a better integration with the rest of AgroPortal (e.g., links back), merging some information (e.g., Figure 6) and some were about pointing out the importance of curating the metadata to create good value in this page.

7 Discussions

According to us, among the main limitations of OMV that might explain why it is not much adopted today are: (1) the fact that it did not reuse any other metadata vocabulary;[26] (2) it was never included in a common ontology editor such as Protégé—it would have highly facilitated the adoption of the vocabulary if ontology developers would have had only to fill out a few forms directly in their preferred ontology edition software; (3) the metadata properties were never really used and valorized by ontology libraries which would have been the best way to incite to fill them up.

In the following, we come back on each of these aspects to discuss the need for a better harmonization of standard vocabularies used to described ontology metadata. Besides our work driven by the AgroPortal project, this effort may be generalized to propose recommendations and guidelines to (1) ontology developers when describing their ontologies; (2) ontology repository or library developers to harmonize their platforms.

[26] Although we acknowledge that in 2005, there was not as vocabularies as today, important standards such as OWL, Dublin Core or FOAF may have been used at that time.

7.1 Need for Metadata Authoring Guidelines and for Harmonization of Existing Metadata Vocabularies

The analysis of the existing metadata vocabularies and practices (Sect. 4) showed there is a clear need for better metadata authoring guidelines for the community of ontology developers and a need of harmonization of existing metadata vocabularies. MOD1.0 [7] was a first attempt to address OMV's limitation of not relying on any other vocabularies but was not "mapped" itself to OMV while being very similar. Plus, it still missed numerous relevant properties to capture information about ontologies. More recently, the authors joined their efforts and proposed a new version of MOD (refer as MOD 1.2) [48].[27] The revision carried out from multiple aspects (e.g., new labels, structural changes, and design principles) to overcome some of the limitations of MOD 1.0 and to enrich it further influenced also by our work on AgroPortal. MOD1.2 contains 88 properties taken from DCAT, DCT, DOAP, FOAF, OMV, OWL, PAV, PROV, RDFS, SD and VOAF but creates only 13 new properties in the MOD namespace. Future extended versions (MOD2.0 and more) shall contain at least equivalent property for each of the 127 of AgroPortal's new metadata model. Note that because MOD development is free from any implementation constraints, we have not always selected in MOD1.2 the same default properties than in AgroPortal's unified metadata model. In [48], we also describe the application goals of MOD1.2 and illustrate our experimental results with SPARQL queries that can be run on properly defined metadata.

MOD 1.2 is a recent initiative and still a temporary proposition. It is understandable that to achieve community adoption, this work needs to engage more people, with the ultimate goal of producing a community standard endorsed by a

[27] https://github.com/sifrproject/MOD-Ontology.

standardization body such as W3C. MOD 1.2 was recently introduced to the Research Data Alliance *Vocabulary and Semantic Services Interest Group* (VSSIG).[28] Future work will now happen in the context of the "ontology metadata" task group of the VSSIG. Among the current studied propositions is to implement MOD2.0 as a profile of DCAT. We shall also make sure we will enforce—or enable operationalization of—the recently published MIRO guidelines [6].

7.2 Metadata Edition

Another important aspect in metadata is that almost no one really like filling them in; therefore, how can we facilitate metadata editing for ontology developers? Within AgroPortal, we have entirely redesigned the ontology information edition page and have tried to build it in a way that will both facilitate the edition and not freak out the editors with a basic list of 127 properties to fill in. However, this page will need improvements. We do envision paths for the future:

- Metadata should be as much as possible generated or predicted automatically either by the ontology edition software or by external tools,[29] e.g., software used, dates, languages.
- It is the role of ontology edition software to actually support (some) metadata edition functionalities. It would highly facilitate the task (and the emergence of a standard vocabulary) if ontology editors would only need to fill out a few forms directly in their preferred ontology edition software. Indeed, as seen in Sect. 4.2 properties available for editing (or even better, automatically generated) within the ontology editor are inclined to be well used.
- It is the role of ontology libraries to facilitate the edition, generation and prediction of ontology metadata for properties that take their senses within a community-based library, e.g., relations between ontologies, reviews, related projects, etc. When relevant, the libraries should offer a mechanism to easily export the metadata edited or generated in order for ontology developers to include it in the original ontology file for other systems to use it. Within AgroPortal, we have developed such a mechanism on the *Ontology Summary* page.[30] In addition of an API call, the "Get my metadata back" buttons allow ontology developers, on a simple click, to download the metadata stored within the portal in RDF/XML, JSON-LD or N-triples syntax to copy/paste within the original ontology. An additional question related to the interest of this functionality was included within the survey presented Sect. 6.3 and obtained 62.5% positive responses. Right now, this feature will return metadata following AgroPortal's model, but when MOD2.0 will be available or any community adopted standard, it will return the metadata with respect to this standard.[31]

In the future, we plan to discuss with the Protégé development team the integration of some of the listed properties in the software, so that developers can edit them in the ontology development process. We are also considering results of the CEDAR project (http://metadatacenter.org) in terms of metadata prediction and edition [38].

7.3 Automatic Ontology Selection and Recommendation

An unified metadata model can also be leveraged by automatic ontology selection tools such as the Recommender also available in Agro/BioPortal [3, 68] which relies mostly on the content of ontologies to recommend them. For instance, the whole network built out of ontologies relations (Fig. 9) will help users to figure out which are the key relevant ontologies to rely on. As another example, searching "for ontologies" often rely on "searching inside" ontologies (method based on coverage) which is not very often satisfactory when instead metadata should be used. For example, searching "anatomy" in BioPortal Search will return a bunch of popular ontologies that contains the term anatomy, but the Foundational Model of Anatomy, which is the reference ontology about human anatomy will not show up in the results. To identify FMA, someone needs to browse the ontologies and filter ontologies with the word anatomy in the ontology name or description. Or better, he or she might use the "Anatomy" ontologies category that BioPortal defines. In both cases, this relies on metadata, not on the content of the ontology.

8 Conclusion and Future Work

In this paper, we have shown the impact of unified and harmonized metadata within an ontology repository. We have explained how it facilitates ontology description, selection and helps to capture the global landscape of ontologies from a given domain. Thanks to this new unified model served by a stable API, metadata descriptions of AgroPortal ontologies have already been automatically harvested by two external ontology libraries: the Agrisemantics Map of Data Standards (http://vest.agrisemantics.org) and FAIRsharing (http://fairsharing.org).

[28] The RDA Interest Group was reconfigured in 2017 (https://www.rd-alliance.org/groups/vocabulary-services-interest-group.html).

[29] For instance, BioPortal uses the OWL-API to generate metrics. Protégé also does but does not save these metrics inside the ontology.

[30] This feature has been recently developed and is still in beta mode.

[31] Before completely changing AgroPortal's model, we believe each library could at least import/export MOD2.0 compliant metadata.

Our motivation was first to make a review of the available vocabularies to describe ontologies or other kinds of resources (dataset, vocabulary, project, document) and pick up the properties that would be relevant for describing ontologies. Since the OMV initiative in 2005, there have been multiple propositions especially with the emergence of the web of data. Our goal was then to identify the redundancy and lacks between these vocabularies by regrouping the properties into a restricted unified model that we have implemented in the AgroPortal ontology repository. We have worked on our side and in partnership with our users to fill the metadata and in parallel developed new user interfaces. This has resulted in multiple new features within AgroPortal that we have presented and have been appreciated by our users as facilitating the ontology identification and selection processes.

We can now come back on addressing some concrete motivational use cases described in Sect. 2:

- The new ontology metadata model has been driven by and finally implemented within AgroPortal and the French SIFR BioPortal. The new model makes the description of the ontologies more complete and is unified for all the ontologies.
- The properties omv:hasLicense, dct:publisher, cc:morePermissions and schema:copyrightHolder can now be used to precisely describe the licensing information about the ontologies endorsed by the Wheat Data Interoperability working group. The endorsement itself is also captured by the property omv:endorsedBy.
- LovINRA ontologies can now be explicitly and unambiguously classified by syntax (omv:hasOntologySyntax), format (omv:hasOntologyLanguage), type (omv:isOfType) and formality level (omv:hasFormalityLevel).
- The alignment relations between the Crop Ontology trait ontologies and the Plant Trait Ontology are now captured by door:isAlignedTo and other relations such as door:comesFromTheSameDomain and door:ontologyRelatedTo.

We did not pursue the goal of integrating all the reviewed vocabularies into a new "integrated vocabulary" that could become a standard for describing ontologies (e.g., a new OMV), although the clear need for metadata authoring guidelines and for harmonization of existing metadata vocabularies has also been discussed. We are currently working in generalizing this work within a new version of MOD that would merge and harmonize existing ones. A generalization of this work is studied in a community-driven standardization effort in the context of the RDA *Vocabulary and Semantic Services Interest Group*. We are also discussing with the Stanford NCBO project how to merge back our contributions to the technology into the NBCO BioPortal.

In the future, we want to be able to describe more the usage of ontologies by defining/extending (1) generic tasks for which ontology are used (annotation, indexing, search, reasoning, etc.) and (2) small examples of usages of the ontologies. We also plan to use the same metadata analysis approach to suggest ontology development guidelines based on community practices. For instance, by looking at the most used properties to describe ontologies or their classes. In the future, by integrating more relevant ontologies and vocabularies into AgroPortal and cautiously describing them, we hope to offer a reference portal to identify and use knowledge organizations systems in agronomy, food, plant sciences and biodiversity. We will continue our metadata edition and curation effort to be sure to provide the community with the best descriptions for ontologies available.

Acknowledgements This work is partly achieved within the Semantic Indexing of French biomedical Resources (SIFR—www.lirmm.fr/sifr) project that received funding from the French National Research Agency (Grant ANR-12-JS02-01001), the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie Grant agreement No 701771, the NUMEV Labex (Grant ANR-10-LABX-20), the Computational Biology Institute of Montpellier (Grant ANR-11-BINF-0002) as well as by University of Montpellier and the CNRS. This work has also been partially funded by the Indian Statistical Institute under the Start-Up Grant project. We also thank the National Center for Biomedical Ontologies for help and time spent with us in deploying the AgroPortal and all our users who have taken some time to author/review the metadata of their ontologies and answer our appreciation survey.

Author Contributions CJ conceived of the project, provided the scientific direction and wrote this manuscript. AT worked on the new metadata model and on description of the ontologies with help of the AgroPortal user community. VE implemented the new metadata model, additional features and the landscape page in AgroPortal. BD worked on vocabulary analysis as well reflections on MOD and also contributed to the survey analysis. All authors declare not conflict of interest and approved the final manuscript.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Ding L, Finin T, Joshi A, Peng Y, Cost RS, Sachs J, Pan R, Reddivari P, Doshi V (2004) Swoogle: a semantic web search and metadata engine. In: Grossman DA, Gravano L, Zhai C, Herzog O, Evans D (eds) 13th ACM conference on information and knowledge management, CIKM'04. ACM, Washington DC, USA
2. Bizer C, Heath T, Berners-Lee T (2009) Linked data—the story so far. Sem Web Inf Syst 5:1–22
3. Martinez-Romero M, Jonquet C, O'Connor MJ, Graybeal J, Pazos A, Musen MA (2017) NCBO Ontology Recommender 2.0: an enhanced approach for biomedical ontology recommendation. Biomed Sem 8:21. https://doi.org/10.1186/s13326-017-0128-y

4. Chao T (2015) Mapping methods metadata for research data. Digit Curation 10:82–94
5. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S-A, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR guiding principles for scientific data management and stewardship. Sci Data 3:160018. https://doi.org/10.1038/sdata.2016.18
6. Matentzoglu N, Malone J, Mungall C, Stevens R (2018) MIRO: guidelines for minimum information for the reporting of an ontology. J Biomed Sem 9:6
7. Dutta B, Nandini D, Kishore G (2015) MOD : metadata for ontology description and publication. In: International conference on Dublin core & metadata applications, DC'15, Sao Paulo, Brazil, pp 1–9
8. Naskar D, Dutta B (2016) Ontology libraries : a study from an ontofier and an ontologist perspectives. In: 19th international symposium on electronic theses and dissertations, ETD'16, Lille, France, pp 1–12
9. Vandenbussche P-Y, Atemezing GA, Poveda-Villalon M, Vatant B (2014) Linked Open Vocabularies (LOV): a gateway to reusable semantic vocabularies on the Web. Sem Web 1:1–5
10. Côté RG, Jones P, Apweiler R, Hermjakob H (2006) The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries. BMC Bioinformatics 7:97
11. Rueda C, Bermudez L, Fredericks J (2009) The MMI ontology registry and repository: a portal for marine metadata interoperability. In: MTS/IEEE Biloxi—marine technology for our future: global and local challenges, OCEANS'09, Biloxi, MS, USA, p 6
12. Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith NB, Jonquet C, Rubin DL, Storey M-A, Chute CG, Musen MA (2009) BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic Acids Res 37:170–173
13. Sabou M, Lopez V, Motta E (2006) Ontology selection on the real semantic web: how to cover the queens birthday dinner? In: Staab S, Svátek V (eds) 15th international conference on knowledge engineering and knowledge management managing knowledge in a world of networks, EKAW'06. Springer, Podebrady, Czech Republic, pp 96–111
14. Park J, Oh S, Ahn J (2011) Ontology selection ranking model for knowledge reuse. Expert Syst Appl 38:5133–5144
15. Malone J, Stevens R, Jupp S, Hancocks T, Parkinson H, Brooksbank C (2016) Ten simple rules for selecting a bio-ontology. PLoS Comput Biol 12:6
16. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg LJ, Eilbeck K, Ireland A, Mungall CJ, Consortium T.O.B.I., Leontis N, Rocca-Serra P, Ruttenberg A, Sansone S-A, Scheuermann RH, Shah NH, Whetzel PL, Lewis S (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol 25:1251–1255
17. Toulet A, Emonet V, Jonquet C (2016) Modèle de métadonnées dans un portail d'ontologies. In: Diallo G, Kazar O (eds) 6èmes Journées Francophones sur les Ontologies, JFO'16, Bordeaux, France
18. Jonquet C, Toulet A, Arnaud E, Aubin S, Dzalé Yeumo E, Emonet V, Graybeal J, Laporte M-A, Musen MA, Pesce V, Larmande P (2018) AgroPortal: a vocabulary and ontology repository for agronomy. Comput Electron Agric 144:126–143
19. Whetzel PL, Team N (2013) NCBO technology: powering semantically aware applications. Biomed Sem 4S1:49
20. Jonquet C, Annane A, Bouarech K, Emonet V, Melzi S (2016) SIFR BioPortal : Un portail ouvert et générique d'ontologies et de terminologies biomédicales françaises au service de l'annotation sémantique. In: 16th Journées Francophones d'Informatique Médicale, JFIM'16, Genève, Suisse, p 16
21. Jonquet C, Shah NH, Musen MA (2009) The open biomedical annotator. In: American medical informatics association symposium on translational bioinformatics, AMIA-TBI'09, San Francisco, CA, USA, pp 56–60
22. Walls RL, Deck J, Guralnick R, Baskauf S, Beaman R, Blum S, Bowers S, Buttigieg PL, Davies N, Endresen D, Gandolfo MA, Hanner R, Janning A, Krishtalka L, Matsunaga A, Midford P, Morrison N, Ó Tuama É, Schildhauer M, Smith B, Stucky BJ, Thomer A, Wieczorek J, Whitacre J, Wooley J (2014) Semantics in support of biodiversity knowledge discovery: an introduction to the biological collections ontology and related ontologies. PLoS ONE 9:e89606
23. Meng X (2012) Special issue—agriculture ontology. Integr Agric 11:1
24. Quesneville H, Dzale Yeumo E, Alaux M, Arnaud E, Aubin S, Baumann U, Buche P, Cooper L, Ćwiek-Kupczyńska H, Davey RP, Fulss RA, Jonquet C, Laporte M-A, Larmande P, Pommier C, Protonotarios V, Reverte C, Shrestha R, Subirats I, Venkatesan A, Whan A (2017) Developing data interoperability using standards: a wheat community use case. F1000Research 6:1843. https://doi.org/10.12688/f1000research.12234.2
25. Cooper L, Meier A, Laporte M-A, Elser JL, Mungall C, Sinn BT, Cavaliere D, Carbon S, Dunn NA, Smith B, Qu B, Preece J, Zhang E, Todorovic S, Gkoutos G, Doonan JH, Stevenson DW, Arnaud E, Jaiswal P (2018) The Planteome database: an integrated resource for reference ontologies, plant genomics and phenomics. Nucleic Acids Res 46:D1168–D1180
26. Shrestha R, Matteis L, Skofic M, Portugal A, McLaren G, Hyman G, Arnaud E (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 3:1–10
27. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT et al (2000) Gene ontology: tool for the unification of biology. Nat Genet 25:25–29
28. Mabee PM, Ashburner M, Cronk Q, Gkoutos GV, Haendel M, Segerdell E, Mungall C, Westerfield M (2007) Phenotype ontologies: the bridge between genomics and evolution. Trends Ecol Evol 22:345–350
29. Zeng ML (2008) Metadata. Neal-Schuman, New York
30. Obrst L, Gruninger M, Baclawski K, Bennett M, Brickley D, Berg-Cross G, Hitzler P, Janowicz K, Kapp C, Kutz O, Lange C, Levenchuk A, Quattri F, Rector A, Schneider T, Spero S, Thessen A, Vegetti M, Vizedom A, Westerinen A, West M, Yim P (2014) Semantic web and big data meets applied ontology. Appl Ontol 9:155–170
31. Graybeal J, Isenor AW, Rueda C (2012) Semantic mediation of vocabularies for ocean observing systems. Comput Geosci 40:120–131
32. Baclawski K, Schneider T (2009) The open ontology repository initiative: requirements and research challenges. In: Tudorache T, Correndo G, Noy N, Alani H, Greaves M (eds) Workshop on collaborative construction, management and linking of structured knowledge, CK'09. CEUR-WS.org, Washington, DC, USA, p 10
33. Till M, Kutz O, Codescu M (2014) Ontohub: a semantic repository for heterogeneous ontologies. In: Theory day in computer science, DACS'14, Bucharest, Romania, p 2
34. Sabou M, Lopez V, Motta E, Uren V (2006) Ontology selection: ontology evaluation on the real semantic web. In: Vrandecic D, Suarez-Figueroa MC, Gangemi A, Sure Y (eds) 4th international

EON workshop, evaluation of ontologies for the web, EON'06. CEUR-WS.org, Edinburgh, Scotland, UK
35. Tejo-Alonso C, Berrueta D, Polo L, Fernández S (2010) Current practices and perspectives for metadata on web ontologies and rules. Metadata Sem Ontol 7:10
36. Palma R, Hartmann J, Haase P (2008) Ontology metadata vocabulary for the semantic web, Report, v2.4, pp 1–85
37. Weibel S, Kunze J, Lagoze C, Wolf M (1998) Dublin core metadata for resource discovery, RFC 2413, Internet Engineering Task Force
38. Musen MA, Bean CA, Cheung K-H, Dumontier M, Durante KA, Gevaert O, Gonzalez-Beltran A, Khatri P, Kleinstein SH, O'Connor MJ, Pouliot Y, Rocca-Serra P, Sansone S-A, Wiser JA (2015) The CEDAR team: the center for expanded data annotation and retrieval. Am Med Inform Assoc 22:1148–1152
39. Dumontier M, Gray AJG, Marshall MS, Alexiev V, Ansell P, Bader G, Baran J, Bolleman JT, Callahan A, Cruz-Toledo J, Gaudet P, Gombocz EA, Gonzalez-Beltran AN, Groth P, Haendel M, Ito M, Jupp S, Juty N, Katayama T, Kobayashi N, Krishnaswami K, Laibe C, Le Novère N, Lin S, Malone J, Miller M, Mungall CJ, Rietveld L, Wimalaratne SM, Yamaguchi A (2016) The health care and life sciences community profile for dataset descriptions. PeerJ 4:e2331
40. McQuilton P, Gonzalez-Beltran A, Rocca-Serra P, Thurston M, Lister A, Maguire E, Sansone S-A (2016) BioSharing: curated and crowd-sourced metadata standards, databases and data policies in the life sciences. Database. https://doi.org/10.1093/database/baw075
41. McGuinness DL (2003) In: Fensel D, Hendler J, Lieberman H, Wahlster W (eds) Spinning the semantic web: bringing the World Wide Web to its full potential, Chapter 6. MIT Press, Cambridge, MA, pp 171–194
42. Zeng ML (2008) Knowledge organization systems (KOS). Knowl Organ 35:160–182
43. Zen ML, Zeng ML (2008) Metadata for terminology/KOS resources. In: 8th networked knowledge organization systems workshop, Washington, DC, USA
44. Hartmann J, Haase P (2005) Ontology metadata vocabulary and applications, pp 906–915
45. Ding Y, Fensel D (2001) Ontology library systems: the key to successful ontology re-use. In: 1st semantic web working symposium, SWWS'01. CEUR-WS.org, Stanford, CA, USA, pp 93–112
46. Hartmann J, Palma R, Gómez-Pérez A (2009) Ontology repositories. In: Staab S, Studer R (eds) Handbook on ontologies. Springer, Berlin, Heidelberg, pp 551–571
47. D'Aquin M, Noy NF (2012) Where to publish and find ontologies? A survey of ontology libraries. Web Sem 11:96–111
48. Dutta B, Toulet A, Emonet V, Jonquet C (2017) New generation metadata vocabulary for ontology description and publication. In: Garoufallou E, Virkus S, Alemu G (eds) 11th metadata and semantics research conference, MTSR'17, Tallinn, Estonia
49. Allocca C, d'Aquin, M, Motta E (2009) DOOR—towards a formalization of ontology relations. In: International conference on knowledge engineering and ontology development, KEOD'09, Madera, Portugal, pp. 13–20
50. Vandenbussche P-Y, Vatant B (2012) Metadata recommendations for linked open data vocabularies. Report, v1.1. https://lov.linkeddata.es/Recommendations_Vocabulary_Design.pdf
51. Min H, Turner S, de Coronado S, Davis B, Whetzel PL, Freimuth RR, Solbrig HR, Kiefer R, Riben M, Stafford GA, Wright L, Ohira R (2016) Towards a standard ontology metadata model. In: Jaiswal P, Hoehndorf R (eds) 7th international conference on biomedical ontologies, ICBO'16, Poster Session, Corvallis, Oregon, USA, p 6
52. Hausenblas KAM (2009) Describing linked datasets—on the design and usage of void, the vocabulary of interlinked datasets. In: Linked data on the web workshop, LDOW'09, Madrid, Spain
53. Juty N, Novère N Le, Laibe C (2012) Identifiers.org and MIRIAM Registry: community resources to provide persistent identification. Nucleic Acids Res 40:580–586
54. Tirmizi SH, Aitken S, Moreira DA, Mungall C, Sequeda J, Shah NH, Miranker DP (2011) Mapping between the OBO and OWL ontology languages. Biomed Sem 2:16
55. Mäkelä E (2014) Aether–Generating and viewing extended VoID statistical descriptions of RDF datasets. In: Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics). Springer, Crete, pp 429–433
56. Ceusters W (2012) An information artifact ontology perspective on data collections and associated representational artifacts. In: Mantas J et al (ed) 24th international conference of the European federation for medical informatics, MIE'12. IOS Press, Pisa, pp 68–72
57. Dumontier M, Baker CJ, Baran J, Callahan A, Chepelev L, Cruz-Toledo J, Del Rio NR, Duck G, Furlong LI, Keath N, Klassen D, McCusker JP, Queralt-Rosinach N, Samwald M, Villanueva-Rosales N, Wilkinson MD, Hoehndorf R (2014) The Semantic-science Integrated Ontology (SIO) for biomedical research and knowledge discovery. J Biomed Sem 5:14
58. Fiorelli M, Stellato A, McCrae JP, Cimiano P, Pazienza MT (2015) LIME: the metadata module for OntoLex. In: Gandon F, Sabou M, Sack H, d'Amato C, Cudré-Mauroux P, Zimmermann A (eds) 12th European semantic web conference, ESWC'15. Springer, Portoroz, pp 321–336
59. McCrae J, Spohr D, Cimiano P (2011) Linking lexical resources and ontologies on the semantic web with lemon. In: Antoniou G, Grobelnik M, Simperl E, Parsia B, Plexousakis D, DeLeenheer P, Pan JZ (eds) 8th extended semantic web conference, ESWC'11. Springer, Heraklion, Crete, pp 245–259
60. Ong E, Xiang Z, Zhao B, Liu Y, Lin Y, Zheng J, Mungall C, Courtot M, Ruttenberg A, He Y (2016) Ontobee: a linked ontology data server to support ontology term dereferencing, linkage, query and integration. Nucleic Acids Res 45:D347–D352
61. Hoehndorf R, Slater L, Schofield PN, Gkoutos GV (2015) Aber-OWL: a framework for ontology-based data access in biology. BMC Bioinformatics 16:1–9
62. Pesce V, Tennison J, Mey L, Jonquet C, Toulet A, Aubin S, Zervas P, Pesce V, Tennison J, Mey L, Jonquet C, Toulet A, Aubin S, Zervas P (2018) A map of agri-food data standards. Technical Report, F1000Research, 7–177
63. D'Aquin M, Baldassarre C, Gridinoc L, Angeletou S, Sabou M, Motta E (2007) Watson: a gateway for next generation semantic web applications. In: 6th international semantic web conference, ISWC'07, Poster & Demo Session, Busan, Korea, p 3
64. Kitchenham B (2004) Procedures for performing systematic reviews. Technical report, TR/SE-0401, Keele University
65. Gangemi A (2005) Ontology design patterns for semantic web content. In: Gil Y, Motta E, Benjamins VR, Musen MA (eds) 4th international semantic web conference, ISWC 2005. Springer, Galway, pp 262–276

66. de Melo G (2015) Lexvo.org: language-related information for the linguistic linked data cloud. Sem Web 6:8
67. Noy NF, Dorf M, Griffith NB, Nyulas C, Musen MA (2009) Harnessing the power of the community in a library of biomedical ontologies. In: Clark T, Luciano JS, Marshall MS, Prud'hommeaux E, Stephens S (eds) Workshop on semantic web applications in scientific discourse, SWASD'09, Washington DC, USA, p 11
68. Jonquet C, Musen MA, Shah NH (2010) Building a biomedical ontology recommender web service. Biomed Sem 1:S1
69. Nédellec C, Bossy R, Valsamou D, Ranoux M, Golik W, Sourdille P (2014) Information extraction from bibliography for marker-assisted selection in wheat. In: International conference on metadata and semantics research, MTSR'14. Springer, Karlsruhe, pp 301–313
70. Shrestha R, Matteis L, Skofic M, Portugal A, McLaren G, Hyman G, Arnaud E (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 3:326. https://doi.org/10.3389/fphys.2012.00326