Assessing the Practice of Ontology Metadata: A Survey Result

Anne Toulet, Biswanath Dutta, Clement Jonquet

Assessing the Practice of Ontology Metadata: A Survey Result.

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

[1] LIRMM, University of Montpellier, France
[2] DRTC, Indian Statistical Institute, Bangalore, India
anne.toulet@lirmm.fr
bisu@drtc.isibang.ac.in
clement.jonquet@lirmm.fr

Abstract. The data deluge – or Big Data – brings us to think differently about data management and shows us the urgent need to move towards FAIR data. Semantic resources (ontologies, vocabularies, thesauri and terminologies) are no exception to this rule. Between February 16 and March 30, 2018, we conducted a survey on the use of metadata vocabularies to describe such resources. We wanted to evaluate the current state of practice and discuss recommendations in terms of metadata standards for ontologies. In this paper, we present and discuss the results of this survey. The analyze shows that the core semantic web languages (RDFS, OWL, SKOS) and vocabularies such as DCAT or Dublin Core are among the most known and used vocabularies to describe ontologies. More surprisingly, most of the numerous vocabularies really relevant for describing ontologies are barely known and never used: DOOR, VANN, ADMS, OMV, MOD. This demonstrates a lack of a clearly identified standard recommendation for this purpose. We then propose perspectives to meet the need for harmonization of metadata vocabularies, placing this survey in the context of the work already done on this question.

Keywords: Ontologies, Thesaurus, Vocabularies, Metadata, Semantic description, Ontology selection, Survey research, Questionnaire.

1 Introduction

In the context of the recent reconfiguration of the RDA Vocabulary and Semantic Services Interest Group (VSSIG)[3], we are leading a task group focused on ontology metadata. The goal of this task group is to review the current state related to metadata description of semantic resources in general (ontologies, vocabularies, thesauri and terminologies); and discuss recommendations and best practices in terms of metadata standards for this type of resource. This paper is based on a survey conducted between February 16 and March 30, 2018. Ontologies are some kind of knowledge artifacts [13] or knowledge organization systems [19]. Hereafter, we often use the word "ontologies" or "ontologies and vocabularies". By

[3] RDA VSSIG: https://www.rd-alliance.org/groups/vocabulary-services-interest-group.html

ontology, we mean not only an OWL structure that respects all the conditions to qualify as a fully formalized ontology: we include every ontology-like structure which formalizes some knowledge (ontology, vocabulary, thesaurus, taxonomy, terminology, dictionary). The point is not to focus on the level of semantics of the ontology[4], but on its metadata description. 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 [18] or the Ontology Metadata Vocabulary working group [5]. The Open Ontology Repository Initiative [1] 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 [12], 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 objective: we believe most information that can be expressed in a scientific article presenting an ontology âĂŞincluding narrative sections such as motivation, knowledge acquisition or change managementâĂŞ can also be captured as appropriate metadata in the ontology itself.

By metadata, we mean any property used to describe the ontology itself or relations between the described ontology and other resources. To avoid confusion with ontologies, we here call metadata vocabularies the resources which offer a list of metadata properties that can be used to describe ontologies (e.g. Dublin Core, VoID, OMV, DCAT, MOD)[5]. The metadata authoring is the process of choosing and editing a metadata property when describing an ontology. Although most ontology producers are still reluctant to use metadata to describe their resources [4], the survey shows that most users are aware of their interest, especially in finding and selecting them. Over 94% of the responses indicated 'yes' to the following question: "Overall, do you think authoring and accessing ontology metadata is important?". However, a recurring problem pointed out by users is the lack of standard or single reference to describe ontologies. How to improve the situation and meet the need of the community? The paper is organized as follow: Section 2 explains the context in which this survey takes place, describing previous and on-going works about ontology metadata. Section 3 presents the survey methodology and analysis. Section 4 is devoted to the presentation and discussion of the results. Finally, Section 5 concludes the paper and lists a few perspectives for the ontology metadata work.

[4] We acknowledge the differences (not discussed here) in all these types of Knowledge Organization Systems (KOS) or knowledge artifacts. The reader may refer to McGuinness's discussion [13].

[5] Please refer to Fig.1 all along the paper for acronym definition of metadata vocabularies.

2 Background

The survey takes place in the perspective of different works we engaged around the question of metadata for ontologies. In this section, we present the most significant researches we have already done in this domain.

2.1 Unified Metadata Model of AgroPortal

The AgroPortal project [7] aims at offering a reference ontology repository for the agronomy and related domains such as food, plant, and biodiversity sciences. The first prototype of the AgroPortal (`http://agroportal.lirmm.fr/`) was introduced in 2015, reusing the NCBO BioPortal technology [9]. The role of metadata in such a repository is essential to facilitate the ontology identification and selection process, which has been assessed as crucial to enable ontology reuse [16]. The original metadata model in AgroPortal, inherited from the NCBO BioPortal, is based on the Ontology Metadata Vocabulary (OMV) [6]. However, our goal was to capture as much information as possible about the ontologies available in AgroPortal, and we decided to go further: we implemented a new metadata model to support better descriptions of ontologies and their relations with respect of the standard metadata vocabularies used in the semantic web community. For this purpose, we reviewed the most standard and relevant vocabularies out-there to describe metadata for ontologies: 23 in total including Dublin Core, VoID, Ontology Metadata Vocabulary, Data Catalog Vocabulary, as shown in Figure 1.

Prefix	Namespace
adms	http://www.w3.org/ns/adms#
cc	http://creativecommons.org/ns#
dc	http://purl.org/dc/elements/1.1/
dcat	http://www.w3.org/ns/dcat#
dct	http://purl.org/dc/terms/
doap	http://usefulinc.com/ns/doap#
door	http://kannel.open.ac.uk/ontology#
foaf	http://xmlns.com/foaf/0.1/
idot	http://identifiers.org/idot/
mod	http://www.isibang.ac.in/ns/mod#
nkos	http://w3id.org/nkos#
oboInOwl	http://www.geneontology.org/formats/oboInOwl#
omv	http://omv.ontoware.org/2005/05/ontology#
owl	http://www.w3.org/2002/07/owl#
pav	http://purl.org/pav/
prov	http://www.w3.org/ns/prov#
rdfs	http://www.w3.org/2000/01/rdf-schema#
schema	http://schema.org/
sd	http://www.w3.org/ns/sparql-service-description#
skos	http://www.w3.org/2004/02/skos/core#
vann	http://purl.org/vocab/vann/
voaf	http://purl.org/vocommons/voaf#
void	http://rdfs.org/ns/void#

Fig. 1. Vocabularies studied: Acronyms and Namespaces.

We then grouped those properties into a unified and simplified model of 124 properties that includes the 45 properties originally offered by the NCBO BioPortal and describe all the new properties with standard vocabularies [17]. AgroPortal now recognizes 346 properties from existing metadata vocabularies that could be used to describe different aspects of ontologies: intrinsic descriptions, people, date, relations to other resources, 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[6]. Most of metadata are automatically extracted from the original ontology file if present and sometime automatically generated by the portal; and a part of them are edited and manually curated by the AgroPortal team. This has resulted in our capability to automatically aggregate information about ontologies & 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 & organization, most active ontologies, etc.). We have now a specific page dedicated to visualizing this landscape in AgroPortal: `http://agroportal.lirmm.fr/landscape`. This work 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 pre-existing ones [8]. A generalization of this work is now launch as part of a standardization effort led, on the one hand through the development of a metadata vocabulary specifically dedicated to the description of ontologies, the Metadata for Ontology Description and publication (MOD), described in Section 2.2, and, on the other hand, in the context of the RDA Vocabulary and Semantic Services Interest Group, presented in Section 2.3.

2.2 Metadata for Ontology Description and publication

This work on ontology metadata first started with the proposition of a metadata vocabulary namely Metadata for Ontology Description and publication (MOD) [3]. The primary aim of MOD was to enable the ontology developers to describe their ontologies in a way so that the ontologies are easily identifiable and reusable for various knowledge engineering tasks. Prior to MOD, there was another vocabulary namely Ontology Metadata Vocabulary (OMV), which was limited in its scope [10, 15] and found to be less used. Our studies have shown that out of the 13 ontology repositories, only four of them were using OMV partially and the rest of the other repositories (about 70%) defined their own metadata elements to describe the ontologies [3, 14]. It was also observed that besides OMV, the community, especially the ontology developers, were using the other general purpose metadata vocabularies, not exclusively designed for ontology description, to describe the ontologies, for instance, Dublin Core, SKOS, RDF Schema, OWL, VOID, DOOR, etc. [4]. All these studies and observations

[6] More details here: `https://github.com/agroportal/documentation/wiki/Ontology-metadata`

motivated the work on MOD. We realized the need for ontology metadata vocabulary for the better description and retrieval of ontology and also to support the interoperability between the ontology repositories. The first version of MOD (1.0) published in 2015 consisting of 15 classes, 18 object properties, and 31 data properties. Later in 2017, a revised version of MOD (V.1.2) was published consisting of 19 classes, 28 object properties, and 60 data properties. The revision was carried out by extending MOD1.0 and also by reusing and realigning MOD with the other relevant metadata vocabularies. Currently, we are working on MOD 1.3. It is to be noted here that MOD is structured as an ontology metadata model, and hence it can be directly used in developing the ontological knowledge base for ontologies besides its use for annotating and documenting the ontologies. A use case of MOD ontology model and its usefulness can be found here [4].

2.3 RDA Vocabulary and Semantic Services Interest Group

The RDA Vocabulary Service Interest Group (VSIG) was created in 2015 and became the Vocabulary and Semantic Services Interest Group (VSSIG)[7] in June 2017. This group seeks to develop community-based approaches and recommendations to make knowledge organization systems (i.e. controlled vocabularies, ontologies, and their associated services) findable, accessible, interoperable, and re-usable (FAIR data principles)[8]. The VSSIG may develop recommendations to address the needs of research communities and software developers for discovering and using multi-disciplinary controlled vocabularies and ontologies published on the web. The key challenge faced by this interest group is to break the barriers between the different community specific vocabularies and improve their findability. The interest group will address this challenge along three main axes: the discoverability of the existing resources; the interoperability of the existing resources at metadata level; and the interoperability at the API level. In this context, 11 task groups were created, including the one working on "Ontology metadata standard" and in which we are involved as leaders. The work of this task group consists in developing together a new ontology metadata standard that can be used to describe the ontologies, vocabularies, terminologies and the similar kinf of resources.

3 Survey methodology

In continuation of the work presented above, we wanted to deepen our knowledge on the real use of metadata in the field of semantic resources. In this perspective, we conducted a survey using an online questionnaire (Google Forms) between February 16 and March 30, 2018. The target audience is the community of ontology users and developers. The survey announcement circulated on about sixty

[7] https://www.rd-alliance.org/group/vocabulary-services-interest-group/wiki/updated-charter-proposed

[8] FAIR Guiding Principles: http://dx.doi.org/10.1038/sdata.2016.18

mailing lists concerned by this subject: W3C lists, RDA groups, ontology portals community (BioPortal, AgroPortal, SIFR BioPortal, OBO Foundry) and researcher mailing lists; as well as on some social networks such as Twitter or Research Gate. The goal of this survey is to understand how the ontology developer community authors metadata to describe their ontologies and how ontology users use or appreciate these metadata. To achieve this objective, the survey is based on a few key questions:

- Do ontology developers actually describe their ontology with metadata and if so, which ones?
- Do ontology users rely on metadata in their use of ontologies?
- What are the way to improve the current situation and make ontologies more FAIR: Findable, Accessible, Interoperable, Reusable ?

The survey was organized in five parts. Part 1 introduces the survey providing the basic information such as the objective of the survey, and also clarifies the meaning of some of the basic terms used in the survey (e.g., ontology, ontology metadata, metadata vocabulary, metadata authoring) to avoid misunderstandings. The meaning of these terms is the one we gave beforehand in this paper. Part 2 focuses on the profile of the respondent: what kind of job he does, what kind of ontology user he is, the level of experience and expertise related to ontologies and main reasons for his interest in ontologies. In part 3-4, we provide a set of questions to find the answer for the above stated key questions. Part 3 aims to explore the community experience related to ontology metadata. It asks questions focusing to the role and usage of metadata in the search and selection of ontologies, practice of the developers in describing their ontology, and the known and used metadata vocabularies. Part 4 seeks to identify the needs of the community and what would enhance the usage of metadata. The goal here is to open up prospects for work already undertaken in this area to meet these needs. Finally, the last facultative part 5 is to know better the respondents. For each question, responses have been studied one by one, using statistics and semantic analysis. For some of them, graphics (bar charts, pie charts) and word clouds have been produced.

4 Results and discussion

We had 168 participants: among them, 91 gave their name, 87 indicated their organization and/or gave a contact email address.

4.1 Profile of participants

It is found that majority of the responders (43%) were the researchers following which around 14% were the professors and 12% software developers. Besides these, there were people from various other domains such as taxonomist, librarian, data manager, linked data architect, metadata specialist, and so forth. The majority of the participants (66%) are both ontology user and developer. This

reflects the fact that the responses we received carry both types of experiences of a participant, one side his experience as an ontology search user, the metadata elements required for finding an ontology, and on the other side his experience and expectations as a developer in describing an ontology to make it FAIR. Note that besides the mixed type of ontology users, there were participants either solely the ontology developer (23%) or the ontology user (11%). Because of the topic and presentation of the survey, most of the respondents were experienced with ontology (to the point of developing some). Indeed, one concretely asks himself about ontology metadata only when facing the problem of describing his own ontology. Among the 168 participants, 52% were the advanced users and 27% were the expert users. Besides these, there were also 20% basic users. The profile of the participants is summarized in the figure 2.

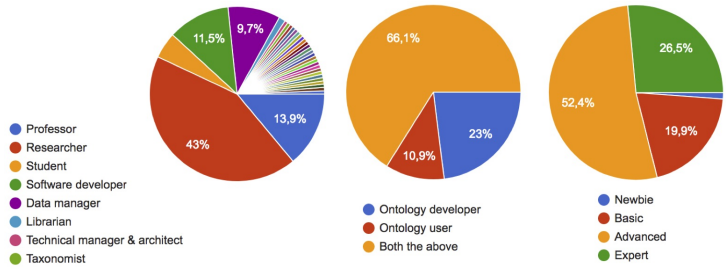

Fig. 2. Profile of participants.

Unsurprisingly, the main reasons for the interest in ontologies reveal the classic and well known use cases for ontologies including for: data management and integration, semantic annotation, knowledge capture, indexing/representing data, etc. Most of the responders put themselves in the position of an ontology user when answering this question. The diversified roles and the background of the participants give us the hints of the coverage and the significance of the content of the responses.

4.2 Experience with ontology metadata

The findings and analysis exploring the experience and practices of the community related to ontology metadata are presented at three different levels as follows.

Criteria for searching and selecting ontologies The big data deluge and the adoption of the semantic 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 [11]. 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 [2] even if it is now identified as a requirement to make the data FAIR. We asked participants how they search and select ontologies to understand the user's ontology search behavior. We found most of the participants (63%) find the ontologies from the literature or through their community; 53% participants responded that they know the ontologies they want to use, and 51% discover the ontologies from ontology libraries or repositories; of course, some of them also search on the Web search engines like Google. Only 11% responded they use any sort of ontology recommender service. From the responses, we can say that the peers play a significant role in disseminating information about the existence of an ontology. This shows in part the lack of a reliable way for users to find recommendations for choosing the resources they need; and their community is possibly the most reliable and efficient way to identify ontologies in terms of time and quality. We then asked people what were the top 5 things they would like to know when searching and selecting an ontology. One of our goals with this question was to identify the information most wanted by users, especially to verify the existence of corresponding metadata properties to express them. The main information the people seek about an ontology is (in order of importance):

- the known usage (including examples and datasets using this ontology) and the ontology task i.e. the purpose for which the ontology was originally designed. This has been pointed by more than one in four people;
- the formality level and the expressivity;
- the maintenance: how often it is updated, is it still used or deprecated, etc.
- the domain, scope and coverage;
- the relations: alignment, mappings, imports, reuse, etc.
- the quality, popularity and endorsement;
- information about community users and community support;
- content : concepts, properties, definitions, labels, languages;
- metrics: number of classes, number of concepts, number of properties.

To a lesser extent, people would also like information on the repository (code source location) and the issue tracker (for example GitHub issues), the documentation and the publications about the ontology, the license and copyright, the standards used. Our experience with metadata vocabularies [8] allows us to confirm that for each of this information, there is a property to make it explicit. This means that users do not easily find the information that would help them to select their ontologies, while the properties to express it exist. What about ontology developers who are best placed to describe their resources?

Ontology developer practices to describe their resources Our different works on ontology metadata have found a lack of description for ontologies [4,8], confirmed by the previous analyze. To better understand the reason, we were interested in how developers describe their ontologies: how they author ontology

metadata, how they express information especially the entities like person, organization, location and how they choose the ontology metadata vocabularies. The first question indicates that although 11% of respondents report not using metadata to describe their ontology, most of them author metadata by adding annotation properties and without importing the metadata vocabulary (Fig. 3). The additional comments for this question make us believe that the relatively important number of answers (27%) for "I formally use metadata vocabularies by importing them within my ontology." seems to come from a miss understanding of the meaning of import (which for us was owl:imports).

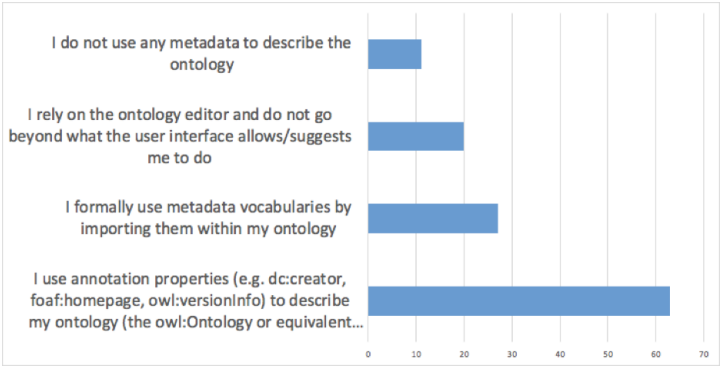

Fig. 3. Illustrates the ontology developers ontology metadata authoring.

To describe classical entities like person, organization or location, 48% of the responders agree on the importance of reusing existing URIs whenever possible. However, when ontology developers create a new value for a property, they equally either create new URIs, or use some kind of string value which semantics is hard to parse further after. We see here that even to give basic information about an ontology, the practices are not harmonized and there is a lack of knowledge of the existing possibilities. But considering the majority of responses that actually declared authoring metadata in some way, we can be optimistic about the impact a relevant and clear guideline for metadata authoring would have. Thus, the following natural question is about the familiarity of users with metadata vocabularies.

Knowledge of metadata vocabularies In Section 2, we presented some of our works: one was to review the metadata vocabularies in order to retain the relevant properties to describe an ontology. Based on this work, we provided participants with a list of 23 vocabularies (See Fig.1). The study shows that the core semantic web languages (RDFS, OWL, SPARQL) are among the most known and used (see Fig.4). DCAT, a relatively new recent metadata recommendation, appears in the top used vocabularies. DC and DCT are considered

classic also. More surprisingly, most of the numerous vocabularies really relevant for describing ontologies (e.g., DOOR, VANN, ADMS, OMV, MOD) are barely known or never used. This demonstrates that most ontology developer certainly finds what they need in these classic metadata vocabularies (e.g., RDFS, OWL, SPARQL, DCAT, DC, DCT, FOAF); or if they do not, they probably don't search further and they limit themselves to a basic description of their ontology.

Unknown	NKOS (104), IDOT (102), DOOR (100), VANN (95), ADMS (91), MOD (91), OMV (81), OboInOwl(80), DCT (48),
Known but never used	CC (45), SD (42), FOAF(38), OMV (33), VOID (29), SKOS (27), SPARQL (25), OMV (33), MOD (24)
Sometimes used	SPARQL (36), MOD (7), OMV (4)
Often used	DC(42), DCT(25), DCAT (16), OMV (6), MOD(2)
Always used (mandatory for me)	OWL (59), RDFS (54), SPARQL (41), FOAF (16), DCT (15), OMV(2)

Fig. 4. Knowledge of metadata vocabularies (the first column indicates the various options made available for the participants.

We wanted to go further and find out if other important vocabulary could have escaped our review. Among the 72 participants who answered, 21 ($\tilde{3}$0%) expressed that they do not know or use any other metadata vocabulary other than the 23 vocabularies that we listed. The rest of the participants mentioned around 27 new vocabularies (e.g., IAO, ISO 19115, LOM, LIME, LOM, Vcard). None of the vocabularies mentioned here appear more than 5 times (i.e. less than 3% of respondents). Interestingly, except IAO and ISO 19115, no other vocabulary has more than two mentions. It is also interesting to state here that some of the vocabularies that were mentioned by the participants are not really metadata vocabularies, but more domain and general purpose ontologies (e.g., DOL, SWEET, Event ontology, GCMD keywords). This means that our vocabulary review done in our previous work has not left out any important ones. There was also the mention of MIRO [12], and the CEDAR metadata workbench (`http://metadatacenter.org`) which design a framework for easy edition of metadata in general. The next question was about the overlap of current metadata properties: how to choose between two properties that have the same meaning, e.g. rdfs:label, dc:title and omv:name? In reply to this question, we received mixed and diversified replies. If two properties look equivalent, 30% of people said they choose according to the consistency of the vocabularies they use, 17% prefer simplicity over the multiplication of vocabularies and favor properties from a recommendation; 12% choose a vocabulary they know(Figure 5). Besides these, there were many other replies, for instance, "select the closest one by analyzing the meaning of the elements in the context", "coming from a large community", and "The ones I trust: who created it, is it a work of quality, will it exist in the foreseeable future?" Some of the participants (4%) replied that they randomly select the properties from the overlapping properties, and some use all

the overlapping properties when there is no clear consensus. The answers show that there is no specific approach to solve this problem of redundancy.

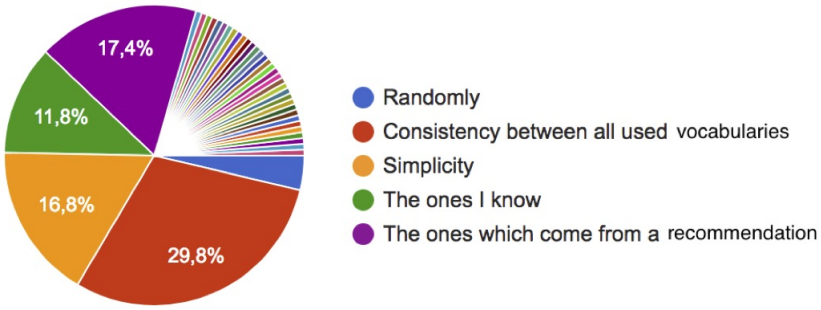

Fig. 5. Illustrates the responses for the question on how to choose between the equivalent properties

We further ask the participants to list out the metadata information they think are missing in the existing metadata vocabularies. The responses (73) to this open question were very educative as: most of the elements mentioned by responders happen to be "already available" inside some metadata vocabularies. This clearly demonstrates a lack of knowledge of the currently existing numerous vocabularies; some responders actually point it out ("There are probably too many vocabularies for the needs, already"). Nevertheless, among the most important information that came up of this response are, by order of importance:

- provenance: how it was built, with what, when, competency questions, being able to describe the inception of an ontology;
- versioning/importing: the classic way of importing an ontology and describing its information as metadata seems not enough, ontology developer needs software libraries like dependencies between ontologies;
- contributing/support: how to send feedback, participate, curate, get info about the ontology, favor open issue trackers;
- scope: domain specificity, coverage;
- use cases: (how it is used, why, examples, applicability, datasets which use)
- expressivity (the need to explicitly capture the level of expressivity of an ontology);
- quality (standard measures, FAIRness level);
- term description (more than just labels, definitions, dates, authors, but also context of use, definition origin, etc.);
- privacy constraints;
- metrics (analytical information, numbers of), mappings (information at the ontology level).

Interesting but not surprising to see that we find here the same properties as those mentioned in the question "What are the top 5 things you would like to

know when searching and selecting an ontology?": People think this information can't be expressed even though properties exist to describe them. This situation is due to the lack of knowledge of vocabularies but also to the fact that it is difficult to find one's way around, even more when one knows the overlap between all the vocabularies.

4.3 How to enhance the usage of metadata for ontologies?

Faced with the difficulty of finding and using metadata adapted to the description of ontologies, we wanted to know what could be the means to make things easier. We first pointed the role of tools: would users find it useful to be supported by a tool to author ontology metadata? Out of the total 161 responders, 70% think that a tool support for authoring ontology metadata would be useful, whereas 7% do not think so. Interestingly, a significant number of responders (23%) remained undecidable. The possible reason could be because among the people who participated to this survey, some do not develop ontologies themselves, and hence they have no opinion on this issue.

But would there be other ways than tools to move towards better practices in terms of ontology metadata? Out of the total 160 responders, almost in equal numbers (56%) think that the **guidelines**, or the **recommendations** on what to describe and how would make them focus on ontology metadata authoring. 50% think that the adoption of community standards like W3C recommendations would enable them to focus on metadata authoring. We expected to receive more replies in favor of the advanced user interface or the copy/paste template, but surprisingly we received a relatively low count (45% and 31%, respectively). It is interesting also to see that the incentives such as citation, reuse, etc. (28%) would also motivate people in authoring the metadata. Besides, the above findings, we also received mixed replies though they are marginal. For instance, some find the existing guidelines inefficient, some think that the ontology editing tools should contain the important metadata elements. 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. Then the last question was about ontology repositories: what do users think the role of ontology libraries and repositories should be with respect to ontology metadata? Most of the responders (121/168) agreed that ontology libraries should facilitate the process of ontology selection and identification using metadata. Then the responses were equalized (around 70) on the other proposed role (facilitate authoring, generate metadata, support metadata export, synchronize one another). Among the open responses, we surprisingly found six suggestions that an ontology repository shall adopt a common model for metadata and resolve internally the vocabulary heterogeneity (what we have done in AgroPortal). A few responses raise the point that on ontology library shall be able to measure a level of metadata description. Some of the other responses also pointed to the FAIRness level and quality metadata metrics. Thus, from the point of view of the community, there are at least three ways to improve the good practices in terms of ontology metadata:

tools, repositories and guidelines or recommendations (community standard). We can then consider that (i) 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; (ii) 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 other systems to use it; (iii) There is a need for a relevant and clear community adopted standard for ontology description.

4.4 Importance and role of metadata

Given these disparate practices, it is legitimate to ask what importance people place on metadata in the context of ontologies. To explore this aspect, we asked the following question: "Overall, do you think authoring and accessing ontology metadata is important?" To this question, 105 participants replied 'yes' with different level of importance. Many respondents acknowledged the importance of ontology metadata for different purposes, for instance, ontology search, discover, identify, selection; version and quality control; ontology reuse; semantic interoperability between machines; reveals the ontology development context; provide essential information to new comers; feedback to ontology designer. A couple of respondents interestingly mentioned that a lack of good metadata is certainly a good demonstrator of a lack of rigor/quality in the process of building the ontology itself. Some of the participants have also made some recommendations on the subject; among them, we can include: need for a single rigorous specification (promoted as a standard); more advanced ontology repositories; alignment between metadata vocabularies; possibility to extend such vocabulary metadata to other semantic resources, not only ontology; use metadata for quality control; the current editing line is too complex; automatically processes metadata inside ontology repositories.

5 Conclusion

Because of their diversified profile, survey participants are a representative sample of the ontology community. The previous analyses show that paradoxically, ontology users, on the one hand, recognize the importance of metadata and hope for rich descriptions when searching for an ontology, but on the other hand, ontology developers do not describe a lot their resource and use only a limited number of properties among the existing ones. Thus, users express that the information sought is generally not present in ontology metadata, even though properties exist to describe it. This situation has been confirmed by all our studies on the subject. An interesting point in this survey is that when developers describe their

ontology, either they do it in a 'personal' way (e.g; with string value), or, if they use a metadata vocabulary, they will choose among the best known ones: core semantic web languages (RDFS, OWL, SPARQL), DCAT, Dublin Core. Very few people use dedicated vocabulary such as DOOR, VANN, ADMS, OMV or MOD. This is not surprising when considering the efforts needed to just identify the potentially relevant vocabularies that could be used to describe ontologies: there is not a single reference to describe ontologies but a quantity of properties taken from several different vocabularies which overlap a lot. Thus, the practices in terms of ontology metadata are not harmonized due to the lack of knowledge about metadata vocabularies and the absence of a clearly and unique recommendation. Our motivation is to better understand the practices and the needs of the ontology community users and then consider the follow-up to our different work. There is a need clearly expressed for metadata authoring guidelines and for harmonization of existing metadata vocabularies. We know now that our vocabulary review has not left out any important ones. Thus, we can rely on our previous work (AgroPortal metadata model & MOD) to propose in the context of the VSSIG a recommendation for ontology description. This recommendation could be a metadata vocabulary (which would be a revised and extended version of the current MOD) or an application profile. Other perspectives focus on improving metadata support through tools and repositories.

Acknowledgements. This work is partly achieved within the Semantic Indexing of French biomedical Resources (SIFR âĂŞ `www.lirmm.fr/sifr`) project that received funding from the French National Research Agency (grant ANR-12-JS02-01001), the European UnionâĂŹs Horizon 2020 research and innovation programme 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. This work has also been partially funded by the Indian Statistical Institute under the Start-Up Grant project.

References

1. Baclawski, K., Schneider, T.: 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 Workshop Proceedings, vol. 514, p. 10. CEUR-WS.org, Washington, DC., USA (oct 2009)
2. Chao, T.: Mapping Methods Metadata for Research Data. Digital Curation 10(1), 82–94 (2015)
3. Dutta, B., Nandini, D., Shahi, G.: Mod: Metadata for ontology description and publication. International Conference on Dublin Core and Metadata Applications pp. 1–9 (2015)
4. Dutta, B., Toulet, A., Emonet, V., Jonquet, C.: New Generation Metadata vocabulary for Ontology Description and Publication. In: Garoufallou, E., Virkus, S.,

Siatri, R., Koutsomiha, D. (eds.) 11th Metadata and Semantics Research Conference, MTSR'17. Communications in Computer and Information Science, vol. 755, pp. 173–185. Springer, Tallinn, Estonia (nov 2017)
5. Hartmann, J., Palma, R., Sure, Y., Suárez-Figueroa, M.C., Haase, P., Gómez-Pérez, A., Studer, R.: Ontology metadata vocabulary and applications. In: Meersman, R., Tari, Z., Herrero, P. (eds.) On the Move to Meaningful Internet Systems 2005: OTM 2005 Workshops. pp. 906–915. Springer Berlin Heidelberg, Berlin, Heidelberg (2005)
6. Hartmann, J., Sure, Y., Haase, P., Suarez-Figueroa, M.: OMV;ontology metadata vocabulary. In: Welty, C. (ed.) Workshop on Ontology Patterns for the Semantic Web, WOP'05. p. 9. Springer, Galway, Irland (nov 2005)
7. Jonquet, C., Toulet, A., Arnaud, E., Aubin, S., Yeumo, E.D., Emonet, V., Graybeal, J., Laporte, M.A., Musen, M.A., Pesce, V., Larmande, P.: AgroPortal: an ontology repository for agronomy. Computers and Electronics in Agriculture 144, 126–143 (jan 2018)
8. Jonquet, C., Toulet, A., Dutta, B., Emonet, V.: Harnessing the power of unified metadata in an ontology repository: the case of AgroPortal. Data Semantics (2018, Under review –2nd round– minor revision)
9. Jonquet, C., Toulet, A., Arnaud, E., Aubin, S., DzaleÌĄ Yeumo, E., Emonet, V., Graybeal, J., Musen, M.A., Pommier, C., Larmande, P.: Reusing the NCBO BioPortal technology for agronomy to build AgroPortal. In: International Conference on Biomedical Ontology and BioCreative (ICBO BioCreative 2016). p. 3p. Corvallis, OR (2016)
10. Khan, Z.C., Keet, C.M.: The foundational ontology library romulus. In: Cuzzocrea, A., Maabout, S. (eds.) Model and Data Engineering. pp. 200–211. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)
11. Martinez-Romero, M., Jonquet, C., O'Connor, M.J., Graybeal, J., Pazos, A., Musen, M.A.: NCBO Ontology Recommender 2.0: An Enhanced Approach for Biomedical Ontology Recommendation. Biomedical Semantics 8(21) (jun 2017)
12. Matentzoglu, N., Malone, J., Mungall, C., Stevens, R.: MIRO: guidelines for minimum information for the reporting of an ontology. Journal of Biomedical Semantics 9(1), 1 (January 2018)
13. McGuinness, D.L.: Spinning the semantic web: bringing the World Wide Web to its full potential, chap. Ontologies, pp. 171–194. MIT Press (2003)
14. Naskar, D., Dutta, B.: Ontology And Ontology Libraries: A Study From An Ontofier And An Ontologist Perspective. In: 19th International Symposium on Electronic Theses and Dissertations, ETD'16. pp. 1–12. Lille, France (jul 2016)
15. 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.: Semantic web and big data meets applied ontology: The ontology summit 2014. Appl. Ontol. 9(2), 155–170 (Apr 2014)
16. Park, J., Oh, S., Ahn, J.: Ontology selection ranking model for knowledge reuse. Expert Systems with Applications 38(5), 5133–5144 (2011)
17. Toulet, A., Emonet, V., Jonquet, C.: Modèle de métadonnées dans un portail d'ontologies. In: Diallo, G., Kazar, O. (eds.) 6èmes Journées Francophones sur les Ontologies, JFO'16. Bordeaux, France (oct 2016)
18. Zen, M.L.: Metadata for Terminology / KOS Resources. In: 8th Networked Knowledge Organization Systems Workshop. Washington, DC, USA (sep 2008)
19. Zeng, M.L.: Knowledge Organization Systems (KOS). Knowledge organization 35(2-3), 160–182 (2008)