New Generation Metadata vocabulary for Ontology Description and Publication

Biswanath Dutta, Anne Toulet, Vincent Emonet, Clement Jonquet

New Generation Metadata vocabulary for Ontology Description and Publication

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

[1]Documentation Research and Training Centre (DRTC)
Indian Statistical Institute, Bangalore, India
`bisu@drtc.isibang.ac.in`
[2]Laboratory of Informatics, Robotics and Microelectronics of Montpellier (LIRMM)
CNRS & University of Montpellier, France
`{anne.toulet,emonet,jonquet}@lirmm.fr`
[3]Center for Biomedical Informatics Research (BMIR)
Stanford University School of Medicine, USA

Abstract. Scientific communities are using an increasing number of ontologies and vocabularies. Currently, the problem lies in the difficulty to find and select them for a specific knowledge engineering task. Thus, there is a real need to precisely describe these ontologies with adapted metadata, but none of the existing metadata vocabularies can completely meet this need if taken independently. In this paper, we present a new version of *Metadata vocabulary for Ontology Description and publication*, referred as MOD 1.2 which succeeds previous work published in 2015. It has been designed by reviewing in total 23 standard existing metadata vocabularies (e.g., Dublin Core, OMV, DCAT, VoID) and selecting relevant properties for describing ontologies. Then, we studied metadata usage analytics within ontologies and ontology repositories. MOD 1.2 proposes in total 88 properties to serve both as (i) a vocabulary to be used by ontology developers to annotate and describe their ontologies, or (ii) an explicit OWL vocabulary to be used by ontology libraries to offer semantic descriptions of ontologies as linked data. The experimental results show that MOD 1.2 supports a new set of queries for ontology libraries. Because MOD is still in early stage, we also pitch the plan for a collaborative design and adoption of future versions within an international working group.

Keywords: Metadata vocabulary, ontology metadata, semantic description, ontology repository, ontology reuse, ontology selection, ontology relation.

1 Introduction

Since a few years, we can observe an increasing number of knowledge artifacts [7][16] or knowledge organization systems [15] that are in use for various semantic applications. Researchers, academicians, practitioners and in general the semantic community across the fields (e.g., computer and information science, medicine, agriculture, economics) are engaged in producing these artifacts (in the rest of the paper these knowledge artifacts are referred with the global term of 'ontologies'). Today, a simple Google Search for "filetype:owl" returns around 34K results. Hence, this is important to describe ontologies with a high degree of accuracy and consistency to find and select them. And for this purpose, we need properly defined metadata. Metadata

will facilitate the manual or automatic search, selection, and elicitation of ontologies. They will allow us to ask various questions on ontologies, for instance, who edited or contributed? When? What methodology or tool was used? Which natural language is used? etc. To describe the ontologies, ontology developers use a variety of metadata vocabularies[1] ranging from general purpose metadata (e.g., DC, DCT, PROV) to dataset specific metadata (e.g., VOID, DCAT, SCHEMA).[2] However, until recently the only ontology specific metadata vocabulary was OMV [4], first published in 2005, which is found to be hardly used by the community. The two main criticisms of OMV are: (i) the current metadata elements primarily allow to capture the provenance information of ontologies, while the other significant aspects, such as development aspect (e.g., curation, evaluation), operational, linguistic, etc. [10] are overlooked; and (ii) it has not reused any other existent relevant metadata vocabularies. Following OMV, in the literature we find a very few studies such as in [3, 9, 8] primarily focused on extending OMV. None of these works address the problems in totality. For instance, the various aspects of ontology descriptions (e.g., ontological relations, community contributions, content-based services) or, the alignment and reuse of other existing metadata vocabularies are not completely exploited. Our earlier work MOD 1.0 (the proposition for Metadata for Ontology Description and publication) in 2015 [2] was a step forward covering some of these limitations. But still there were issues with MOD 1.0 (as discussed in section 2.1). In the current paper, we present our most recent and revised work on MOD (we refer it as MOD 1.2) acquired with the experience of building a brand new metadata model for the AgroPortal ontology repository [13, 5]. 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. In the current work, we also describe the application goals of the vocabulary and illustrate our experimental results with queries that can be run on properly defined metadata.

The main contributions of this work are: (1) the analysis of current ontology metadata practices by looking at the currently existing metadata vocabularies and how they are explicitly used by ontology developers and ontology libraries; (2) the introduction of MOD 1.2, a metadata vocabulary which is a new proposition to the community to harmonize and clarify ontology metadata descriptions, and (3) a use-case describing how to exploit MOD 1.2 ontology with a knowledge base consisting of metadata of eight ontologies originally downloaded from AgroPortal (http://agroportal.lirmm.fr) and new types of queries enabled.

The rest of the paper is organized as follows: Section 2 discusses the current ontology metadata practices; Section 3 discusses MOD 1.2 design methodology and illustrates the MOD OWL model. Section 4 illustrates experimental results. Finally, section 5 concludes and discusses our proposition for community involvement.

---

[1] In this paper, 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 objects used to described ontologies (e.g., OMV, DC, DCAT, etc.).

[2] Please refer to column 'prefix' of Table 2 all along the paper for acronyms definitions of metadata vocabularies.

2 Analysis of current ontology metadata practices

2.1. Analysis of existing metadata vocabularies to describe ontologies

Here, we describe the vocabularies that to some extent have been proposed to capture metadata about ontologies or that could be used with this purpose. Capturing the metadata about 'electronic objects' has been the original motivation of the DCMI [14]. The *Dublin Core* (DC) and *DCMI Metadata Terms* (DCT) are the results of these initiatives. Then we can cite the W3C Recommendations such as *Resource Description Framework Schema* (RDFS), *Web Ontology Language* (OWL) and *Simple Knowledge Organization System* (SKOS). Then the *Ontology Metadata Vocabulary* (OMV) produced in the context of several EU projects and published in 2005 [4]. OMV consists of 16 classes, 33 object properties, and 29 data properties. Unfortunately, the initiative stopped in 2007. One of the limitations of OMV was not to be aligned (or reuse) standard vocabularies at that time. This limitation has been partially addressed in our earlier work on the *Metadata for Ontology Description and publication* (now referred as MOD 1.0) [2]. It has been designed as an ontology consisting of 15 classes (mod:Ontology + 10 others + 4 from FOAF), 18 object properties (7 new ones compare to OMV) and 31 data properties. In MOD 1.0, some of the properties from SKOS, FOAF, and DC and DCT have been used and the vocabulary was not relying on OMV. However, MOD 1.0 still missed out numerous relevant properties as discussed later. 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) [1] 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. More recently, the *Vocabulary of a Friend* (VOAF) [10] was created to "describe vocabularies (RDFS vocabularies or OWL ontologies) used in the Linked Data (LD) Cloud." Although VOAF was developed to capture relations between ontologies, it makes no use or reference to OWL or DOOR (with which it shares a few properties). In 2014, the NKOS working group of the Dublin Core proposed the *NKOS Application Profile* which introduces 6 new properties and reused 22 properties from other vocabularies.

Ontologies share some characteristics with web datasets or data catalogs. 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) [6]. It allows describing two main objects void:Dataset and void:Linkset. *Data Catalog Vocabulary* (DCAT), a most recent W3C recommendation for metadata (and uses DCT), and its profile *Asset Description Metadata Schema* (ADMS), used to describe semantic assets (such as data models, code lists, taxonomies, etc.). Finally, *Schema.org* (SCHEMA) 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). OboInOwl mappings [12] which convert 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 in annotating the ontologies by the ontology developers.

Lessons learned: this review clearly shows that there is a strong overlap in all the vocabularies studied. It shows that no currently existing vocabularies really covers enough aspects of ontologies to be used solely. We also see that despite a few exceptions metadata vocabularies do not rely on one another and redefine things that have already been described several times before (such as dates for which 25 properties are available among the previously listed metadata vocabularies). It is therefore important that a new effort such as MOD 1.2 shall focus on integrating and harmonizing previous ones rather than adding a new one to the list. It is crucial that MOD 1.2 relies on existing metadata vocabularies (preferably official recommendations as we will see later) and also proposes to fusion (and simplify) with the vocabularies that are specific to ontologies (i.e., OMV, MOD 1.0, DOOR, VOAF, VANN) and are not recommendations.

2.2. Analysis of current use of metadata vocabularies for ontology descriptions

To get a sense of the existing metadata vocabularies actually used by ontology developers, we downloaded and manually reviewed 222 OWL ontologies taken randomly from different sources (108 from the NCBO BioPortal (https://bioportal.bioontology.org), 53 from AgroPortal, 61 from searching on Google). We provide here the analysis of the study.

We found 23 ontologies (10%) without any description or annotation. For rest of the 199 ontologies, the number of properties used in describing the ontologies ranging from 1 to 20. For instance, out of the 53 ontologies retrieved from AgroPortal, there are two ontologies having only one metadata for each. There are eight ontologies, for which ten or more properties (and maximum 20) are observed. For rest of the 21 ontologies, the number of metadata per ontology ranges from 2 to 9. The similar trend is observed in ontology descriptions retrieved from BioPortal and the Web. We have also observed in total 32 metadata vocabularies that are being used to describe the ontologies. The 12 most frequently used vocabularies are exemplified in Table 1. Notice that of these most frequently used vocabularies, half of them are W3C or Dublin Core recommended vocabularies. The rest of the other 20 vocabularies form the long tail of the curve with a couple of uses or mostly only once. These include recommended standards (e.g., PROV, SCHEMA), community standards (e.g., VOID, ADMS, DOAP) or very specific vocabularies (e.g., PRISM, EFO, IRON, CITO). The readers may also refer here [11] for a similar study consisted of a corpus of total 23 RDFS/OWL metadata vocabularies which came to comparable conclusions.

Lessons learned: (1) most of all these 32 vocabularies are general in purpose. The metadata vocabularies which were especially proposed with the purpose of annotating/describing ontologies (e.g., OMV, DOOR), are completely absent from the selected sample of our study; (2) two vocabularies among the most used (oboInOwl and protege) are present because they are automatically included in ontologies by ontology development softwares. We can see that rdfs:comment, owl:versionOf and owl:imports are the most frequently used metadata elements. The reason could be 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 which is quite handy; (3) multiple properties capture the same information. For example, in providing the name of the ontology, some have used dc:license, while some other have used cc:license; (4) there is a confusion between the use of DC and DCT as the latter includes and refine the 15 primary properties from the former;[3] (5) generic properties such as rdfs:comment or dc:date are used over more specific ones such as dc:description or dc:created/modified, respectively.

Table 1. Most frequent used vocabularies over a corpus of 222 ontologies.

Prefix	Number	Properties used (number)
dc	294	creator (60), title (51), contributor (34), description (32), rights (20), date (19), subject (15), publisher (14), format (10), identifier (10), license (10), language (9), source (5), coverage (2), issued (1), modified (1), type (1)
rdfs	196	comment (110), seeAlso (23), label (58), isdefinedby (5)
owl	194	versionInfo (105), imports (70), versionIRI (16), priorVersion (3)
oboInOwl	181	hasOboFormatVersion (38), date (35), default-namespace (35), savedBy (31), auto-generated-by (27), namespaceIdRule (3), synonymtypedef (3), hassubset (2), typeref (2), data-version (1), id_space (1), subsetdef (1), treat-xrefs-as-genus-differentia (1), treat-xrefs-as-is_a (1)
dct	105	license (15), modified (15), creator (12), description (12), created (9), issued (8), title (8), subject (6), rights (4), contributor (3), identifier (3), publisher (3), alternative (1), available (1), hasPart (1), hasVersion (1), language (1), lastModified (1), type (1)
skos	27	definition (8), altLabel (6), prefLabel (6), editorialNote (4), historyNote (2), changeNote (1)
vann	21	preferredNamespacePrefix (11), preferredNamespaceUri (10)
cc	12	license (12)
protege	11	defaultLanguage (11)
dcat	9	landingpage (5), downloadURL (2), contactPoint (1), mediaType (1)
foaf	5	primaryTopic (2), homepage (1), maker (1), page (1)
pav	5	version (5)
void	5	subset (2), dataBrowse (1), dataDump (1), sparqlEndpoint (1)

2.3. Analysis of metadata representation within ontology repositories

We have studied some of the most common ontology libraries and repositories available in the semantic web community to understand: (i) how they are dealing with ontology metadata; and (ii) to which extent they rely on previously analyzed metadata vocabularies. We have explicitly reviewed: repository or portals including the NCBO BioPortal, Ontobee (www.ontobee.org), EBI Ontology Lookup Service (www.ebi.ac.uk/ols), MMI Ontology Registry and Repository (https://marinemetadata.org/orr), the ESIP Portal (http://semanticportal.esipfed.org), AberOWL (http://aber-owl.net/). Registries or catalogs including the OKFN Linked Open Vocabularies (http://lov.okfn.org), OBO Foundry (www.obofoundry.org), WebProtégé (http://webprotege.stanford.edu), VEST/AgroPortal Map of Standards (http://vest.agrisemantics.org), and BioSharing (https://biosharing.org).

---

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

Lesson learned: We have reviewed the metadata properties captured by all these libraries. The NCBO BioPortal which uses 66 metadata properties and partially reuse OMV served as reference as it was also our baseline when implementing a new metadata model with AgroPortal [13, 5], before MOD 1.2. We observe that each of the reviewed libraries uses, to some extent, some metadata elements but do not always use standard metadata vocabularies. For a recent review of ontology libraries and their metadata, the readers might also refer to [2] where we showed that ontology metadata vocabularies are rarely used by ontology libraries: 4 ontology libraries over the 13 studied have partially used the OMV.

3 Presentation of MOD 1.2

3.1. Design methodology

From our previous reviews and analysis, we have come up with a list of 88 properties forming MOD 1.2 vocabulary (whereas MOD 1.0 offered 25) that would capture the information about an ontology. The criteria for inclusion were the following, consider by order of importance:

1. Relevance for describing an ontology – the property may have a sense if used to describe an ontology. For this purpose, we prepared a list of queries (aka competency questions) considering the varieties of use scenarios (or, tasks), for instance, an application developer searching for an ontology to use in an application he is developing, a user making a survey to find the existence of ontologies in his domain of interest, and an ontology developer searching for an ontology that he can refer as a gold standard to evaluate his ontology.
2. 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 of type omv:Ontology, foaf:Document, owl:Ontology, prov:Entity.
3. Being included in a W3C or Dublin Core recommendation.
4. The frequency of use as found in the study presented in Section 2.2.
5. Priority to vocabularies specific for ontologies rather than to the ones specialized for the more general objects (cc:Work, dcat:DataSet, sd:Service, etc.).

From each of these vocabularies, we have selected the significant properties to describe objects where 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. From RDFS and OWL, we have reviewed properties that can be used to describe rdfs:Resource, and owl:Ontology. From RDFS, we selected only one property rdfs:comment, whereas we considered the properties rdfs:seeAlso and rdfs:label are better represented by other properties. For instance, rdfs:label is better represented by dct:title. From OWL, we selected all the considered properties. From Dublin Core (assuming the domain of its properties is rdfs:Resource), we have selected 28 properties giving priority to DCT. From OMV, we have considered all the 37 properties for omv:Ontology but selected in MOD 1.2 only 20 as others can be represented by other properties matching more of our criteria. In a similar way, we have selected the number of properties from the other vocabularies as indicated in Table 2. Out of the total considered 244 properties (column #C) from 23 vocabularies, MOD 1.2 consisted of

88 properties (column #S) from 11 vocabularies, which includes 13 properties defined (Table 3) in mod namespace.

Table 2. Vocabularies studied and used within MOD 1.2. R column states if it is a W3C or DC recommendation (R) or note (N). Colon (#S) is the number of property selected in MOD1.2 from this vocabulary. Colon (#C) is the number of property considered that are either selected or considered equivalent of another selected one.

Prefix	Namespace	Name	R	#S	#C
adms	http://www.w3.org/ns/adms#	Asset Description Metadata Schema	N	0	9
cc	http://creativecommons.org/ns#	Creative Commons Rights Expression Language		0	3
dc	http://purl.org/dc/elements/1.1/	Dublin Core	R	0	4
dcat	http://www.w3.org/ns/dcat#	Data Catalog Vocabulary	R	11	15
dct	http://purl.org/dc/terms/	DCMI Metadata Terms	R	18	34
doap	http://usefulinc.com/ns/doap#	Description of a Project		3	11
door	http://kannel.open.ac.uk/ontology#	Descriptive Ontology of Ontology Relations		0	6
foaf	http://xmlns.com/foaf/0.1/	Friend of a Friend Vocabulary	N	5	10
idot	http://identifiers.org/idot/	Indentifiers.org		0	4
mod	http://www.isibang.ac.in/ns/mod#	Metadata for Ontology Description & Publication 1.0		13	25
nkos	http://w3id.org/nkos#	Networked KOS Application Profile		0	4
oboInOwl	http://www.geneontology.org/formats/oboInOwl#	OboInOwl Mappings		0	4
omv	http://omv.ontoware.org/2005/05/ontology#	Ontology Metadata Vocabulary		20	37
owl	http://www.w3.org/2002/07/owl#	OWL 2 Web Ontology Language	R	7	7
pav	http://purl.org/pav/	Provenance, Authoring and Versioning		2	10
prov	http://www.w3.org/ns/prov#	Provenance Ontology	R	3	9
rdfs	http://www.w3.org/2000/01/rdf-schema#	RDF Schema	R	1	3
schema	http://schema.org/	Schema.org		0	31
sd	http://www.w3.org/ns/sparql-service-description#	SPARQL 1.1 Service Description	R	1	1
skos	http://www.w3.org/2004/02/skos/core#	Simple Knowledge Organization System	R	0	1
vann	http://purl.org/vocab/vann/	Vocabulary for annotating vocabulary descriptions		0	4
voaf	http://purl.org/vocommons/voaf#	Vocabulary of a Friend		4	5
void	http://rdfs.org/ns/void#	Vocabulary of Interlinked Datasets	N	0	7
TOTAL		**23 vocabularies, 12 used**	**12**	**88**	**244**

3.2. MOD 1.2

MOD 1.2 is defined in OWL with the namespace http://www.isibang.ac.in/ns/mod#. Fig. 1 provides a representation of the model in terms of its main classes, object & data properties, including the constraints on its primary class mod:Ontology. The OWL file and versions are publicly available (https://github.com/sifrproject/MOD-Ontology).

Table 3. MOD1.2 new created metadata properties (naae=not available anywhere else, ifbp=inspired from BioPortal, povnyi=present in other vocabularies not yet integrated in MOD).

Properties	Definition	Reason for creating
mod:competencyQuestion	A set of questions asked at design time to explain why the ontology is needed and explain its design.	naae
mod:group	A group of ontologies that the ontology is usually considered into.	naae, ifbp
mod:translation	A pointer to the translated ontology(ies) for an existing ontology.	povnyi
mod:rootClasses	The root class(es) of an ontology. This could be automatically populated by taking the direct subclasses of owl:Thing. If the ontology is also defined as a unique skos:ConceptScheme, then this property becomes the equivalent of skos:hasTopConcept	naae
mod:browsingUI	The user interface (URL) where the ontology may be browsed or searched.	naae, ifbp
mod:vocabularyUsed	The vocabularies that are used and/or referred to create the current ontology.	povnyi
mod:sampleQueries	A set of queries (may be SPARQL, DL Queries) that are provided along with an ontology to illustrate use cases.	naae
mod:ontologyInUse	An ontology that is used in a project.	naae, ifbp
mod:evaluation	An ontology that has been evaluated by an agent.	ifbp, povnyi
mod:numberOfObjectProperties	The total number of object properties in an ontology. Refines omv:numberOfProperties.	naae
mod:numberOfDataProperties	The total number of data properties in an ontology. Refines omv:numberOfProperties.	naae
mod:numberOfLabels	Number of defined labels for any resources in an ontology (classes, properties, etc).	naae
mod:byteSize	The byte size of an ontology file.	naae

Classes: MOD 1.2 consists of 19 classes (where a class is a collection of things sharing common attributes) including three subclasses. The classes are derived by analyzing the selected properties and identifying the reusable ontological resources. Some of the exemplary classes (and their reusable resources) are mod:Ontology (e.g., Gene Ontology, Disease Ontology), mod:Group (e.g., OBO Foundry or OBO Library), foaf:Project (e.g., Planteome, AgroLD).

Object Property: MOD 1.2 consists of 28 object properties (which connects two resources belonging to two different, or same classes), for instances, mod:evaluatedBy, omv:endoresedBy, mod:ontologyInUse. Each object properties are defined with its

domain and range e.g., the mod:ontologyInUse has a domain class mod:Ontology and a range class foaf:Project. An object property can have more than one domain and range.

Data Property: MOD 1.2 consists of 60 data properties (which connects a resource to a literal) to describe ontologies, for instances, omv:URI, mod:competencyQuestion, mod:sampleQueries. Besides, there are 9 other properties (e.g., foaf:name) included to facilitate the description of the related resources such as, foaf:Agent. Each data properties is specified with its domain and range. For instance, property mod:competencyQuestion has a domain class mod:Ontology and a range xsd:string. A data property can have more than one domain.
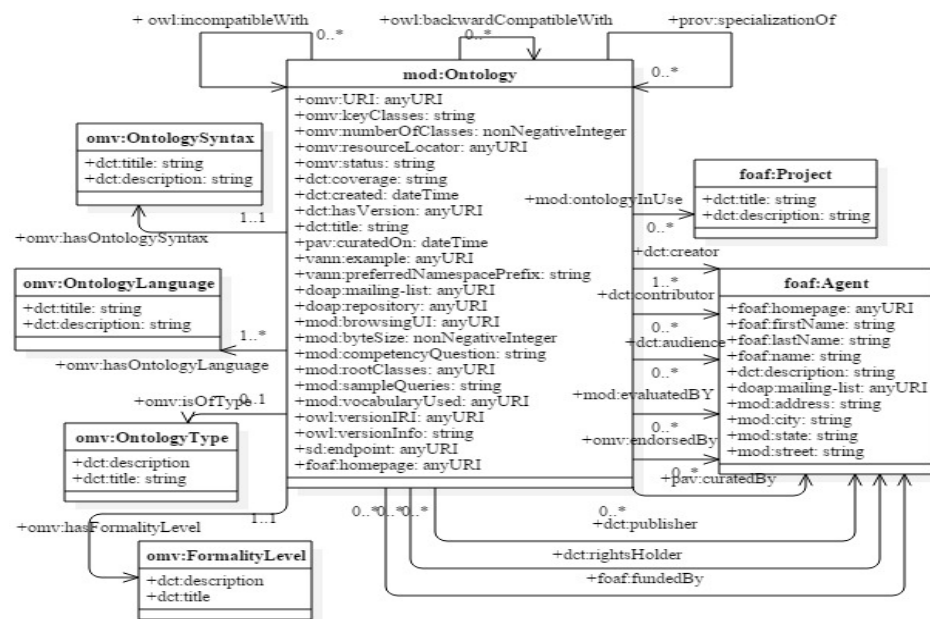
Fig. 1. A snapshot of MOD 1.2 (a complete diagrammatic representation is available here https://github.com/sifrproject/MOD-Ontology).

4 Illustration of experimental results

Using the above MOD OWL model, we have created a knowledge base consisting of metadata about eight agronomical ontologies selected from AgroPortal and defined as instance of omv:Ontology These are AGROVOC, Gene Ontology, National Agricultural Library Thesaurus, NCBI Organismal Classification, Protein ontology, AnaEE Thesaurus, IBP Crop Research Ontology, and Sequence Types and Features Ontology. These ontologies were chosen from AgroPortal because of the new metadata model recently developed within this ontology repository [13]. Indeed, the AgroPortal's team has spent a significant amount of time to edit the metadata of the ontologies with the goal to facilitate the comprehension of the agronomical ontology

landscape [5].[4] Therefore, ontologies and vocabularies within AgroPortal are very precisely described still with another set of properties not yet aligned with MOD.

The knowledge base has been created manually using Protégé (https://protege.stanford.edu). Most of the metadata for the selected ontologies originally came from AgroPortal. In some cases, we have also consulted with the original source of those ontologies and other information online. In the knowledge base, we decided to reuse, wherever available, the existing URIs of the resources instead of creating them in mod namespace. For instances, the OBO foundry ontologies offer persistent URLs for each of their ontologies e.g., Gene Ontology. Also, for creating the organizational resources, we have preferred to use DBPedia defined URIs. In the case of unavailability, we have used the organizational homepage URL as the resource URI. Similarly, in the case of people, we have preferred to use the ORCID IDs as URIs. In the case of unavailability, or any kind of ambiguity, we have created the resources in mod namespace. For language, we have used Lexvo vocabulary (www.lexvo.org). The same approach is followed for creating the other related resources, for examples, licensing (https://creativecommons.org), vocabulary formality level (OMV), ontology types (http://w3id.org/nkos/nkostype), projects, and so forth. The current knowledge base consists of in total 1962 axioms, 20 classes, 33 objects and 69 data properties, and 217 individuals. The knowledge base is available and can be downloaded from https://github.com/sifrproject/MOD-Ontolog with the main MOD 1.2 file.

The knowledge base supports the varieties of new queries, for instance, which is the most popular ontology editing tool? Who are the key contributors in a domain? How many ontologies are produced by OBO Foundry group? What are the projects using the Protein Ontology? *What are the ontologies endorsed by the RDA Wheat Data Interoperability Group (RDA WDI) and the National Science Foundation (NSF)*? These queries were expressed in SPARQL and successfully run over the knowledge base. The above *italicized* query is shown below. It returns the title and the creator of the ontologies endorsed by RDA WDI and NSF. A couple of such sample SPARQL queries are also available on GitHub.

```
SELECT DISTINCT ?Ontology ?Author
   WHERE {
       {?x a mod:Ontology; omv:endorsedBy <https://www.rd-alliance.org/groups/wheat-data-interoperability-wg.html> ; dct:title ?Ontology .}
   UNION
       {?x a mod:Ontology; omv:endorsedBy <http://dbpedia.org/resource/Category:National_Science_Foundation> ; dct:title ?Ontology .}
     OPTIONAL {?x dct:creator ?Author .} }
```

Furthermore, our future goals are: (i) to automatize the process of creating mod:Ontology instances using the application programming interfaces of the main ontology libraries (e.g., BioPortal, AgroPortal, OBO Foundry). This will enable to export the content of these libraries without doing any change to their internal data models; (ii) to release knowledge base as Linked Open Data consisting of metadata for

---

[4] AgroPortal has now a specific page (http://agroportal.lirmm.fr/landscape) dedicated to visualizing this landscape. It displays highly valuable synthetized information with diagrams and charts about the ontologies in agriculture. This was made possible by the new metadata model.

ontologies covering a significant amount of ontologies; and (iii) to offer a SPARQL endpoint to provide local and remote advanced queries on the knowledge base.

5 Conclusion and proposition for community involvement

From our study and analysis (Section 2), we have seen that so far, the only ontology metadata vocabulary OMV (until the recent publication of MOD 1.0 in 2015) published in 2005 could make a very little impact on the community at least in terms of its use. According to us, among the main limitations of OMV that might explain why it is not really adopted today are: (i) it did not reuse any other existent relevant metadata vocabularies; (ii) it was never included in a common ontology editor like Protégé. It would have highly facilitated the adoption of the vocabulary if ontology editors would have had only to fill out a few forms directly in their preferred ontology edition software; (iii) 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; (iv) after 2009, there was no update and the development team has become less active.

MOD 1.2 is an initiative, a joint effort from ISI and LIRMM, which attempts to overcome some of the limitations of OMV and overall proposes a solution which shows the promises of satisfying the community needs for describing the ontologies from multiple aspects (e.g., provenance, developmental, linguistic, community). However, MOD 1.2 is 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 standardization body such as W3C. One of our objectives is to introduce MOD 1.2 to the Research Data Alliance recently re-configured Vocabulary and Semantic Services Interest Group (VSSIG - https://www.rd-alliance.org/groups/vocabulary-services-interest-group.html). This group is a follow-up of the EUDAT Semantic Working Group Workshop from April 2017, where multiple interests on standardizing ontology metadata have been publicly expressed by the major ontology repository developers.

MOD is an open project described on GitHub and ResearchGate, so that the community can view, and participate in the discussions. We envision that the MOD 1.2, currently consisted of 88 properties, in the near future will turn to a collaborative extended version (MOD 2.0). One of our short-term objectives is also to propose an "application profile" for the description of ontologies that will be based on MOD and will serve ontology developers more easily than by creating mod:Ontology instances. Section 2.2 analysis has shown that it is the way adopted by ontology developers.

References

[1] C. Allocca, M. d'Aquin, and E. Motta. Towards a Formalization of Ontology Relations in the Context of Ontology Repositories. In *Knowledge Discovery, Knowlege Engineering and Knowledge Management*, vol 128 of *Communications in Computer and Information Science*, pages 164–176. Springer, 2011.

[2] B. Dutta, D. Nandini, and G. Kishore. MOD: Metadata for Ontology Description and Publication. In *International Conference on Dublin Core & Metadata Applications, DC'15*, pages 1–9, Sao Paulo, Brazil, September 2015.

[3] M. Fiorelli, A. Stellato, J. P. McCrae, P. Cimiano, and M. T. Pazienza. LIME: The Metadata Module for OntoLex. In *12th European Semantic Web Conference, ESWC'15*, volume 9088 of *LNCS*, pages 321–336, Portoroz, Slovenia, May 2015. Springer.

[4] J. Hartmann, Y. Sure, P. Haase, and M. Suarez-Figueroa. OMV: ontology metadata vocabulary. In C. Welty, editor, *Workshop on Ontology Patterns for the Semantic Web, WOP'05*, page 9, Galway, Irland, November 2005. Springer.

[5] C. Jonquet, A. Toulet, and V. Emonet. Two years after: a review of vocabularies and ontologies in AgroPortal. In *International Workshop on sources and data integration in agriculture, food and environment using ontologies, IN-OVIVE'17*, page 13, Montpellier, France, July 2017. EFITA.

[6] M. H. Keith Alexander. 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, April 2009.

[7] D. L. McGuinness. *Spinning the semantic web: bringing the World Wide Web to its full potential*, chapter Ontologies come of age, pages 171–194. MIT Press, 2003.

[8] H. Min, S. Turner, S. de Coronado, B. Davis, P. L. Whetzel, R. R. Freimuth, H. R. Solbrig, R. Kiefer, M. Riben, G. A. Stafford, L. Wright, and R. Ohira. Towards a Standard Ontology Metadata Model. In *7th International Conference on Biomedical Ontologies, ICBO'16, Poster Session*, volume 1747 of *CEUR*, page 6, Corvallis, Oregon, USA, August 2016.

[9] E. Montiel-Ponsoda, G. A. de Cea, M. C. Suarez-Figueroa, R. Palma, A. Gomez-Pérez, and W. Peters. LexOMV: an OMV extension to capture multilinguality. In *Lexicon/Ontology Interface Workshop, OntoLex'07*, page 10, Busan, South-Korea, November 2007.

[10] L. Obrst, M. Gruninger, K. Baclawski, M. Bennett, D. Brickley, G. Berg-Cross, P. Hitzler, K. Janowicz, C. Kapp, O. Kutz, C. Lange, A. Levenchuk, F. Quattri, A. Rector, T. Schneider, S. Spero, A. Thessen, M. Vegetti, A. Vizedom, A. Westerinen, M. West, and P. Yim. Semantic Web and Big Data Meets Applied Ontology: The Ontology Summit 2014. *Applied Ontology*, 9(2):155–170, April 2014.

[11] C. Tejo-Alonso, D. Berrueta, L. Polo, and S. Fernandez. Current practices and perspectives for metadata on web ontologies and rules. *Metadata, Semantics and Ontologies*, 7(2):10, 2012.

[12] S. H. Tirmizi, S. Aitken, D. A. Moreira, C. Mungall, J. Sequeda, N. H. Shah, and D. P. Miranker. Mapping between the OBO and OWL ontology languages. *Biomedical Semantics*, 2(S1/S3):16, March 2011.

[13] A. Toulet, V. Emonet, and C. Jonquet. Modèle de métadonnées dans un portail d'ontologies. In *6èmes Journées Francophones sur les Ontologies, JFO'16*, Bordeaux, France, October 2016. Best paper award.

[14] S. Weibel, J. Kunze, C. Lagoze, and M. Wolf. Dublin core metadata for resource discovery. Technical Report RFC 2413, Internet Engineering Task Force, September 1998.

[15] M. L. Zeng. Knowledge Organization Systems (KOS). *Knowledge organization*, 35(2-3):160–182, 2008.

[16] Dutta, B.: Examining the interrelatedness between ontologies and linked data. *Library Hi Tech*, 35(2), 312-321, 2017.