Challenges for ontology repositories and applications to biomedicine & agronomy

Clement Jonquet

Laboratory of Informatics, Robotics, and Microelectronics of Montpellier (LIRMM), University of Montpellier & CNRS, France &
Center for BioMedical Informatics Research (BMIR), Stanford University, USA
jonquet@lirmm.fr
(ORCID: 0000-0002-2404-1582)

Abstract

The explosion of the number of ontologies and vocabularies available in the Semantic Web makes ontology libraries and repositories mandatory to find and use them. Their functionalities span from simple ontology listing with more or less of metadata description to portals with advanced ontology-based services: browse, search, visualization, metrics, annotation, etc. Ontology libraries and repositories are usually developed to address certain needs and communities. BioPortal, the ontology repository built by the US National Center for Biomedical Ontologies BioPortal relies on a domain independent technology already reused in several projects from biomedicine to agronomy and earth sciences. In this position paper, we describe six high level challenges for ontology repositories: metadata & selection, multilingualism, alignment, new generic ontology-based services, annotations & linked data, and interoperability & scalability. Then, we present some propositions to address these challenges and point to our previously published work and results obtained within applications –reusing NCBO technology– to biomedicine and agronomy in the context of the NCBO, SIFR and AgroPortal projects.

Keywords

Ontologies, ontology libraries & repositories, ontology metadata, ontology-based services, ontology selection, semantic annotation, BioPortal.

1 Introduction

The Semantic Web produces many vocabularies and ontologies to represent and annotate any kind of data. However, those ontologies are spread out, in different formats, of different size, with different structures and from overlapping domains. The scientific community has always been interested in designing common platforms to list and sometime host and serve ontologies, align them, and enable their (re)use (Ding and Fensel, 2001; Hartmann et al., 2009; D'Aquin and Noy, 2012; , 1995). These platforms range from simple ontology listings or libraries with structured metadata, to advanced repositories (or portals) which feature a variety of services for multiple types of semantic resources (ontologies, vocabularies, terminologies, taxonomies, thesaurus) such as browse/search, visualization, metrics, recommendation, or annotation. In this paper, we will focus on ontology repositories, they allow to address important questions:

- If you have built an ontology, how do you let the world know and share it?
- How do you connect your ontology to the rest of the semantic world?
- If you need an ontology, where do you go to get it?
- How do you know whether an ontology is any good?
- If you have data to index, how do you find the most appropriate ontology for your data?
- If you look for data, how may the semantics of ontologies help you locate them?

More generally, ontology repositories help "ontology users" to deal with ontologies without managing them or engaging in the complex and long process of developing them.

However, with big number of ontologies new problems have raised such as describing, selecting, evaluating, trusting, and interconnecting them. From our experience working first on the US National Center for Biomedical Ontologies (NBCO) BioPortal, the most widely adopted biomedical ontology repository and later on the SIFR BioPortal, a specific sub-portal to address the French biomedical community and AgroPortal, an ontology repository for agronomy, we review and discuss six challenges in designing such platforms:

1. **Metadata & selection.** Ultimately, ontology repositories are made to share and reuse ontologies. But which ontology should I reuse? With too many different and overlapping ontologies, properly describing them with metadata and facilitate their identification and selection becomes and important issue. We believe, as any other data, ontologies must be FAIR.
2. **Multilingualism.** We live in a multilingual world, so are the concepts and entities from this world. The Semantic Web offers now tools and standards to develop multilingual and lexically rich ontologies. Repositories must be able to deal with multiple languages also.
3. **Ontology alignment.** No conceptualization is an island. It is now commonly agreed data interoperability cannot be achieved by means of a single common ontology for a domain, and interlinking ontologies is the way forward. But the more ontologies are being produced, the more the need for ontology alignment becomes important.
4. **Ontology-based services.** On reason to adopt Semantic Web standards and use ontology repositories is to benefit from multiple services for –and based on– ontologies. No one likes to reimplement something already existing and that can be generalized to another ontology just by dropping it in a repository. The portfolio of services for ontologies available in repositories should then grows.
5. **Annotations and linked data.** Ontologies and vocabularies are the backbone of semantically rich data (Linked Open Data, knowledge bases, etc.) as they are used to semantically annotate and interlink datasets. It is also important to facilitate semantic indexing, search and data access directly from the repositories.
6. **Scalability & interoperability.** The community of ontology developers and users is growing both horizontally (i.e., new domains) and vertically (i.e., new adopters inside a domain). Ontology repositories shall therefore scale to high number of ontologies, while facilitating their alignments, and when multiple repositories are created, they must be interoperable.

In the following, we will detail these challenges and briefly describe/point to results obtained in the context of our multiple ontology repository projects. In some sense, this article is an index of 10-years of published research in the domain of ontology repositories. We do not report hereafter all related work for each challenge neither we claim to have addressed them all. However, we believe our results illustrate potential solutions to move forward in that domain of research.

2 Background

2.1 Ontology libraries & repositories

With the growing number of ontologies developed, ontology libraries and repositories have always been of interest in the Semantic Web community. Ding and Fensel (2001) introduced the notion of *ontology library* and presented a review of libraries at that time:

> "A library system that offers various functions for managing, adapting and standardizing groups of ontologies. It should fulfill the needs for re-use of ontologies. In this sense, an ontology library system should be easily accessible and offer efficient support for re-using existing relevant ontologies and standardizing them based on upper-level ontologies and ontology representation languages."

The terms "collection", "listing" or "registries" are also used to describe ontology libraries. All correspond to systems that help reuse or find ontologies by simply listing them (e.g., DAML or DERI listings) or by offering structured metadata to describe them (e.g., FAIRSharing, BARTOC). But those systems do not support any services beyond description, especially based on the content of the ontologies.

Hartmann et al., (2009) introduced the concept of *ontology repository*, with advanced features such as search, browsing, metadata management, visualization, personalization, and mappings and an application programming interface to query their content/services:

> "A structured collection of ontologies (…) by using an Ontology Metadata Vocabulary. References and relations between ontologies and their modules build the semantic model of an ontology repository. Access to resources is realized through semantically-enabled interfaces applicable for humans and machines. Therefore, a repository provides a formal query language."

By the end of the 2000's, the topic was of high interest as illustrated by the 2010 ORES workshop (d'Aquin et al., 2010) or the 2008 OntologySummit.[1] The Open Ontology Repository Initiative (Baclawski and Schneider, 2009) was a collaborative effort to develop a federated infrastructure of ontology repositories. At that time, the effort already reused the NCBO technology (Whetzel and Team, 2013) that was the most advanced open source technology for managing ontologies but not yet packaged in an "virtual appliance" as it is today. More recently the effort also studied OntoHub (Till et al., 2014) technology for generalization but the OOR initiative is now discontinued.

In parallel, there have been effort do index any Semantic Web data online (including ontologies) and offer search engines such as Swoogle and Watson (Ding et al., 2004; D'Aquin et al., 2007). We cannot talk about ontology library or repositories for those "Semantic Web indexes", even if they support some features of ontology libraries or repositories (e.g., search).

In the biomedical or agronomic domains there are several standards and/or ontology libraries such as FAIRSharing (fairsharing.org) (McQuilton et al., 2016), the FAO's VEST Registry (aims.fao.org/vest-registry), and the agINFRA linked data vocabularies (vocabularies.aginfra.eu). They usually register ontologies and provide a few metadata attributes about them. However, because they are registries not especially focused on vocabularies and ontologies, they do not support the level of features that an ontology repository offers. In the biomedical domain, the OBO Foundry (Smith et al., 2007) is a reference community effort to help the biomedical and biological communities build their ontologies with an enforcement of design and reuse principles that have made the effort very successful. The OBO Foundry Web application (http://obofoundry.org) is not an ontology repository per se, but relies on other applications that pull their data from the foundry, such as the NCBO BioPortal (Noy et al., 2009), OntoBee (Ong et al., 2016), the EBI Ontology Lookup Service (Côté et al., 2006) and more recently AberOWL (Hoehndorf et al., 2015). In addition, there exist other ontology libraries and repository efforts unrelated to biomedicine, such as the Linked Open Vocabularies (Vandenbussche et al., 2014), OntoHub (Till et al., 2014), and the Marine Metadata Initiative's Ontology Registry and Repository (Rueda et al., 2009). More recently, the SIFR BioPortal (Jonquet et al., 2016a) prototype was created at University of Montpellier to build a French Annotator and experiment multilingual issues in BioPortal (Jonquet et al., 2015). The same university is also developing AgroPortal, an ontology repository for agronomy and neighboring domains such as food, plant sciences and biodiversity (Jonquet et al., 2017a).

D'Aquin and Noy, (2012) and Naskar and Dutta, (2016) provided the latest reviews of ontology repositories. In Table 1, we provide a non-exhaustive –but quite rich– list of ontology libraries, repositories and Web indexes available today.

Ontology libraries
OBO Foundry
WebProtégé
Romulus
DAML ontology library
Colore
VEST/AgroPortal Map of standards
FAIRsharing
DERI Vocabularies
OntologyDesignPatterns
SemanticWeb.org
W3C Good ontologies
TaxoBank
BARTOC
GFBio Terminology Service
agINFRA Linked Data Vocabularies
oeGOV
Ontology repositories
NCBO BioPortal*
Ontobee
EBI Ontology Lookup Service
AberOWL
CISMEF HeTOP
SIFR BioPortal*
OKFN Linked Open Vocabularies
ONKI Ontology Library Service
MMI Ontology Registry and Repository*

[1] http://ontolog.cim3.net/wiki/OntologySummit2008.html

ESIPportal*
AgroPortal*
OntoHub
Finto
EcoPortal (proposition end 2017)*
Semantic Web indexes
Swoogle
Watson
Sindice
Falcons
Technology
NCBO Virtual Appliance (Stanford)
OLS technology (EBI)
LexEVS (Mayo Clinic)
Intelligent Topic Manager (Mondeca)
SKOSMOS (Nat. Library of Finland)
Abandoned projects include: Cupboard, Knoodl, Schemapedia, SchemaWeb, OntoSelect, OntoSearch, OntoSearch2, TONES, SchemaCache, Soboleo

Table 1. Non-exhaustive list of ontology libraries, repositories and Web indexes available today. We also included some known “technology” that can be reused to setup an ontology library. Blue cells are projects in biomedicine and health sciences. A * identifies ontology repositories which reuse(d) NCBO technology.

2.2 Focus on the NCBO BioPortal: a “one stop shop” for biomedical ontologies

In the biomedical domain, BioPortal (http://bioportal.bioontology.org) (Noy et al., 2009), developed by the National Center for Biomedical Ontologies (NBCO) at Stanford is a well-known open repository for biomedical ontologies originally spread out over the Web and in different formats. There are +650 public ontologies in this collection as of end 2017. By using the portal’s features, users can browse, search, visualize and comment on ontologies both interactively through a Web interface, and programmatically via Web services. Within BioPortal, ontologies are used to develop an annotation workflow (Jonquet et al., 2009) that indexes several biomedical text and data resources using the knowledge formalized in ontologies to provide semantic search features that enhance information retrieval experience (Jonquet et al., 2011). The NCBO BioPortal functionalities have been progressively extended in the last 12 years, and the platform has adopted Semantic Web technologies (e.g., ontologies, mappings, metadata, notes, and projects are stored in an RDF[2] triple store) (Salvadores et al., 2013).

An important aspect is that NCBO technology (Whetzel and Team, 2013) is domain-independent and open source. A BioPortal virtual appliance[3] is available as a server machine embedding the complete code and deployment environment, allowing anyone to set up a local ontology repository and customize it. The NCBO virtual appliance is quite regularly reused by organizations which need to use services like the NCBO Annotator but have to process sensitive data in house e.g., hospitals. Via the virtual appliance, NCBO technology has already been adopted for different ontology repositories in related domains and was also originally chosen as foundational software of the OOR Initiative (Baclawski and Schneider, 2009). The MMI Ontology Registry and Repository (Rueda et al., 2009) used it as its backend storage system for over 10 years, and the Earth Sciences Information Partnership earth and environmental semantic portal (Pouchard L. Huhns M., 2012) was deployed several years ago. We are also currently working on the SIFR BioPortal (Jonquet et al., 2016a) and AgroPortal (Jonquet et al., 2017a) projects described hereafter.

2.3 Two collaborative ontology repository projects

In the context of our projects, to avoid building new ontology repositories from scratch, we have considered which of the previous technologies are reusable. While most of them are “open source,” only the NCBO BioPortal[4] and OLS[5] are really meant for reuse, both in their construction, and with their documentation provided. Although we

[2] The Resource Description Framework (RDF) is the W3C language to described data. It is the backbone of the semantic web. SPARQL is the corresponding query language. By adopting RDF as the underlying format, an ontology repository based on NCBO technology can easily make its data available as linked open data and queryable through a public SPARQL endpoint. To illustrate this, the reader may consult the Link Open Data cloud diagram (http://lod-cloud.net) that since 2017 includes ontologies imported from the NCBO BioPortal (most of the Life Sciences section).

[3] www.bioontology.org/wiki/index.php/Category:NCBO_Virtual_Appliance

[4] The technology has always been open source, and the appliance has been made available since 2011. However, the product became concretely and easily reusable after BioPortal v4.0 end of 2013.

[5] The technology has always been open source but some significant changes (e.g., the parsing of OWL) facilitating the reuse of the technology for other portals were done with OLS 3.0 released in December 2015.

cannot know all the applications of other technologies, the visibly frequent reuse of the NCBO technology definitively confirmed it is a good candidate for reuse when building a new ontology repository. Also, of the two candidate technologies, we believe NCBO technology implements the highest number of required features in our projects (Jonquet et al., 2017a).

SIFR BioPortal

In the context of the Semantic Indexing of French Biomedical Data Resources (SIFR) project, we have developed the SIFR BioPortal (http://bioportal.lirmm.fr) (Jonquet et al., 2016a), an open platform to host French biomedical ontologies and terminologies based on the technology developed by the NCBO. The portal facilitates use and fostering of terminologies and ontologies which were only developed in French or translated from English resources and are not well served in the English-focused NCBO BioPortal. As of today, the portal contains 25 public ontologies and terminologies (+ 6 private ones) that cover multiple areas of biomedicine, such as the French versions of standards terminologies (e.g., MeSH, MedDRA, ATC, ICD-10) but also multilingual ontologies. In this later cases, we use the NCBO BioPortal as a source repository –so users do not have to upload their multilingual ontologies twice– and only parse and index the French content on the SIFR BioPortal.

The original motivation in building the SIFR BioPortal was to develop the SIFR Annotator (http://bioportal.lirmm.fr/annotator) to address the lack of out-of-the-shelve openly and easily accessible semantic annotation system for French (Jonquet et al., 2016a; Tchechmedjiev et al., 2017a). The service is originally based on the NCBO Annotator [8], a Web service allowing scientists to utilize available biomedical ontologies for annotating their datasets automatically, but was significantly enhanced and customized for French. The annotator service processes raw textual descriptions, tags them with relevant biomedical ontology concepts and returns the annotations to the users in several formats such as JSON-LD, RDF or BRAT.

AgroPortal: a vocabulary and ontology repository for agronomy

We have been reusing the NCBO BioPortal technology to design AgroPortal, an ontology repository for agronomy, food, plant sciences, and biodiversity (http://agroportal.lirmm.fr) (Jonquet et al., 2016c; Jonquet et al., 2017a). AgroPortal, is an advanced prototype featuring all BioPortal services and new ones implemented to address the requirements of the agronomy community. The platform currently hosts 77 ontologies among which 50 are not present in any comparable repository. We have identified 93 other candidate ontologies that will be loaded in the future to complement this valuable resource.

3 Challenges, propositions and results

In the following sections, we describe some challenges we identified by working on ontology repository and exchanging with our user communities. In each case, we describe a few results obtained on the relevant topic.

3.1 Metadata & selection

The first questions we ask ourselves when entering a bookstore are often: “Where is the book I am looking for?” or “Which book will I discover and pick up today?” The same questions are true for ontology libraries. To address them, **we need better description of the ontologies, with precise and harmonized metadata and we need also means to facilitate the identification and selection of the ontologies of interest.** Ontologies serve to make data FAIR (Wilkinson et al., 2016), ontology repositories shall serve to make ontologies FAIR.

As any resources, ontologies, vocabularies and terminologies need to be described with relevant metadata to facilitate their identification and selection. However, none of the existing metadata vocabularies can completely meet this need if taken independently. Indeed, some metadata properties are intrinsic to the ontology (name, license, description); others, such as community feedbacks, or relations to other ontologies are typically information that an ontology library shall capture, populate and consolidate to facilitate the ontology landscape comprehension (e.g., selection of an ontology).

In Jonquet et al., (2017b), we have reviewed the most standard and relevant vocabularies (23 totals) currently available to describe metadata for ontologies (such as Dublin Core, Ontology Metadata Vocabulary, VoID, etc.) as well as the different metadata implementation in multiple ontology libraries or repositories. We have then built a new metadata model for AgroPortal. The repository now parses 346 standard properties that could be used to describe different aspects of ontologies: intrinsic descriptions, people, date, relations, content, metrics, community, administration, and access. We use them to populate a model of 127 properties implemented in the portal and harmonized for all the ontologies. We have 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 by displaying diagrams and charts about all the ontologies on the portal. We have now a specific page (http://agroportal.lirmm.fr/landscape) dedicated to visualizing the ontology landscape in AgroPortal that facilitates analysis of the repository content. The landscape page helps to figure out what are some of the main domain of interests as well as common development practices when creating an ontology in agronomy.

In Dutta et al., (2017), we have generalized our work done within AgroPortal to propose a new Metadata vocabulary for Ontology Description and publication, called MOD (https://github.com/sifrproject/MOD-Ontology). MOD 1.2 is defined in OWL and consists of 19 classes and 88 properties most of them to describe the mod:Ontology object. MOD 1.2 may serve as (i) a vocabulary to be used by ontology developers to annotate and describe their ontologies, or (ii) an explicit OWL ontology to be used by ontology libraries to offer semantic descriptions of ontologies as linked data. MOD 1.2 is an initiative which attempts to overcome some of the limitations of the Ontology Metadata Vocabulary (Suarez-Figueroa et al., 2005) but is still a temporary proposition that will be discussed in the next months within the Research Data Alliance recently re-configured Vocabulary & Semantic Services Interest Group.[6]

Automatic ontology selection or recommendation has been a subject of interest to facilitate ontology reuse (Sabou et al., 2006)(Butt et al., 2016). The number and variety of ontologies in certain domains is now so large that choosing one for an annotation task or for designing a specific application is quite cumbersome.

In Martinez-Romero et al., (2017), we developed the NCBO Ontology Recommender. This service suggests relevant ontologies from the repository for annotating text data. The new recommendation approach evaluates the relevance of an ontology to biomedical text data according to four different criteria: (1) the extent to which the ontology covers the input data; (2) the acceptance of the ontology in the community; (3) the level of detail of the ontology classes that cover the input data; and (4) the specialization of the ontology to the domain of the input data. This new version of a service originally released in 2010 (Jonquet et al., 2010) combines the strengths of its predecessor with a range of adjustments and new features that improve its reliability and usefulness. Because it is integrated in the NCBO technology, the Recommender is already available within the SIFR BioPortal and AgroPortal. We shall note that these services do not yet rely on the new metadata model previously cited.

3.2 Multilingualism

Scientific discoveries that could be made with help of ontologies to annotate, integrate, mine and search data, are often limited by the availability of ontology-based tools and services only for one natural language, usually English, for which there exist the most ontologies. Recently, ontology localization, i.e., "the process of adapting an ontology to a concrete language and culture community" (Cimiano et al., 2010), has become very important in the ontology development lifecycle, but when efforts are made to properly represent lexical (e.g., using Lemon (McCrae et al., 2011)) or multilingual information (e.g., using LexOMV (Montiel-Ponsoda et al., 2007) or Lemon translation module (Gracia et al., 2014)) are made, it is rarely leveraged by ontology libraries and repositories. In the future, **we need ontology repositories to entirely support interface and content internationalization** (i.e., both displaying user interfaces (e.g., menu names, help, etc.) in different languages and displaying their content (e.g., ontology labels, mappings, etc.) in different languages) **and be multilingual by enabling a complete use of their functionalities and services for multilingual ontologies or monolingual ontologies** linked one another.

In Jonquet et al., (2015), we presented a roadmap for addressing the issues of dealing with multilingual or monolingual ontologies in the NCBO BioPortal, which takes English as primary language. We proposed a set of representations to support multilingualism in the portal and to enable a complete use of the functionalities and services

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

for any kind of ontologies and data: (i) Representation of natural language property for an ontology; (ii) Representation of translation relations between ontologies; (iii) Representation of the distinction between ontologies with multilingual content i.e., multilingual and mono lingual ontologies; (iv) Representation of multilingual mappings. Those aspects have been addressed now within MOD and/or the new AgroPortal metadata model previously cited. In addition, in Annane et al., (2016b), we reconciled more than 228K mappings between ten English ontologies hosted on NCBO BioPortal and their French translations hosted on the SIFR BioPortal. The next big step is now to internationalize the portal.

3.3 Ontology alignment

Ontologies, or other semantic resources, will inevitably overlap in coverage. Therefore, the need for ontology alignment. This need has been explicitly expressed by almost all our partner organizations in biomedicine, agronomy or ecology. Surprisingly, it seems there is a gap between the state-of-the-art results obtained at each edition of the Ontology Alignment Evaluation Initiative (OAEI - http://oaei.ontologymatching.org) and the day-to-day reality of ontology developers. Tools are often hardly reusable, and results cannot be easily reproduced outside of the benchmarking effort. Another key role of ontology repositories is to store mappings (or alignments) between ontologies. **Ontology repositories shall support the extraction, generation, validation, evaluation, storage and retrieval of mappings between the ontologies they host**. Automatic mapping generation within ontology repositories shall go beyond simple lexical or ID-based approaches[7] and state-of-the-art tools shall be incorporated within repositories. An equivalent effort, such as the one made to harvest ontologies, must be made to harvest the mappings between these ontologies and describe them with metadata and provenance information to facilitate trust and reuse.

In Ghazvinian et al., (2009), we have analyzed the mappings automatically generated within BioPortal and what they tell us about the ontologies themselves, the structure of the ontology repository, and the ways in which the mappings can help in the process of ontology design and evaluation. This study demonstrated the value of having a mapping repository goes beyond ontology-to-ontology alignment, but concretely helps analyze the structures, dependencies and overlap of ontologies in the same domain. A similar, more recent study about ontology terms reuse have been done by Kamdar et al. (Kamdar et al., 2017). In Annane et al., (2016a), we have also demonstrated that existing mappings between ontologies can also be used to improve ontology alignment methods based on background knowledge; in other words, a centralized mapping repository will also be an excellent resource to curate and generate new mappings.

3.4 Generic ontology-based services

Ontology repositories offer a large span of services: file hosting, versioning, search and browse content, visualization, metrics, notes, mapping, etc. These services are ‘generic’ if they are domain independent i.e., not specific to a domain, group of ontologies, specific format or design principles. It is important that **ontology repositories continue to enhance ontology-based services and offer new generic ones to enlarge the spectrum of possible use of ontologies**. Using standard formats such as OWL or SKOS has facilitated the development of a wide range of tools and services for semantic resources. The challenge is now to package them inside ontology repositories and keep vertical quality (i.e., one ontology) while enabling quantitative horizontal use.

One important use of ontologies is for annotating and indexing text data (Spasic et al., 2005; Handschuh and Staab, 2003). Therefore, we often see aside of ontology repositories, ontology-based annotation services. For instances, BioPortal has the NCBO Annotator (Jonquet et al., 2009), OLS had Whatizit (Rebholz-Schuhmann et al., 2008) and now moved to ZOOMA, HeTOP had FMTI (Sakji et al., 2010) and UMLS has MetaMap (Aronson, 2001). Hereafter, we focus on services for text data (annotation & terminology extraction).

In Lossio-Ventura et al., (2014), we presented BioTex, a Web application that implements state-

[7] To the best of our knowledge, only the NCBO technology automatically computes ontology alignments when ontologies are hosted within the portal. The portal automatically creates some mappings when two classes share the same identifiers properties, or when they share a common normalized preferred label or synonym. Although basic lexical mapping approaches can be inaccurate and should be used with caution (Faria et al., 2014; Pathak and Chute, 2009), they usually work quite well to interconnect ontologies (Ghazvinian et al., 2009).

of-the-art measures for automatic extraction of biomedical terms from English and French free text. The application includes a new methodology for automatic term extraction mixing linguistic, statistical, graph and Web-based approaches that have been demonstrated quite efficient (Lossio-Ventura et al., 2015). Among other use of BioTex, we have shown it can be part of an ontology enrichment workflow that could be highly valuable for ontology developers (Lossio-Ventura et al., 2016). However, this work has not yet been incorporated within an ontology repository technology.

In Tchechmedjiev et al.,(2017), we present multiple enhancement to the semantic annotation workflow that we have developed on top of the NCBO Annotator and when building a French version of the service. Some of these new functionalities are particularly relevant to process electronic health records. These new features include: annotation scoring (Melzi and Jonquet, 2014), additional output formats (for evaluation and integration with standard clinical systems), clinical context detection (negation, experiencer and temporality through the integration of the NegEx/ConText algorithm) (Abdaoui et al., 2017), coarse-grained entity type annotations (using UMLS Semantic Groups, e.g., anatomy, disorders, devices).

3.5 Annotations and Linked Data

Data integration and semantic interoperability enable new scientific discoveries that could be made by merging different currently available data. These is one major reason for adopting ontologies. They are used to design semantic indexes of data and linked open datasets that could be used for various type of cross datasets studies (Handschuh and Staab, 2003; Bizer et al., 2009). **Ontology repositories must facilitate indexing/annotation, search and access to semantically described, interoperable, actionable, open, rich linked data directly from the within the repositories.** Working with big data represents a set of challenges for ontology repositories when designing these semantic indexes: scalability, consistency, completeness in a context where both ontologies and data constantly evolve. In addition, cross ontologies semantics and indexed data consistency shall be checked by ontology repositories using OWL reasoning.

In Jonquet et al., (2011), we have built the NCBO Resource Index, an ontology-based index of more than twenty heterogeneous biomedical resources (later extended to 50) included within BioPortal. Directly when browsing the ontologies or using a dedicated search engine, users can discover datasets of interest. The indexing relied on the NCBO Annotator workflow and used the semantics that the ontologies encode, such as synonyms, class hierarchies, and the mappings between ontologies, to improve the search experience. The Resource Index, was a tentative developed before 2010 that did not rely neither on big data technologies and did not followed linked open data principles. Both were in their infancies at that time. More recently, in agronomy, we have followed new efforts such as AgroLD project (Venkatesan et al., 2015) to build a database of resources described in RDF, and annotated with ontologies. We are currently working on the interoperation of AgroLD and AgroPortal.

3.6 Scalability & interoperability

In 2007, Swoogle claimed to “Search over 10.000 ontologies”. Today, a simple Google Search for “filetype:owl” returns around 34K results. The NCBO BioPortal, which is generally considered has the biggest ontology repository (not library) contains +650 ontologies as of end of 2017. More and more vocabularies are being developed and hosted by the LOV platform. Multiple domain specific ontology repository efforts have started often inspired by results in the biomedical domain and usually by reusing NCBO technology (e.g., MMI OOR, AgroPortal, ESIPPortal). **The more ontologies and ontology repositories are being developed, the more scalability and interoperability issues become important**. Some ontologies are useful to different communities and shall then be hosted in multiple repositories e.g., domain ontologies such as the Gene Ontology (Ashburner et al., 2000), or the Environment Ontology (Buttigieg et al., 2013). Because no repository will host them all, ontology repositories have to offer a certain level of interoperability to ensure their users that they will not have to work with multiple web applications and programming interfaces if their ontologies of interest are not all hosted by the same repositories. As previously explained standard ontology metadata is a crucial aspect to achieve this.

In Jonquet et al., ; Jonquet et al., (2016b), our projects described Section 2.3, we have been particularly careful in not redeveloping features and functionalities that to our knowledge were already available. We have designed and implemented two advanced prototype ontology repositories for the French biomedical community and for the agronomy domain. Our choice to reuse the NCBO technology was justified by the large spectrum of features and services, but in addition our motivation was: (i) to avoid re-developing tools that have already been designed and extensively used and contribute to long term support of the commonly used technology; and (ii) to offer the same tools, services and formats to different but still interconnected communities, to facilitate the interface and interaction between their domains (agro, bio, health (French)). Relying on the same original technology enhance both technical reuse (for example, enabling queries to either systems with the same code), and semantic reuse. Then, we have developed new functionalities –as previously described– while keeping our systems backward compatible with the original technology to facilitate a convergence of the efforts. We strongly believe that sharing the technology is the best way to guaranty long term support and development by engaging different ontology practitioners and communities all around the world with their respective funding and supporting schemes. Also, sharing the technology is the best way to make ontology repositories interoperable. As explained in Tchechmedjiev et al.,(2017), all of the new features implemented (e.g., NCBO Annotator + or the new Recommender) are available across any other NCBO based platform at minimum cost.

4 Conclusions

In this paper, we have presented our vision on challenges and issues in building ontology repositories. We have illustrated our thoughts with results obtained over the last 10 years within our projects in biomedicine and agronomy. By adopting NCBO technology, we inherit some advantages and inconvenients but we can now contribute to this field of research with concrete use cases, communities and outcomes. NCBO-based ontology repositories adopted a vision where multiple semantic resources are made available in a common place (though not combined and consistency checked), and cast to a common model. While doing so, the repositories arguably limits the full power of ontologies –which has been a recurrent criticism– constraining their use to features supported by the common model. We see two general scenarios of use for these repositories:

- The repositories provide basic ontology library services for users with a “vertical need” —those who want to do very precise things (e.g., reasoning, using specific relations) using only suitable ontologies (developed by the same communities and in the same format). Such users may just use the repositories as libraries to find and download ontologies, and work in their own environment.
- The repositories provide many ontology-based services to users with “horizontal needs” —those who wants to work with a wide range of ontologies and vocabularies useful in their domain but developed by different communities, overlapping and in different formats. Such users greatly appreciate the unique endpoints (Web application and programmatic for REST and SPARQL queries) offered by the repositories under a simplified common model.

In this position paper, we have unfortunately not covered all related work on the cited challenges and we have certainly skipped other important challenges: semantic consistency, ontology evaluation, visualization, community feedback. But we offered a short summary of multiple various contributions on ontology repository and ontology-based service research. In the future, we will continue our efforts to address the identified challenges (and others), while continue to offer to various scientific communities the means to share and leverage their ontologies or semantic resources and enable new science in their fields.

Acknowledgments

This work is partly achieved within the Semantic Indexing of French biomedical Resources (SIFR – www.lirmm.fr/sifr) project that received funding from the French National Research Agency (grant ANR-12-JS02-01001), the European Union’s Horizon 2020 research and innovation 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 the University of Montpellier and the CNRS. I also acknowledge the National Center for Biomedical Ontologies for their insights and thanks all my collaborators in Montpellier or Stanford interested like me on ontology repositories.

References

Amine Abdaoui, Andon Tchechmedjiev, William Digan, Sandra Bringay, and Clement Jonquet. 2017. French ConText: a Publicly Accessible System for Detecting Negation, Temporality and Experiencer in French Clinical Notes. *Biomedical Informatics*, UNDER REVI. Under review.

Amina Annane, Zohra Bellahsene, Faical Azouaou, and Clement Jonquet. 2016a. Selection and Combination of Heterogeneous BK to Enhance Biomedical Ontology Matching. In *20th International Conference on Knowledge Engineering and Knowledge Management, EKAW'16*, volume 10024 LNAI, Bologna, Italy.

Amina Annane, Vincent Emonet, Clement Jonquet, Faical Azouaou, and Clement Jonquet. 2016b. Multilingual Mapping Reconciliation between English-French Biomedical Ontologies. In *6th International Conference on Web Intelligence, Mining and Semantics, WIMS'16*, volume 13–15–June, page 12, Nimes, France. ACM.

Alan R Aronson. 2001. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. In *American Medical Informatics Association Annual Symposium, AMIA'01*, pages 17–21, Washington, DC, USA.

Michael Ashburner, Catherine A Ball, Judith A Blake, David Botstein, Heather Butler, J Michael Cherry, Allan P Davis, Kara Dolinski, Selina S Dwight, Janan T Eppig, and others. 2000. Gene Ontology: tool for the unification of biology. *Nature genetics*, 25(1):25–29.

Kenneth Baclawski and Todd Schneider. 2009. The open ontology repository initiative: Requirements and research challenges. In T Tudorache, G Correndo, N Noy, H Alani, and M Greaves, editors, *Workshop on Collaborative Construction, Management and Linking of Structured Knowledge, CK'09*, volume 514, page 10, Washington, DC., USA. CEUR-WS.org.

Christian Bizer, Tom Heath, and Tim Berners-Lee. 2009. Linked Data - The Story So Far. *Semantic Web and Information Systems*, 5(3):1–22.

Anila Sahar Butt, Armin Haller, and Lexing Xie. 2016. RecOn: Ontology recommendation for structureless queries. *Applied Ontology*, 11(4):301–324.

Pier Luigi Buttigieg, Norman Morrison, Barry Smith, Christopher J Mungall, and Suzanna E and Lewis. 2013. The environment ontology: contextualising biological and biomedical entities. *Biomedical Semantics*, 4(1):43.

Philipp Cimiano, Elena Montiel-Ponsoda, Paul Buitelaar, Mauricio Espinoza, and Asuncion Gàmez-Pérez. 2010. A note on ontology localization. *Applied Ontology*, 5(2):127–137.

Richard G Côté, Philip Jones, Rolf Apweiler, and Henning Hermjakob. 2006. The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries. *BMC bioinformatics*, 7:97, January.

Mathieu D'Aquin, Claudio Baldassarre, Laurian Gridinoc, Sofia Angeletou, Marta Sabou, Enrico Motta, Mathieu Aquin, Claudio Baldassarre, Laurian Gridinoc, Sofia Angeletou, Marta Sabou, Enrico Motta, Mathieu D'Aquin, Claudio Baldassarre, Laurian Gridinoc, Sofia Angeletou, Marta Sabou, Enrico Motta, Mathieu Aquin, et al. 2007. Watson: A Gateway for Next Generation Semantic Web Applications. In *Poster & Demonstration Session at the 6th International Semantic Web Conference, ISWC'07*, page 3, Busan, Korea.

2010. 1st Workshop on Ontology Repositories and Editors for the Semantic Web, ORES'10. In Mathieu d'Aquin, Alexander G Castro, Christoph Lange, and Kim Viljanen, editors, number 596, Hersonissos, Greece. CEUR-WS.org.

Mathieu D'Aquin and Natasha F Noy. 2012. Where to Publish and Find Ontologies? A Survey of Ontology Libraries. *Web semantics*, 11(August):96–111, March.

Li Ding, Tim Finin, Anupam Joshi, Yun Peng, R Scott Cost, Joel Sachs, Rong Pan, Pavan Reddivari, and Vishal Doshi. 2004. Swoogle: A Semantic Web Search and Metadata Engine. In D A Grossman, L Gravano, C Zhai, O Herzog, and D Evans, editors, *13th ACM Conference on Information and Knowledge Management, CIKM'04*, Washington DC, USA. ACM.

Ying Ding and Dieter Fensel. 2001. Ontology Library Systems: The key to successful Ontology Re-use. In *1st Semantic Web Working Symposium, SWWS'01*, pages 93–112, Stanford, CA, USA. CEUR-WS.org.

Biswanath Dutta, Anne Toulet, Vincent Emonet, and Clement Jonquet. 2017. New Generation Metadata vocabulary for Ontology Description and Publication. In E Garoufallou, S Virkus, and G Alemu, editors, *11th Metadata and Semantics Research Conference, MTSR'17*, Tallinn, Estonia.

Daniel Faria, Ernesto Jiménez-Ruiz, Catia Pesquita, Emanuel Santos, and Francisco M Couto. 2014. Towards Annotating Potential Incoherences in BioPortal Mappings. In *13th International Semantic Web Conference, ISWC'13*, volume 8797, pages 17–32, Riva del Garda, Italy. Springer.

Amir Ghazvinian, Natasha F Noy, Clement Jonquet, Nigam H Shah, and Mark A Musen. 2009. What

Four Million Mappings Can Tell You about Two Hundred Ontologies. In A Bernstein, D R Karger, T Heath, L Feigenbaum, D Maynard, E Motta, and K Thirunarayan, editors, *8th International Semantic Web Conference, ISWC'09*, volume 5823, pages 229–242, Washington DC, USA. Springer.

Jorge Gracia, Elena Montiel-Ponsoda, Daniel Vila-Suero, and Guadalupe Aguado-de Cea. 2014. Enabling Language Resources to Expose Translations as Linked Data on the Web. In N Calzolari, K Choukri, T Declerck, H Loftsson, B Maegaard, J Mariani, A Moreno, J Odijk, and S Piperidis, editors, *9th International Conference on Language Resources and Evaluation, LREC'14*, pages 409–4013, Reykjavik, Iceland. European Language Resources Association.

Siegfried Handschuh and Stephen Staab, editors. 2003. *Annotation for the Semantic Web*.volume 96 of *Frontiers in Artificial Intelligence and Applications*. IOS Press.

Jens Hartmann, Raúl Palma, and Asunción Gómez-Pérez. 2009. Ontology Repositories. *Handbook on Ontologies*:551–571.

Robert Hoehndorf, Luke Slater, Paul N Schofield, and Georgios V Gkoutos. 2015. Aber-OWL: a framework for ontology-based data access in biology. *BMC Bioinformatics*, 16(1):1–9.

Clement Jonquet, Amina Annane, Khedidja Bouarech, Vincent Emonet, and Soumia Melzi. 2016a. SIFR BioPortal : Un portail ouvert et générique d'ontologies et de terminologies biomédicales françaises au service de l'annotation sémantique. In *16th Journées Francophones d'Informatique Médicale JFIM'16*.

Clement Jonquet, Amina Annane, Khedidja Bouarech, Vincent Emonet, and Soumia Melzi. 2016b. SIFR BioPortal : Un portail ouvert et générique d'ontologies et de terminologies biomédicales françaises au service de l'annotation sémantique. In *16th Journées Francophones d'Informatique Médicale, JFIM'16*.

Clement Jonquet, Vincent Emonet, and Mark A Musen. 2015. Roadmap for a multilingual BioPortal. In J Gracia, J P McCrae, and G Vulcu, editors, *4th Workshop on the Multilingual Semantic Web, MSW4'15*, volume 1532, pages 15–26, Portoroz, Slovenia.

Clement Jonquet, Paea LePendu, Sean Falconer, Adrien Coulet, Natalya F Noy, Mark A Musen, and Nigam H Shah. 2011. NCBO Resource Index: Ontology-Based Search and Mining of Biomedical Resources. *Web Semantics*, 9(3):316–324. 1st prize of Semantic Web Challenge at the 9th International Semantic Web Conference, ISWC'10, Shanghai, China.

Clement Jonquet, Mark A Musen, and Nigam H Shah. 2010. Building a Biomedical Ontology Recommender Web Service. *Biomedical Semantics*, 1(S1). Selected in Pr. R. Altman's 2011 Year in Review at AMIA TBI.

Clement Jonquet, Nigam H Shah, and Mark A Musen. 2009. The Open Biomedical Annotator. In *American Medical Informatics Association Symposium on Translational BioInformatics, AMIA-TBI'09*, pages 56–60, San Francisco, CA, USA.

Clément Jonquet, Anne Toulet, Elizabeth Arnaud, Sophie Aubin, Esther Dzalé-Yeumo, Vincent Emonet, John Graybeal, Mark A. Musen, Cyril Pommier, and Pierre Larmande. 2016c. Reusing the NCBO BioPortal technology for agronomy to build AgroPortal. In P Jaiswal and R Hoehndorf, editors, *7th International Conference on Biomedical Ontologies, ICBO'16, Demo Session*, volume 1747, pages 4–6, Corvallis, Oregon, USA.

Clement Jonquet, Anne Toulet, Elizabeth Arnaud, Sophie Aubin, Esther Dzalé Yeumo, Vincent Emonet, John Graybeal, Marie-Angélique Laporte, Mark A Musen, Valeria Pesce, and Pierre Larmande. 2017a. AgroPortal: an ontology repository for agronomy. *Computers and Electronics in Agriculture*, IN PRESS:30. Under review.

Clement Jonquet, Anne Toulet, Biswanath Dutta, and Vincent Emonet. 2017b. Harnessing the power of unified metadata in an ontology repository: the case of AgroPortal. *Data Semantics*, UNDER REVI. Under review.

Maulik R Kamdar, Tania Tudorache, and Mark A Musen. 2017. A Systematic Analysis of Term Reuse and Term Overlap across Biomedical Ontologies. *Semantic web*, 8(6):853–871.

Juan-Antonio Lossio-Ventura, Clement Jonquet, Mathieu Roche, Maguelonne Teisseire, Juan Antonio, Lossio-ventura Clement Jonquet, Mathieu Roche, and Maguelonne Teisseire. 2015. Biomedical term extraction: overview and a new methodology. *Information Retrieval, Special issue on Medical Information Retrieval*, 19(1):59–99.

Juan Antonio Lossio-Ventura, Clement Jonquet, Mathieu Roche, and Maguelonne Teisseire. 2016. A Way to Automatically Enrich Biomedical Ontologies. In *19th International Conference on Extending Database Technology, EDBT'16, Poster Session*, number 305, page 2, Bordeaux, France. OpenProceedings.org.

Juan Antonio Juan Antonio Lossio-Ventura, Clement Jonquet, Mathieu Roche, and Maguelonne Teisseire. 2014. BIOTEX: A system for Biomedical Terminology Extraction, Ranking, and Validation.

In M Horridge, M Rospocher, and J Ossenbruggen, editors, *13th International Semantic Web Conference, Demonstration, ISWC'14*, volume 1272, pages 157–160, Riva del Garda, Italy.

Marcos Martinez-Romero, Clement Jonquet, Martin J O'Connor, John Graybeal, Alejandro Pazos, and Mark A Musen. 2017. NCBO Ontology Recommender 2.0: An Enhanced Approach for Biomedical Ontology Recommendation. *Biomedical Semantics*, 8(21).

John McCrae, Dennis Spohr, and Philipp Cimiano. 2011. Linking lexical resources and ontologies on the semantic web with lemon. In G Antoniou, M Grobelnik, E Simperl, B Parsia, D Plexousakis, P DeLeenheer, and J Z Pan, editors, *8th Extended Semantic Web Conference, ESWC'11*, number 6643, pages 245–259, Heraklion, Crete, Greece. Springer.

Peter McQuilton, Alejandra Gonzalez-Beltran, Philippe Rocca-Serra, Milo Thurston, Allyson Lister, Eamonn Maguire, and Susanna-Assunta Sansone. 2016. BioSharing: curated and crowd-sourced metadata standards, databases and data policies in the life sciences. *Database*, baw075.

Soumia Melzi and Clement Jonquet. 2014. Scoring semantic annotations returned by the NCBO Annotator. In A Paschke, A Burger, P Romano, M S Marshall, and A Splendiani, editors, *7th International Semantic Web Applications and Tools for Life Sciences, SWAT4LS'14*, volume 1320, page 15, Berlin, Germany. CEUR-WS.org.

Elena Montiel-Ponsoda, Guadalupe Aguado de Cea, Mari Carmen Suarez-Figueroa, Raul Palma, Asuncion Gomez-Pérez, and Wim Peters. 2007. LexOMV: an OMV extension to capture multilinguality. In P Buitelaar, K Choi, A Gangemi, C Huang, and A Oltramari, editors, *Lexicon/Ontology Interface Workshop, OntoLex'07*, page 10, Busan, South-Korea.

Debashis Naskar and Biswanath Dutta. 2016. Ontology And Ontology Libraries: A Study From An Ontofier And An Ontologist Perspective. In *19th International Symposium on Electronic Theses and Dissertations, ETD'16*, pages 1–12, Lille, France.

Natalya F Noy, Nigam H Shah, Patricia L Whetzel, Benjamin Dai, Michael Dorf, Nicholas B Griffith, Clement Jonquet, Daniel L Rubin, Margaret-Anne Storey, Christopher G Chute, Mark A Musen, Cherie H Youn, Adrien Coulet, Chris Callendar, and Barry Smith. 2009. BioPortal: ontologies and integrated data resources at the click of a mouse. In B Smith, editor, *Nucleic Acids Research*, volume 37, pages 170–173, Stockholm, Sweden, July.

Edison Ong, Zuoshuang Xiang, Bin Zhao, Yue Liu, Yu Lin, Jie Zheng, Chris Mungall, Mélanie Courtot, Alan Ruttenberg, and Yongqun He. 2016. Ontobee: A linked ontology data server to support ontology term dereferencing, linkage, query and integration. *Nucleic acids research*, 45(D1):D347–D352, October.

Jyotishman Pathak and Christopher G Chute. 2009. Debugging Mappings between Biomedical Ontologies: Preliminary Results from the NCBO BioPortal Mapping Repository . In B Smith, editor, *International Conference on Biomedical Ontology*, pages 95–98, Buffalo, NY, USA.

Depriest A Pouchard L. Huhns M. 2012. Lessons learned in deploying a cloud-based knowledge platform for the ESIP Federation. *American Geophysical Union Fall Meeting, poster session*:22725.

Dietrich Rebholz-Schuhmann, Miguel Arregui, Sylvain Gaudan, Harald Kirsch, and Antonio Jimeno. 2008. Text processing through Web services: Calling Whatizit. *Bioinformatics*, 24(2):296–298, January.

Carlos Rueda, Luis Bermudez, and Janet Fredericks. 2009. The MMI Ontology Registry and Repository: A Portal for Marine Metadata Interoperability. In *MTS/IEEE Biloxi - Marine Technology for Our Future: Global and Local Challenges, OCEANS'09*, page 6, Biloxi, MS, USA.

Marta Sabou, Vanessa Lopez, and Enrico Motta. 2006. Ontology Selection on the Real Semantic Web: How to Cover the Queens Birthday Dinner? In S Staab and V Svátek, editors, *15th International Conference on Knowledge Engineering and Knowledge Management Managing Knowledge in a World of Networks, EKAW'06*, volume 4248, pages 96–111, Podebrady, Czech Republic. Springer.

Saoussen Sakji, Q Gicquel, S Pereira, I Kergoulay, D Proux, Darmoni SJ, and M H Metzger. 2010. Evaluation of a French Medical Multi-Terminology Indexer for the Manual Annotation of Natural Language Medical Reports of Healthcare-Associated Infections. In C Safran et al., editor, *13th World Congress on Medical Informatics, MedInfo'10*, volume 160, pages 252–256, Cape Town, South Africa. IOS Press.

Manuel Salvadores, Paul R Alexander, Mark A Musen, and Natalya F Noy. 2013. BioPortal as a dataset of linked biomedical ontologies and terminologies in RDF. *Semantic Web*, 4(3):277–284.

Barry Smith, Michael Ashburner, Cornelius Rosse, Jonathan Bard, William Bug, Werner Ceusters, Louis J Goldberg, Karen Eilbeck, Amelia Ireland, Christopher J Mungall, The O B I Consortium, Neocles Leontis, Philippe Rocca-Serra, Alan Ruttenberg, Susanna-Assunta Sansone, Richard H Scheuermann, Nigam H Shah, Patricia L Whetzel, and Suzanna Lewis. 2007. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, 25(11):1251–1255.

Irena Spasic, Sophia Ananiadou, John McNaught, and Anand Kumar. 2005. Text mining and ontologies in biomedicine: making sense of raw text. *Briefings in Bioinformatics*, 6(3):239–251.

Suarez-Figueroa, Jens Hartmann, York Sure, Peter Haase, and M. Suarez-Figueroa. 2005. OMV–ontology metadata vocabulary. In C. Welty, editor, *Workshop on Ontology Patterns for the Semantic Web, WOP'05*, page 9, Galway, Irland. Springer.

Andon Tchechmedjiev, Amine Abdaoui, Vincent Emonet, and Clement Jonquet. 2017a. ICD10 Coding of Death Certificates with the NCBO and SIFR Annotator(s) at CLEF eHealth 2017 Task 1. In *Working Notes of CLEF eHealth Evaluation Lab*, volume 1866, page 16, Dublin, Ireland.

Andon Tchechmedjiev, Amine Abdaoui, Vincent Emonet, Soumia Melzi, Jitendra Jonnagaddala, and Clement Jonquet. 2017b. Enhanced Functionalities for Annotating and Indexing Clinical Text with the NCBO Annotator+. *Bioinformatics*, UNDER REVI. Under Review.

Mossakowski Till, Oliver Kutz, and Mihai Codescu. 2014. Ontohub: A semantic repository for heterogeneous ontologies. In *Theory Day in Computer Science, DACS'14*, page 2, Bucharest, Romania.

Pierre-Yves Vandenbussche, Ghislain A Atemezing, Maria Poveda-Villalon, and Bernard Vatant. 2014. Linked Open Vocabularies (LOV): a gateway to reusable semantic vocabularies on the Web. *Semantic Web*, 1:1–5.

Aravind Venkatesan, Nordine El Hassouni, Florian Phillipe, Cyril Pommier, Hadi Quesneville, Manuel Ruiz, and Pierre Larmande. 2015. Exposing French agronomic resources as Linked Open Data. In James Malone, Robert Stevens, Kerstin Forsberg, and Andrea Splendiani, editors, *8th International Conference on Semantic Web Applications and Tools for Life Sciences, SWAT4LS'15*, pages 205–207, Cambridge, UK. CEUR.

Patricia L Whetzel and NCBO Team. 2013. NCBO Technology: Powering semantically aware applications. *Biomedical Semantics*, 4S1(S8):49.

Mark D Wilkinson, Michel Dumontier, Ijsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, Jildau Bouwman, Anthony J Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T Evelo, Richard Finkers, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3.

1995. A case study in ontology library construction. *Artificial Intelligence in Medicine*, 7(3):227–255, June.