Modèle de métadonnées dans un portail d'ontologies

Anne Toulet, Vincent Emonet, Clement Jonquet

Modèle de métadonnées dans un portail d'ontologies

Anne Toulet* — Vincent Emonet* — Clément Jonquet*,**

* *Laboratoire d'informatique, de Robotique et de Microélectronique de Montpellier (LIRMM) CNRS – Université de Montpellier*
161, rue Ada – 34095 Montpellier
{jonquet, toulet, emonet}@lirmm.fr

** *Center for Biomedical Informatics Research, Stanford University, USA*

RÉSUMÉ. *Les communautés scientifiques utilisent un nombre croissant d'ontologies. Pour les mettre à disposition, il existe des portails d'ontologies, à l'exemple du NCBO BioPortal qui regroupe actuellement plus de 500 ontologies biomédicales. Mais face à cette avalanche de ressources, comment trouver l'ontologie qui répondra à nos besoins ? Une solution consiste à décrire chaque ontologie avec des métadonnées appropriées. Or, il n'existe pas à ce jour de vocabulaire de métadonnées suffisamment exhaustif pour répondre à ce besoin. Nous avons passé en revue un grand nombre de vocabulaires, tels que Dublin Core, OMV, DCAT ou VOID ainsi que les propriétés implémentées par les portails d'ontologies les plus courant. Nous en avons produit un modèle simplifié composé de 124 propriétés. Nous présentons ici quelques exemples d'utilisation de ces propriétés dans AgroPortal, un portail d'ontologies dédié à l'agronomie, et nous expliquons comment elles sont gérées et utilisées pour la description et l'identification d'ontologies.*

ABSTRACT. *Scientific communities are using an increasing number of ontologies. Repositories make them available, like the NCBO BioPortal which currently hosts more than 500 biomedical ontologies. Now the question is how to find the ontology we need? One solution is to describe each ontology with appropriate metadata. However, none of the existing metadata vocabularies can completely meet this need if taken independently. We have reviewed a large number of vocabularies, such as Dublin Core, OMV, DCAT, or VOID, as well as the properties implemented by common ontology repositories. We then listed those properties into a simplified model of 124 properties. We present a few examples of use of these properties within the AgroPortal, an ontology repository for agronomy, and explain how the portal handles these properties to facilitate ontology description and selection.*

MOTS-CLÉS : *vocabulaire de métadonnée, ontologie, web sémantique, portail d'ontologies, BioPortal, AgroPortal, standard W3C.*

KEYWORDS: *metadata vocabulary, ontology, semantic description, ontology repository, BioPortal, AgroPortal, ontology relation, W3C standards.*

1. Introduction

L'explosion des volumes de données et la proposition du web sémantique pour les décrire et les lier [2] ont fait croître exponentiellement la création et l'utilisation de terminologies, d'ontologies et autres vocabulaires contrôlés. Actuellement, la page d'accueil de Swoogle [6], moteur de recherche du web sémantique, exploite au moins 10000 ontologies. Pour héberger et publier ces ontologies, les portails

d'ontologies ont fleuri au cours de la dernière décennie, à l'exemple du NCBO BioPortal (http://bioportal.bioontology.org) qui héberge aujourd'hui plus de 500 ontologies du domaine biomédical. Dans ce contexte, les ontologies se doivent d'être envisagées comme des ressources à part entière, qui nécessitent d'être découvertes, rendues accessibles, utilisées avec efficacité et à bon escient. Mais comment répondre à cette exigence ?

L'une des solutions est de décrire précisément chaque ontologie avec des métadonnées appropriées. Cependant, si l'on regarde le contenu du NCBO BioPortal, l'utilisation de métadonnées par les développeurs d'ontologies est encore très minoritaire ou se limite à quelques champs de base, en général obligatoires pour l'hébergement de l'ontologie dans un portail ou automatiquement générés par l'outil d'édition. Cela n'a rien de surprenant lorsque l'on considère les efforts nécessaires pour simplement identifier les vocabulaires potentiellement pertinents pour décrire une ontologie[1]. En outre, les éditeurs et portails d'ontologies n'ont pas pris ce problème suffisamment au sérieux pour encourager la description des ontologies avec des vocabulaires standards. Ainsi, lors de la recherche d'une ontologie, maintes questions peuvent se poser : qui a édité ou contribué, à quelle date ? Quelle langue naturelle, quelle méthodologie, quels outils ont été utilisés ? Quels sont les formats disponibles ? Quels sont les droits attachés à cette ontologie ? Ces interrogations peuvent également concerner un ensemble d'ontologies. Par exemple, pour un domaine concerné, quelles sont les ontologies les plus utilisées ? Quelles relations entretiennent-elles ? Qui sont les principaux contributeurs du domaine ? Toutes ces informations peuvent être renseignées à l'aide de métadonnées et ont pour but de faciliter le processus de sélection d'une ontologie, ce qui a été évalué comme essentiel pour en permettre une utilisation efficace [19,18,15]. Par ailleurs, des métadonnées harmonisées devraient également fournir des informations globales sur les ontologies d'un domaine spécifique, et aider à comprendre « l'écosystème » formé par ces ontologies. Par exemple, si l'information de l'outil utilisé pour éditer une ontologie n'est pas indispensable pour l'usage d'une ontologie donnée, cette information renseignée pour toutes les ontologies d'un domaine pourrait nous permettre de répondre à la question : quel est l'éditeur d'ontologie le plus fréquemment utilisé dans ma communauté ?

Dans la suite de cet article, nous présentons en section 2 un état de l'art des travaux liés aux vocabulaires de métadonnées, y compris dans les portails d'ontologies. La Section 3 détaille notre approche pour établir une liste exhaustive de propriétés de métadonnées. La section 4 présente comment ce modèle est intégré au portail d'ontologies AgroPortal, comment les métadonnées y sont gérées, utilisées et valorisées. Les sections 5 et 6 abordent les perspectives et les enjeux des métadonnées dans ce contexte et concluent le papier.

[1] Nous considérons les termes ontologies, terminologies et vocabulaires comme équivalents pour nommer des systèmes d'organisation de connaissances. Toutefois, nous utiliserons le mot *ontologie* pour identifier le sujet décrit par les métadonnées et le mot *vocabulaire* pour les ressources qui permettent de les décrire.

2. État de l'art : quelles métadonnées pour décrire une ontologie ?

Dans ce paragraphe, nous expliquons comment nous avons identifié les propriétés de métadonnées permettant de décrire de façon très complète une ontologie : (i) en explorant les vocabulaires existants ; (ii) en étudiant les métadonnées proposées par des portails d'ontologies disponibles.

2.1. *Revue des vocabulaires existants*

Nous avons passé en revue un grand nombre de vocabulaires existants. De fait, il en existe de plus en plus, permettant de décrire des entités très variées. Nous n'avons retenu que ceux qui pouvaient nous fournir des propriétés pertinentes dans le cadre de cette étude i.e., pour décrire une ontologie. Ceux-ci ont été identifiés soit dans la littérature du web sémantique, soit parce qu'ils apparaissent explicitement dans les portails d'ontologies, soit encore parce qu'ils sont cités dans des études similaires comme celle sur les jeux de données proposée par le groupe de travail HCLS (https://www.w3.org/TR/hcls-dataset).

Parmi eux (cf. Tableau 1), très peu sont dédiés spécifiquement à la description d'ontologies (*OMV*, *DOOR*), de vocabulaires (*VOAF*) ou de jeux de données (*VOID*, *DCAT*, *SCHEMA*). La plupart sont généralistes, à l'exemple de *DC*, *DCT*, *PROV* ou *DOAP*. Il n'existe qu'un seul vocabulaire spécifiquement dédié à la description d'ontologies, *OMV* [9], qui a été développé et publié en 2005 dans le cadre de projets européens et qui propose 37 propriétés. Malheureusement, cette initiative s'arrêta en 2007. Une limitation d'*OMV* est de ne pas s'appuyer sur les vocabulaires standards existants. Cette faiblesse a été partiellement corrigée par la publication en 2015 du vocabulaire *MOD* [8], très similaire à *OMV* (mais qui ne l'utilise pourtant pas) qui intègre des propriétés existantes de *SKOS*, *FOAF* et *DC*. *MOD* a ajouté 7 propriétés à celles existantes dans *OMV*, mais comme nous le verrons dans la section 3.1, il en manque encore beaucoup pour atteindre nos objectifs. En 2005, le très simple mais pertinent *VANN* fut rendu disponible et est assez utilisé depuis. En 2009, un autre vocabulaire fut publié pour décrire les relations entre ontologies, *DOOR* [1], mais jamais réellement utilisé en dehors du projet NeON. Il décrit très précisément et d'un point de vue logique, 32 relations entre ontologies, organisées hiérarchiquement. *DOOR* réutilise les relations offertes par *OWL*, comme par exemple owl:imports. Plus récemment, *VOAF* [23] fut produit pour « décrire des vocabulaires utilisés dans le monde des données liées et ouvertes. En particulier, il propose des propriétés qui expriment les différents liens qui existent entre vocabulaires (..) Il repose lui-même sur *DC* et *VOID*. » Mais il n'utilise pas du tout les propriétés offertes par *OWL* ou *DOOR*.

Parmi les travaux récents pour décrire les jeux de données, on trouve différents vocabulaires offrant des propriétés intéressantes : *VoID* [13], qui peut être utilisé « pour exprimer des métadonnées générales basées sur *DC*, des métadonnées décrivant les accès, la structure, et les liens entre jeux de données » ; *IDOT* [12], développé par l'Institut Européen de Bioinformatique pour spécifier les modèles d'URI ; *DCAT* et sa spécialisation *ADMS*, qui sont les recommandations du W3C les plus récentes pour décrire des ressources de données. Enfin, nous citerons *Schema.org*, proposé et adopté en 2011 par Google, Bing and Yahoo! et qui inclut

une classe Dataset. Nous citerons également les initiatives suivantes : *FOAF* ou *DOAP* pour décrire des documents et des projets ; *CC* pour expliciter les différents types de licences ; *SD* pour décrire les SPARQL endpoints ; *PROV* ou *PAV* pour préciser la provenance d'une ressource ; *OboInOwl* [21] le support qui permet de convertir les ontologies (y inclus leur métadonnées) du format OBO au format OWL. Parmi les vocabulaires que nous avons étudiés mais pas réutilisés (non listés dans le Tableau 1), nous pouvons également mentionner *Extension to the VoID* [14], une extension de *VoID*, principalement pour des descriptions statistiques ; *Citation Typing Ontology* (*CiTO*) ; *Protocol for Web Description Resources* (*POWDER*) qui propose un modèle pour décrire les ressources du web ; *DDI-RDF Discovery Vocabulary* (*DISCO*), un vocabulaire de métadonnées pour la documentation des données de recherche ; *Information Artifact Ontology* (*IAO*) [3], qui s'attache à décrire des contenus d'information tels que les documents, bases de données ou images numériques ; *Semanticscience Integrated Ontology* (*SIO*) [7] pour décrire les ressources biomédicales. Pour finir, nous mentionnerons également le document ISO/IEC 19763-3 (*Metamodel framework for interoperability* (*MFI*) – Part 3: Metamodel for ontology registration) dont la version la plus récente remonte à 2010 mais n'est pas publique.

2.2. *Quid des métadonnées dans les portails d'ontologies ?*

Nous nous sommes intéressés à tout type d'outils web dédié spécifiquement aux ontologies. Que ce soit des : (i) portails comme le *NCBO BioPortal, OBO Foundry Ontobee, EBI Ontology Lookup Service, MMI Ontology Registry and Repository, AgroPortal* ou *AberOWL* ; (ii) catalogues comme l'*OKFN Linked Open Vocabularies, WebProtégé, FAO VEST Registry, BioSharing ;* (iii) moteurs de recherche comme *Watson, Swoogle,* ou *Sindice.* Pour une revue partielle de certains de ces outils : voir [5]. Un recensement récent et une classification de certains de ces outils sont également proposés et discutés dans [16] (dont on trouvera un condensé dans [8]). Dans tous ces outils, les vocabulaires de métadonnées pour décrire les ontologies sont rarement utilisés. Cependant, si l'un d'eux est utilisé, il s'agit en général d'OMV. À l'exception de la plateforme ouverte Linked Open Vocabularies [22] et dans une certaine mesure du NCBO BioPortal [17], la question de la normalisation des descriptions d'une ontologie n'est pas vraiment une question centrale. BioPortal utilise 65 propriétés de métadonnées (dont 45 non spécifiques au portail) qui servent de base à notre propre listing et qui seront détaillées dans la section 3. Le modèle sous-jacent dans BioPortal est partiellement fourni par OMV et dans certains cas utilise les mêmes noms de propriétés. Du fait qu'ils utilisent le même code source, cette situation est similaire pour le portail MMI-ORR qui a ajouté d'autres champs qui lui sont propres. Le catalogue LOV quant à lui utilise explicitement VoID et VOAF, ce dernier ayant d'ailleurs été conçu dans ce cadre-là. Le site OBO Foundry [20] réutilise des métadonnées empruntées à une vingtaine de vocabulaires tels que DC, FOAF, IDOT, VoID, DOAP (http://obofoundry.github.io/registry/context.jsonld). Par ailleurs, l'application associée Ontobee [24] propose quelques métadonnées très générales telles que IRI, home, contact, etc, et affiche aussi comme propriétés d'annotations les métadonnées incluses dans l'ontologie elle-même. Ce portail calcule aussi quelques métriques pour chaque ontologie comme BioPortal. AberOWL [10] et OLS [4] proposent

quelques métadonnées générales ainsi que l'affichage de celles contenues dans l'ontologie (également sous forme de propriétés d'annotations). Contrairement à ce qui est fait dans OBO Foundry, leurs métadonnées ne sont pas décrites à l'aide de vocabulaires standards. Les autres outils listés ci-dessus n'utilisent aucun vocabulaire standard pour les métadonnées d'ontologies. Nous confirmons donc en partie l'analyse faite dans [8, 16].

Prefixe	Nom	W3C	#
rdfs	RDF Schema	R	3
owl	OWL 2 Web Ontology Language	R	7/11
skos	Simple Knowledge Organization System	R	5/14
dc	Dublin core	R	15/15
dcterms	Dublin Core qualifié	R	39/55
omv	Ontology Metadata Vocabulary		37/37
mod	Metadata for Ontology Description and Publication		24/25
door	Descriptive Ontology of Ontology Relations		11/32
voaf	Vocabulary of a Friend		12/16
void	Vocabulary of Interlinked Datasets	N	16/24
vann	Vocabulary for annotating vocabulary descriptions		5/6
idot	Identifiers.org		6/9
dcat	Data Catalog Vocabulary	R	4/5
adms	Asset Description Metadata Schema	N	9/41
schema	Schema.org		41/90
foaf	Friend of a Friend Vocabulary	N	9/11
doap	Description of a Project		17/25
cc	Creative Commons Rights Expression Language		5/5
prov	Provenance Ontology	R	10/22
pav	Provenance, Authoring and Versioning		16/30
oboInOwl	OboInOwl Mappings		9/13

Tableau 1. *Vocabulaires étudiés. La colonne W3C mentionne si le vocabulaire est une recommandation (R) ou s'il s'agit d'une note (N). La dernière colonne (#) est le nombre de propriétés retenues parmi celles qui pourraient être utilisées.*

3. Construction d'un modèle simplifié de métadonnées

La démarche adoptée pour construire notre modèle s'est faite en deux temps : nous avons d'abord établi une liste de toutes les propriétés de métadonnées candidates ; nous en avons ensuite tiré un modèle simplifié permettant de prendre en charge l'ensemble des propriétés retenues. Les critères d'inclusion des propriétés dans notre listing sont les suivants :

Consistance sémantique – les propriétés doivent pouvoir s'appliquer à une ontologie ; par exemple, suivant le vocabulaire et la propriété ciblée, une ontologie sera envisagée comme un jeu de données (dataset), une ressource informatique (asset), un projet, un document, etc.

Pertinence – la propriété doit avoir un sens dans le contexte de description d'une ontologie c'est-à-dire, renseigner une information en lien avec l'ontologie décrite.

Non spécificité – la propriété ne doit pas être spécifique à un portail d'ontologies et doit pouvoir être exploitée pour décrire l'ontologie en dehors du portail.

3.1. *Liste complète des propriétés retenues*

À partir des vocabulaires du Tableau 1, nous avons dressé une liste de 316 propriétés pertinentes. Concernant les portails d'ontologies, nous avons pris le NCBO BioPortal comme source de propriétés car c'est celui qui en fournit le plus. Si ces propriétés ne sont pas couvertes par ailleurs, nous les incorporons dans notre liste avec le préfixe bpm pour « Bioportal metadata ». La liste des 316 propriétés recensées est disponible à https://github.com/agroportal/documentation/ (metadata).

3.2. *Modèle simplifié*

Parmi les propriétés constituant la liste principale, beaucoup d'entre elles définissent exactement la même chose. Par exemple, les informations de version d'une ontologie peuvent être décrites par : omv:version, owl:versionInfo, mod:version, doap:release, pav:version and schema:version. D'autres propriétés ont un sens très proche, à l'exemple de bpm:homepage, foaf:homepage, cc:attributionURL, mod:homepage, doap:blog, and schema:mainEntityOfPage.

Dans l'objectif de simplifier notre liste et de l'implémenter dans un portail d'ontologies, nous avons groupé en fonction de leur description et/ou usage les propriétés de sens similaire ou proche, puis choisi pour chaque groupe ainsi constitué une propriété représentative (« propriété par défaut »). Ce choix s'est effectué de la façon suivante : lorsqu'elles existent, les propriétés sont celles d'OMV[2] ; dans le cas contraire, la préférence est donnée à un vocabulaire qui est soit une recommandation, soit une note du W3C. De cette façon, nous obtenons 124 propriétés qui constituent le modèle simplifié sur lequel nous nous appuyons pour gérer l'ensemble des 316 métadonnées candidates.

Le rôle des équivalences[3] posées entre propriétés d'un même groupe n'est pas d'offrir un vocabulaire unique pour décrire les ontologies (cette question sera

[2] Dans le contexte de notre projet AgroPortal, nous avons fait ce choix pour le moment dans la continuité de ce qui avait été fait sur le NCBO BioPortal. Cela ne préfigure en rien les propriétés par défaut qu'il faudrait prendre dans un autre contexte.

[3] Nous laissons volontairement de côté l'usage des mots "mapping" ou "alignement."

discutée dans la Section 5) mais de proposer un modèle capable de prendre en charge (ou capturer) des métadonnées décrites avec n'importe lequel des vocabulaires identifiés précédemment au sein d'un portail d'ontologies. En effet, comme nous le verrons dans la section 4, cette liste réduite permet de valoriser les métadonnées au sein d'un portail d'ontologies (quelle que soit la propriété originale utilisée) et offre plus de lisibilité à quelqu'un qui voudrait décrire son ontologie. Pour une meilleure compréhension, nous avons regroupé ci-après les propriétés de notre modèle simplifié en catégories (Tableau 2).[4]

Catégorie	Propriétés
Intrinsèque	omv:acronym, omv:name, omv:hasOntologyLanguage, omv:URI, omv:naturalLanguage, omv:version, omv:description, omv:status, omv:keywords, omv:hasOntologySyntax, omv:conformsToKnowledgeRepresentationParadigm, omv:designedForOntologyTask, omv:hasFormalityLevel, omv:hasLicense, omv:usedOntologyEngineeringMethodology, omv:usedOntologyEngineeringTool, omv:isOfType, owl:deprecated, owl:versionIRI, cc:morePermissions, cc:useGuidelines, dcterms:coverage, dcterms:identifier, dcterms:alternative, dcterms:abstract, foaf:depiction, foaf:logo, mod:competencyQuestion, skos:hiddenLabel, vann:preferredNamespaceUri, vann:preferredNamespacePrefix, voaf:toDoList
Description	omv:reference, omv:documentation, omv:hasDomain, omv:resourceLocator, omv:notes, bpm:prefLabelProperty, bpm:definitionProperty, bpm:synonymProperty, bpm:authorProperty, bpm:hierarchyProperty, bpm:obsoleteProperty, bpm:group, bpm:obsoleteParent, dcterms:accrualMethod, dcterms:accrualPeriodicity, dcterms:accrualPolicy, doap:repository, doap:bug-database, doap:mailing-list, idot:exampleIdentifier, foaf:homePage, schema:award, schema:copyrightHolder, schema:associatedMedia, schema:includedInDataCatalog, void:uriRegexPattern
Personnes	omv:hasContributor, omv:hasCreator, dcterms:publisher, dcat:contactPoint, schema:translator
Relation	omv:useImports, omv:hasPriorVersion, omv:isBackwardCompatibleWith, omv:isIncompatibleWith, bpm:viewOf, bpm:views, dcterms:hasVersion, dcterms:hasPart, dcterms:isFormatOf, dcterms:hasFormat, door:ontologyRelatedTo, door:similarTo, door:comesFromTheSameDomain, door:explanationEvolution, door:hasDisparateModelling, door:isAlignedTo, schema:translationOfWork, schema:workTranslation, voaf:usedBy, voaf:metadataVoc, voaf:generalizes, voaf:hasDisjunctionsWith
Contenu	omv:keyClasses, bpm:csvDump, schema:url, sd:endpoint, void:propertyPartition, void:classPartition, void:rootResource, void:dataDump, void:uriLookupEndpoint, void:openSearchDescription, vann:example
Communauté	omv:knownUsage, omv:endorsedBy, bpm:projects, bpm:analytics, dcterms:audience, foaf:fundedBy, schema:review, schema:comments
Dates	omv:creationDate, omv:modificationDate, dcterms:dateSubmitted, dcterms:valid, pav:curatedOn
Métriques	omv:numberOfClasses, omv:numberOfIndividuals, omv:numberOfProperties, omv:numberOfAxioms, bpm:maxDepth, bpm:maxChildCount, bpm:averageChildCount, bpm:classesWithOneChild, bpm:classesWithMoreThan25Children, bpm:classesWithNoDefinition, void:entities
Provenance	dcterms:source, pav:curatedBy, prov:wasGeneratedBy, prov:wasInvalidatedBy

Tableau 2. *Propriétés par défaut du modèle simplifié regroupées en catégories.*

[4] Ces catégories ne font pas partie du modèle, elles servent juste à regrouper les propriétés.

4. Exemple d'application dans le domaine agronomique

Comme en biomédecine, beaucoup d'ontologies sont produites pour annoter des données dans le domaine agronomique. L'objectif principal du projet AgroPortal [11] est de développer et proposer un portail d'ontologies de référence dans ce domaine. En réutilisant la technologie du NCBO BioPortal, nous avons développé un premier prototype (http://agroportal.lirmm.fr). Il contient une cinquantaine d'ontologies du domaine agronomique.

Le portail d'ontologies est le lieu par excellence où donner un sens et valoriser les métadonnées. Nous avons donc intégré à AgroPortal notre modèle simplifié en modifiant et en adaptant le code d'origine du portail, tout en restant compatible avec le NCBO BioPortal, de façon à prendre en charge toutes les métadonnées de l'ontologie, qu'elles soient : (i) saisies par l'utilisateur au moment du dépôt ; (ii) calculées par le portail ; (iii) présentes dans le fichier d'origine de l'ontologie. Dans ce dernier cas, nous avons rendu possible l'extraction automatique des métadonnées lorsqu'elles sont présentes dans le fichier d'origine (OWL, SKOS ou OBO) de l'ontologie, ce qui permet de peupler les propriétés par défaut du modèle indépendamment du vocabulaire de métadonnées utilisé tout en gardant l'information à la source c'est-à-dire dans le fichier de l'ontologie. L'utilisateur peut visualiser (et éditer) les métadonnées associées à son ontologie dans l'interface utilisateur (Figure 1). Il peut aussi obtenir l'ensemble des informations concernant sa dernière soumission via le web service REST du portail et récupérer ainsi toutes les métadonnées associées à sa dernière soumission au format JSON-LD. E.g.,

http://data.agroportal.lirmm.fr/ontologies/BIOREFINERY/latest_submission?display=all

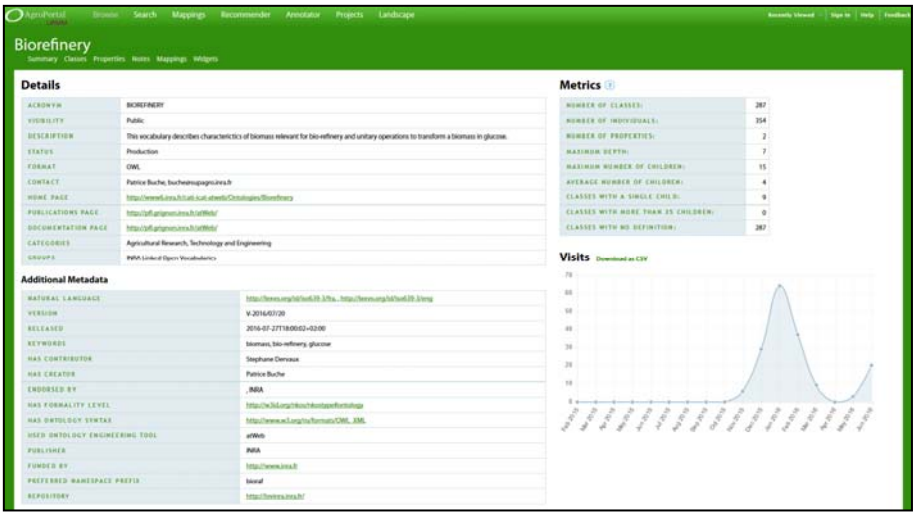

Figure 1. *Un exemple de visualisation de métadonnées dans l'interface utilisateur pour l'ontologie Biorefinery (http://agroportal.lirmm.fr/ontologies/BIOREFINERY). Les nouvelles métadonnées sont inclues dans la partie 'Additional Metadata'.*

Ainsi, par comparaison avec AberOWL, Ontobee et OLS, AgroPortal n'affiche pas directement les métadonnées de l'ontologie sous forme de propriétés

d'annotation, mais il utilise les métadonnées d'origine pour remplir un modèle de métadonnées commun à toutes les ontologies du portail, qui utilise lui-même des vocabulaires standards. Cela nous permet de traiter uniquement avec la liste réduite au sein d'AgroPortal, ce qui facilite la valorisation et l'utilisation des métadonnées. Cependant, le choix technique parfait serait probablement celui de LOV, qui ne traite que les métadonnées d'un vocabulaire annoncé spécifique. Mais l'objectif de cet article est entre autre de démontrer que le choix d'un vocabulaire standard seulement ne suffit pas à couvrir tous les champs requis pour les ontologies.

Pour utiliser et valoriser les métadonnées de l'ontologie dans le portail, nous sommes en train de développer de nouvelles fonctionnalités dans l'interface utilisateur (et dans les services web correspondants). Notre objectif est de faciliter :

– la description d'une ontologie, en affichant (et en utilisant des diagrammes) beaucoup plus d'informations que ne le faisait le système d'origine (Figure 1) ;

– la sélection d'ontologies, en enrichissant le processus de recherche par facettes ou de classement des ontologies par propriété (page « Browse ») ;

– la compréhension de l'écosystème des ontologies du portail, en affichant des statistiques ou synthèse sur une propriété donnée pour toutes les ontologies du portail. La figure 2 illustre un premier prototype de la page « Landscape ».

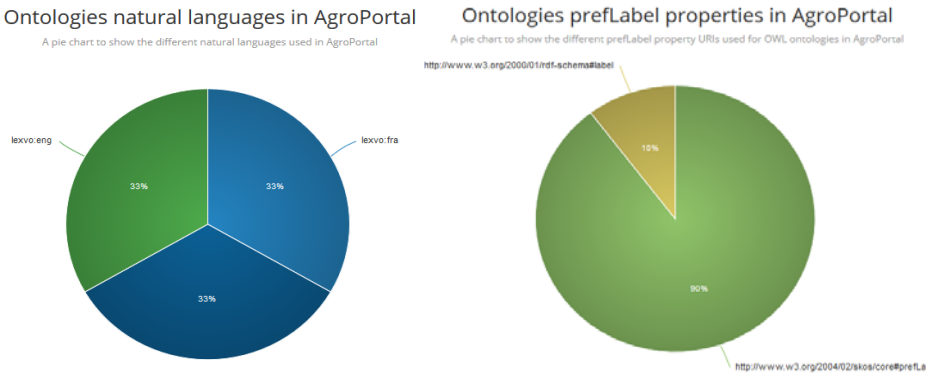

Figure 2. *Diagramme affichant la synthèse des propriétés* omv:naturalLanguage *et* bpm:prefLabelProperty *pour toutes les ontologies dans AgroPortal.*

5. Discussion

L'un des objectifs de cette étude est d'offrir aux développeurs d'ontologies une liste dans laquelle ils peuvent choisir des propriétés pour mieux décrire leurs ontologies. Nous sommes partis du constat qu'il n'y avait pas de vocabulaire adapté pour répondre à ces besoins, mais que le vocabulaire OMV fournissait une base intéressante sur laquelle assoir notre modèle. Selon nous, les principales limites d'OMV qui pourraient expliquer pourquoi il n'est pas vraiment utilisé aujourd'hui sont : (i) la non réutilisation d'autres vocabulaires de métadonnées[5] ; (ii) le fait qu'il

[5] Notons qu'en 2005, même s'il n'existait pas autant de vocabulaires qu'aujourd'hui, des standards importants tels qu'OWL, DC ou FOAF auraient pu être utilisés.

n'ait jamais été inclus à un éditeur d'ontologie tel que Protégé par exemple. Cela aurait grandement facilité l'adoption du vocabulaire si les développeurs d'ontologie n'avaient qu'à remplir quelques formulaires directement dans leur logiciel d'édition préféré ; (iii) les propriétés de métadonnées n'ont jamais été vraiment utilisées et valorisées par les portails d'ontologie, ce qui aurait été le meilleur moyen d'inciter les utilisateurs à en remplir. En particulier, cela n'a pas été exploité dans le NCBO BioPortal alors qu'OMV fait partie du modèle sous-jacent.

5.1. *Vers un nouveau vocabulaire pour décrire des ontologies*

La liste complète des propriétés que nous avons construite montre clairement qu'il existe de nombreux recoupements entre les différents vocabulaires étudiés. Par ailleurs, à quelques exceptions près, beaucoup de ces vocabulaires ne réutilisent pas l'existant et redéfinissent des choses qui ont déjà été décrites à plusieurs reprises (comme par exemple les dates pour lesquelles au moins 25 propriétés sont disponibles). En outre, il existe également une dizaine de propriétés dans BioPortal qui ne sont dans aucun vocabulaire (e.g., la notion de groupe, de vue, de projets associés). Dans ce contexte, nous pouvons donc nous demander si développer un nouveau vocabulaire de métadonnées est une bonne idée. Cependant, une piste intéressante est de se tourner vers l'utilisation de profils d'application [25] tels que proposés par le Dublin Core Metadata Initiative[6]. Cette solution permet de créer un nouveau vocabulaire tout en intégrant ceux déjà existants. Si un tel vocabulaire était communément adopté et devenait une spécification claire, les portails d'ontologies l'adopteraient, ce qui clarifierait la prise en charge des propriétés à la fois pour les développeurs d'ontologies et les portails. Dans tous les cas, nous pensons que la création d'un tel vocabulaire se doit d'être envisagée en collaboration avec différents groupes d'experts et nous voyons ceci comme une perspective.

5.2. *Édition de métadonnées*

Un autre aspect important dans les métadonnées est que peu de personnes aiment les remplir. Quelle que soit la taille de notre liste réduite, la question essentielle est : comment pouvons-nous faciliter l'édition des métadonnées pour les développeurs de l'ontologie ? Les métadonnées doivent être autant que possible générées automatiquement par les éditeurs d'ontologies ou par des outils externes[7], pour prédire ou calculer des métadonnées telles que le langage naturel utilisé, des métriques, des dates, etc. Dans cette perspective :

– il appartient aux logiciels d'édition d'ontologies de jouer un rôle majeur pour motiver et faciliter l'édition de métadonnées. Nous envisageons de nous rapprocher de l'équipe qui développe Protégé pour y intégrer de telles fonctionnalités ;

– les portails d'ontologies peuvent contribuer à cette démarche en facilitant l'édition, la prédiction et le calcul de métadonnées et, lorsque cela a du sens, de permettre à l'utilisateur de récupérer ces métadonnées directement dans un format à inclure dans le fichier original de l'ontologie.

---

[6] http://dublincore.org/documents/profile-guidelines

[7] Par exemple, BioPortal utilise le « Manchester Ontology Metrics » pour calculer certaines métadonnées.

6. Conclusion et perspectives

Dans cette étude, notre première motivation était de passer en revue les vocabulaires disponibles pour décrire une ontologie et de lister les propriétés pertinentes dans ce contexte. Depuis l'initiative d'OMV en 2005, il y a eu de nombreuses propositions, en particulier avec l'apparition du web de données, mais aucune qui puisse répondre à nos exigences. Ce travail nous a permis d'identifier à la fois la redondance et les manques dans les propriétés des vocabulaires existants, d'en produire une liste complète puis un modèle simplifié en utilisant une structuration en groupes de propriétés similaires. Nous avons implémenté cette liste dans le portail d'ontologies AgroPortal et sommes en train de valoriser ces résultats (travaux de visualisations en cours de développement). Nous avons expliqué l'objectif de produire un nouveau vocabulaire ou profil d'application pour décrire les ontologies et cela est clairement identifié comme l'étape suivante à ce travail.

Le nouveau modèle étant d'ores et déjà implémenté dans AgroPortal, nous allons maintenant amorcer un processus d'enrichissement des métadonnées déjà présentes en partenariat avec nos utilisateurs. Ensuite, nous envisageons de faire une étude qualitative pour mesurer l'efficacité de la valorisation de ces métadonnées pour la sélection d'ontologies. Une autre de nos perspectives est de travailler avec l'équipe de développement de Protégé pour intégrer une partie des propriétés de notre liste. Cela permettrait aux développeurs d'éditer au même moment les métadonnées et le contenu de leur ontologie. Nous nous intéressons également au projet CEDAR (http://metadatacenter.org) qui travaille sur la prédiction et l'édition de métadonnées.

Remerciements. Ce travail a pu être réalisé grâce au soutien du projet Semantic Indexing of French biomedical Resources (SIFR – www.lirmm.fr/sifr) financé par l'Agence Nationale de la Recherche française ANR-12-JS02-01001, le Labex NUMEV (ANR-10-LabX-20), l'Institut de Biologie Computationnelle de Montpellier (ANR-11-BINF-0002), l'Université de Montpellier, et le CNRS. Nous remercions également le National Center for Biomedical Ontology (NCBO).

Bibliographie

[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*, pp. 164–176. 2011.

[2] C. Bizer, T. Heath, and T. Berners-Lee. Linked Data - The Story So Far. *Semantic Web and Information Systems*, 5(3):1–22, 2009.

[3] W. Ceusters. An information artifact ontology perspective on data collections and associated representational artifacts. In J. M. et al., editor, *24th International Conference of the European Federation for Medical Informatics, MIE'12*, pp. 68–72, Pisa, 2012.

[4] R. G. Côté, P. Jones, R. Apweiler, and H. Hermjakob. The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries. *BMC Bioinformatics*, 7(97):7, Feb. 2006.

[5] M. d'Aquin and N. F. Noy. Where to Publish and Find Ontologies? A Survey of Ontology Libraries. *Web Semantics*, 11:96–111, March 2012.

[6] L. Ding, T. Finin, A. Joshi, Y. Peng, R. S. Cost, J. Sachs, R. Pan, P. Reddivari, and V. Doshi. Swoogle: A Semantic Web Search and Metadata Engine. In, *13th ACM Conference on Information and Knowledge Management*, Washington DC, Nov 2004.

[7] M. Dumontier, C. J. Baker, J. Baran, et al., The Semanticscience Integrated Ontology for biomedical research and knowledge discovery. *Biomedical Semantics*, 5(14):11, 2014.

[8] B. Dutta, D. Nandini, and G. Kishore. MOD: Metadata for Ontology Description and Publication. In *Int. Conf. on Dublin Core & Metadata Applications,* pp. 1–9, Sao Paulo, Brazil, Sept. 2015.

[9] J. Hartmann, Y. Sure, P. Haase, and M. Suarez-Figueroa. OMV–ontology metadata vocabulary. *Workshop on Ontology Patterns for the Semantic Web,* Galway, Nov. 2005.

[10] R. Hoehndorf, L. Slater, P. N. Schofield, and G. V. Gkoutos. Aber-OWL: a framework for ontology-based data access in biology. *BMC Bioinformatics*, 16(1):1–9, 2015.

[11] C. Jonquet, E. Dzalé-Yeumo, E. Arnaud, and P. Larmande. AgroPortal: a proposition for ontology-based services in the agronomic domain. *3ème atelier INtégration de sources/masses de données hétérogènes et Ontologies, dans le domaine des sciences du VIVant et de l'Environnement, IN-OVIVE'15*, page 5, Rennes, June 2015.

[12] N. Juty, N. L. Novère, C. Laibe. Identifiers.org/MIRIAM Registry: community resources to provide persistent identification. *Nucleic Acids Research*, 40(D1):580–586, 2011.

[13] 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,* Madrid, 2009.

[14] E. Makela. Aether – Generating and Viewing Extended VoID Statistical Descriptions of RDF Datasets. In *11th Extended Semantic Web Conference, ESWC'14, demo session*, vol. 8798 of *LNCS*, pp. 429–433, Crete, Greece, May 2014. Springer.

[15] J. Malone, R. Stevens, S. Jupp, T. Hancocks, H. Parkinson, and C. Brooksbank. Ten Simple Rules for Selecting a Bio-ontology. *PLOS Computational Biology*, 12(2):6, 2016.

[16] D. Naskar. Ontology and Ontology Libraries: A critical study. Master thesis, Indian Statistical Institute, Bangalore, India, July 2014.

[17] N. F. Noy, N. H. Shah, P. L. Whetzel, M. Dorf, C. Jonquet, D. L. Rubin, M.-A. Storey, C. G. Chute, and M. A. Musen. BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research*, 37((web server)):170–173, May 2009.

[18] J. Park, S. Oh, and J. Ahn. Ontology selection ranking model for knowledge reuse. *Expert Systems with Applications*, 38(5):5133–5144, 2011.

[19] M. Sabou, V. Lopez, and E. Motta. Ontology Selection on the Real Semantic Web: How to Cover the Queens Birthday Dinner? *15th International Conference on Knowledge Engineering and Knowledge, EKAW'06*, LNCS 4248, pp. 96–111, Podebrady, 2006.

[20] B. Smith, M. Ashburner, C. Rosse, et al., The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, 25(11):1251–1255, Nov. 2007.

[21] 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.

[22] P.-Y. Vandenbussche, G. A. Atemezing, M. Poveda-Villalón, and B. Vatant. Linked Open Vocabularies (LOV): a gateway to reusable semantic vocabularies on the Web. *Semantic Web*, 2014.

[23] P.-Y. Vandenbussche and B. Vatant. Metadata recommendations for linked open data vocabularies. Report 1.1, Mondeca, 2012.

[24] Z. Xiang, C. Mungall, A. Ruttenberg, and Y. He. Ontobee: A Linked Data Server and Browser for Ontology Terms. In, *2nd International Conference on Biomedical Ontology, ICBO'11*, 833 of *CEUR Workshop* , page 3, Buffalo, NY, USA, July 2011.

[25] I. Mougenot, J-C. Desconnets and H. Chahdi, A DCAP to Promote Easy-to-Use Data for Multiresolution and Multitemporal Satellite Imagery Analysis, Int. Conf. on Dublin Core and Metadata Applications, DC'15, pp. 10-19, Sao Paulo, Brazil, Sept. 2015.