Multilingual Mapping Reconciliation between English-French Biomedical Ontologies

Amina Annane, Vincent Emonet, Faiçal Azouaou, Clement Jonquet

Multilingual Mapping Reconciliation between English-French Biomedical Ontologies

Amina ANNANE
Ecole nationale Supérieure d'Informatique, Algiers, Algeria
a_annane@esi.dz

Vincent EMONET
LIRMM, Univ. of Montpellier & CNRS, France
vincent.emonet@lirmm.fr

Faiçal AZOUAOU
Ecole nationale Supérieure d'Informatique, Algiers, Algeria
f_azouaou@esi.dz

Clement JONQUET
LIRMM, Univ. of Montpellier & CNRS, France
BMIR, Stanford University, USA
jonquet@lirmm.fr

ABSTRACT
Even if multilingual ontologies are now more common, for historical reasons, in the biomedical domain, many ontologies or terminologies have been translated from one natural language to another resulting in two potentially aligned ontologies but with their own specificity (e.g., format, developers, and versions). Most often, there is no formal representation of the translation links between translated ontologies and original ones and those mappings are not formally available as linked data. However, these mappings are very important for the interoperability and the integration of multilingual biomedical data. In this paper, we propose an approach to represent translation mappings between ontologies based on the NCBO BioPortal format. We have reconciled more than 228K mappings between ten English ontologies hosted on NCBO BioPortal and their French translations. Then, we have stored both the translated ontologies and mappings on a French customized version of the platform, called the SIFR BioPortal, making the whole thing available in RDF. Reconciling the mappings turned more complex than expected because the translations are rarely exactly the same than the original ontologies as discussed in this paper.

CCS Concepts
• **Information systems~Ontologies** • Information systems~Web data description languages • Information systems~Multilingual and cross-lingual retrieval • Theory of computation~Data integration.

Keywords
Ontology repository; ontology localization; biomedical ontologies; BioPortal; semantic web; multilingual mapping; ontology alignment; mapping reconciliation; linked data.

WIMS '16, June 13-15, 2016, Nîmes, France.

DOI: http://dx.doi.org/10.1145/2912845.2912847

1. INTRODUCTION
The biomedical domain is rich in terms of ontology[1]. However, the majority of these ontologies are in English [1] and even when ontologies are available in other languages like French, there is a strong lack of related tools and services to exploit them. This lack does not reflect the huge amount of biomedical data produced, especially in the clinical world (e.g., electronic health records). The repository of biomedical ontologies NCBO BioPortal (http://bioportal.bioontology.org) [2] includes, as of end 2015, more than 433 ontologies, only six are not in English, five in French and one in Spanish [3]. Furthermore, the UMLS (Unified Medical Language System) Metathesaurus [4], even if it covers 21 languages, 75.1% of its terms are in English and only 1.82% of its terms are in French [5]. There have been initiatives in the past to reinforce the involvement of French language in the UMLS [6, 7] but most of these French ontologies are still not included, they are most often aggregated and translated by the CISMeF group[2] [8] (324.000 French concepts in HeTOP vs. 85,000 in the native UMLS)[3].

The lack of support for ontologies in different languages represents a real barrier for non-English-speaking communities that produce and manage biomedical data in their own languages. Indeed, when biomedical resources contain text content, it is important that these resources' languages are the same as the language of ontologies that will help to index or exploit them. Hence there is the need to have multilingual or translated ontologies [9-11]. The translation of MeSH by the French organization INSERM[4] is a good example and has greatly enriched the French biomedical vocabulary in UMLS [1].

However, except in Metathesaurus approaches such as the UMLS or CISMeF where ontologies are integrated in a common model, when someone gets a translated ontology to work with, it is never formally aligned to the original one and there is no standard format or resource to get such alignments. It definitively prevents multilingual exploiting of ontologies for annotation, semantic search, and data indexing neither for integration or knowledge extraction from these data.

To ensure semantic interoperability it is not enough to just translate ontologies, we must also formally keep the link between objects of the translated ontologies and the original ones [12, 13]. Re-establishing this link is the aim of this work, which we have

[1] In this work we use ontology to identify both of the (biomedical) terminological and ontological resources.
[2] Rouen's University Hospital (http://www.chu-rouen.fr/cismef/)
[3] http://www.hetop.eu/hetop/
[4] http://www.inserm.fr/

called reconciliation[5] of multilingual mappings. These multilingual mappings[6], once established and represented in a formal way, can have multiple applications [14]. For example, they allow performing a multilingual indexing of biomedical resources, which allow multilingual semantic search. A user types in a query using French terms and retrieves results within English data resources (and vice-versa). Multilingual mappings also allow integrating biomedical data of different languages. For example, resistance to diseases differs from one population to another, and it is a research problem that could be studied at a larger scale thanks to the multilingual mappings which enable cross-language databases integration. Indeed, the correlation study between genotypes and diseases [15] across different populations databases, annotated each in its original language with biomedical ontologies, linked by multilingual mappings, allow researchers to have a better vision of the problem and potentially, to discover new knowledge.

Our work is part of the SIFR project (Semantic Indexing of French Biomedical Data Resources - http://www.lirmm.fr/sifr) in which we are interested in exploiting ontologies in construction of services like indexing, mining, and information retrieval for French biomedical resources. In this project, we develop a semantic indexing workflow (called the French Annotator) based on ontologies similar to that existing for English resources [16], but focused on the French resources. To improve the workflow and connect the used French ontologies to their English equivalents, the project focuses on the reconciliation of multilingual mappings. The present study concerns ten French ontologies hosted on the SIFR BioPortal (http://bioportal.lirmm.fr) (a local instance of BioPortal dedicated to French) that we wish to align formally with their original English ontologies hosted on the NCBO BioPortal. The idea is to be able to retrieve from a French concept in the SIFR BioPortal, its corresponding English concept in the NCBO BioPortal and vice versa. As of now we are mainly focusing on "monolingual" ontologies but in parallel we are studying how to manage multilingualism in BioPortal [17].

The rest of the paper is organized as follows: Section 2 is dedicated to the presentation of related work in the field. Section 3 describes our approach to represent multilingual mappings within the BioPortal architecture. Section 4 presents the ontologies used in our work. Section 5 explains the followed methodology. Then, section 6 exposes obtained results. Section 7 discusses the study and its results. Finally, section 8 concludes and presents the perspectives of this work.

2. RELATED WORK

Multilingualism has always been considered as an important issue for the semantic web [18], that has even become more important with the explosion of data. Several challenges are identified [13], in particular cross-lingual ontology alignment and the representation of multilingual lexical information in ontologies, which are the starting points for cross-lingual access and querying of linked data. In the following, we briefly review related work on these two issues.

In the literature [19], several approaches have been proposed to extract multilingual mappings. The first was the manual approach where mappings are extracted by human experts as in the work of Liang and Sini [20], who manually aligned the English version of the AGROVOC thesaurus to the Chinese Agriculture Thesaurus. Despite the good quality of mappings generated by this approach, it cannot be used to process large and complex ontologies. Therefore, researchers have turned to automated approaches using different techniques: machine learning [21], machine translation [22], extraction mappings using multilingual background [23], etc. Overall, the ontology alignment community mostly focuses on the topic of generating mappings between different ontologies in different languages [19, 24] and ignores the problem of mapping reconciliation considered (truly) as a more easy issue. However, the reality shows us that: (i) it is not that trivial: ontologies and their translation are always different (they do not follow the same evolution after the process of translation) and (ii) the community still needs those mappings out there for use in concrete applications.

On the other hand, there have been several attempts to define models representing the linguistic description of terminological resources on the web (thesaurus, ontologies, etc.). The RDFS model allows to represent labels of concepts through the *rdfs:label* property without more information. SKOS model refine this property and decompose it into three properties which are "preferred label", "alternative label" and "hidden label". However, these properties are not enough to describe the linguistic characteristics and in particular cross-lingual specifications. To fill these gaps, other models were proposed such as: the GOLD ontology (General Ontology for Linguistic Description) [25], which allows to represent formal linguistic concepts using an OWL ontology. The Lemon model (LExicon Model for ONtologies) model [26], which is now the most widespread representation for the publication of lexical resources as linked data. Indeed, Lemon is the result of the evolution of several models: LMF (Lexical markup framework) model, LexInfo model [27] and Linguistic Information Repository [28]. Lemon allows describing more information on lexica, in particular: morphology, phrasing structure and subcategorization information. It also allows representing lexical information relative to an ontology that is shared on the semantic web. It has been gradually expanded to include new modules such as translations [29] resulting in the newly developed model OntoLex/Lemon [30]. It is really good to have such models to represent all linguistic details of lexica, but we also need to think about the use of proposed models. Rich models such as Lemon are complex to implement. Indeed, details as parts of speech, morphology, etc. need linguistic experts to determine them and formalize them. This task is very hard, especially for large and complex ontologies like SNOMED-CT. Consequently, there is a need to specific tools to support these models use in order to convince stakeholders in the web of data to adopt them [31].

As of now, the biomedical domain is one of plenty of ontologies that are not being lexically grounded, and are not multilingual. For which a translation has sometimes been produced by another group/project than the group that has developed the original one (e.g., MeSH, MedlinePlus, ICD, MEDDRA, and ICPC). Many of these ontologies are made available within the NCBO BioPortal [2] but this platform is not multilingual even if it accepts both

[5] We use the term reconciliation to avoid the confusion with ontology alignment extraction or creation approaches, which challenges the fact of aligning (some time multilingual) different ontologies [24].

[6] In the rest of this paper, the two terms mapping and alignment are interchangeable and mean a one-to-one correspondence between concepts of two ontologies.

multilingual and monolingual ontologies [32]. Another important resource in the biomedical domain, the UMLS Metathesaurus, which is a set of terminologies manually integrated and distributed (mostly publicly) by the United States National Library of Medicine [4]. It does contain terminologies in other languages than English and therefore, explicitly store the mappings between them. However, the number of French resources in the UMLS is not sufficient to cover the diversity of the biomedical domain. The HeTOP portal [8] also offers translated terms in multiple languages, especially French, and enables cross lingual search but most of its content is not publicly or easily accessible (e.g., No web service API or ontology download functionality). Furthermore, in both cases, the underlying approach is one of a common meta-model for all the integrated ontologies which means that there exists a unique abstraction for concepts in different sources (e.g., the UMLS Concept Unique Identifiers (CUI)) and label properties offer translations to multiple languages. This is different from the BioPortal approach that we are also following. This approach does not build a global thesaurus but keep each ontology separated and use mappings to interconnect them [33, 34]. Another difference with BioPortal, is that neither UMLS nor HeTOP are built natively with semantic Web technologies and thus do not offer semantic representation to make multilingual ontologies or multilingual mappings available as linked data.

The review of the state of the art identifies (i) the need (at least for French) for an explicit reconciliation of the multilingual mappings between translated ontologies and their origin ones and (ii) the need for making them available as linked data.

3. MULTILINGUAL MAPPING IN BIOPORTAL

Our aim is to link the French ontologies hosted on SIFR BioPortal with their English counterparts hosted on the NCBO BioPortal. For this purpose, we need to represent multilingual mappings [17] in a way that will ensure the interoperability between the two portals and avoid duplicating the data.

3.1 Choice of the Mapping Properties

BioPortal stores mappings in a particular format that reifies a mapping as a RDF resource. These mappings can have several properties including provenance information (process, note, date, who created, etc.). Especially, BioPortal uses one property of standard semantic web vocabularies to tag / describe a mapping between two concepts of ontologies[7]. For example, the property *skos:exactMatch* to indicate that two concepts are identical and the property *skos:closeMatch* to mark a strong bond of similarity between two concepts without being completely identical. With the mapping being reified, the triple e.g., (Cls1 skos:exactMatch Cls2) is not explicitly materialized but can be retrieved any time. Indeed, all mappings (as any other data in the portal) are available either via a SPARQL endpoint or via JSON-LD.

We propose to represent multilingual mappings as any mapping in the portal, but with specific additional semantic properties to mark the linguistic aspect and formalize the translation relationship between two concepts. For example, the concept *Mélanome* (http://purl.lirmm.fr/ontology/MSHFRE/D008545) in the French version of MeSH within the SIFR BioPortal should be mapped to concept *Melanoma* in the English version of MeSH (http://purl.bioontology.org/ontology/MESH/D008545) in the NCBO BioPortal. This allows to specify that the two concepts *Mélanome* and *Melanoma* have the same meaning and that the first label is the French translation of the second one.

We can still use SKOS[8] properties to represent that the two concepts have the same meaning. Nevertheless, for the linguistic information, we need another property to describe the translation relationship. For this purpose, we propose to use Lemon or GOLD models [3] . In the following, we have chosen to use the GOLD properties, in particular *gold:freeTranslation* property to represent an accurate translation, and *gold:translation* to represent a less precise translation (see Figure 1). As of now, we did not use Lemon as no other lexical or linguistic layer was necessary for our biomedical ontologies.

3.2 Changes in BioPortal Architecture

In order to store our multilingual mappings, we had to change their representation in BioPortal's architecture, especially: (1) Allow to tag the same mapping with several semantic web properties to avoid duplicating the mappings (semantic mapping and translation mapping); (2) Allow a BioPortal virtual appliance[9] to store mappings that target ontologies (i) in another instance of BioPortal (inter-portal), (ii) that are not in any BioPortal instance (external mappings).[10]

In order to formally figure out the source and target languages of a translation, we assumed the metadata of the monolingual ontologies would mention the natural language used for labels. Indeed, BioPortal offers the property *omv:naturalLanguage* included in the OMV ontology (http://omv2.sourceforge.net) which uses ISO-639-3 to specify the appropriate language for each ontology.

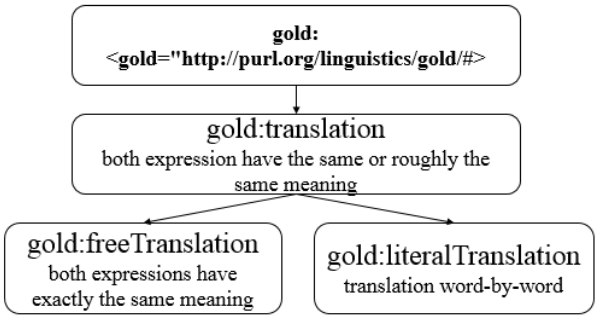

Figure 1. Translation properties of GOLD ontology.

4. ONTOLOGIES TO ALIGN

We have treated a set of 20 ontologies, 10 in French and 10 in English. These ontologies are widely used in the biomedical field both in French and in English. For example, the International Classification of Diseases (CIM-10/ICD10) is used in hospitals to code medical acts, the Medical Subject Heading (MeSH) is used for indexing documents both by the NLM (English) and CISMeF (French). In our study, all English ontologies come from the UMLS Metathesaurus (version 2015AA) and were imported by the NCBO team in the NCBO BioPortal using the umls2rdf tool

[7] http://www.bioontology.org/wiki/index.php/BioPortal_Mappings

[8] http://www.w3.org/2004/02/skos/

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

[10] This feature is not explicitly used in this study.

Table 1. Ontologies processed in this study (acronyms are identifiers from the NCBO BioPortal and the SIFR BioPortal).

N°	Ontology	Acronym	Version	Format	Source
01	Systematized Nomenclature of MEDicine	**SNMI**	2015AA	RDF\TTL	UMLS
	Systematized Nomenclature of MEDicine, version française	**SNMIFRE**	3.5	OWL	CISMeF
02	International Classification of Functioning, Disability and Health	**ICF**	1.0.2	OWL	UMLS
	Classification Internationale du Fonctionnement, du handicap et de la santé	**CIF**	2001	OWL	CISMef
03	MedlinePlus Health Topics	**MEDLINEPLUS (EN)**	2015AA	RDF\TTL	UMLS
	MEDLINEPLUS FR	**MEDLINEPLUS (FR)**	-	OWL	CISMeF
04	Minimal Standard Terminology of Digestive Endoscopy	**MSTDE**	2015AA	RDF\TTL	UMLS
	Terminologie minimale standardisée en endoscopie digestive	**MTHMSTFRE**	2011ab	RDF\TTL	UMLS
05	Semantic Types Ontology	**STY (EN)**	2015AA	RDF\TTL	UMLS
	Réseau sémantique UMLS	**STY (FR)**	2014AB	RDF\TTL	CISMeF
06	Medical Subject Headings	**MESH**	2015AA	RDF\TTL	UMLS
	Medical Subject Headings, version francaise	**MSHFRE**	2015AA	RDF\TTL	UMLS
07	Medical Dictionary for Regulatory Activities	**MEDDRA**	2015AA	RDF\TTL	UMLS
	Dictionnaire médical pour les activités règlementaires en matière de médicaments	**MDRFRE**	2015AA	RDF\TTL	UMLS
08	World Health Organization (WHO) Adverse Reaction Terminology	**WHO**	2015AA	RDF\TTL	UMLS
	World Health Organization (WHO) Adverse Reaction Terminology, version française	**WHO-ARTFRE**	1997	OWL	CISMeF
09	International Classification of Diseases, Version 10	**ICD10**	2015AA	RDF\TTL	UMLS
	Classification Internationale des Maladies, version 10	**CIM-10**	10	OWL	CISMeF
10	International Classification of Primary Care - 2 PLUS	**ICPC2P**	2015AA	RDF\TTL	UMLS
	Classification Internationale de Soins Primaires	**CISP-2**	1998	OWL	CISMeF

(https://github.com/ncbo/umls2rdf). The French ontologies come from the UMLS or were provided by the CISMeF group as an OWL file. In this second case, the translations were generally produced or synthesized by CISMeF. All processed ontologies are stored in the SIFR BioPortal (see Table1).

5. METHODOLOGY

The followed methodology consists of: (1) Download ontology files in .ttl or .owl formats from the NCBO and SIFR BioPortals. (2) Parse them with the Jena API to extract the necessary data for multilingual alignment. (3) Store the data in SQL table (one table per ontology). (4) Make the relevant "join" queries between the two tables on the field/property used to reconcile the mappings. (5) Finally, post the produced mappings to SIFR BioPortal after choosing the relevant GOLD and SKOS properties (see Figure 2).

5.1 Downloading Files

For this study, we have chosen ten ontologies for which we have a French version in the SIFR BioPortal and that contains labels that will be easily used by the SIFR Annotator for identifying biomedical words in text. These ontologies have been downloaded from English and French BioPortals. As an example, files of the English and the French version of the SNOMED International terminology (SNMI) are respectively available at: https://bioportal.bioontology.org/ontologies/SNMI, http://bioportal.lirmm.fr/ontologies/SNMIFRE .

5.2 Retrieving Data from Ontologies Files

We used the Jena API (https://jena.apache.org) to extract RDF triplets (concept, property, propertyValue) from the ontology files. To extract only the needed triplets, we filtered them according to a specific property. Mostly, this property is the field that we are going to use to reconcile mappings. Indeed, ontologies always provide a unique identifier or code for the concepts / classes they define. With recent OWL ontologies, this is of course an URI, but this was not the case for ontologies that have been originally designed not following a semantic web principles. To determine the appropriate property, we had to study the ontologies one by one. In most cases, the alignment was based on the internal code of concepts in ontologies. Except MedlinePlus, as there is no internal code for its concepts, we had to use the UMLS code (CUI). However, the code property name differs from one ontology to another, the most frequent was *skos:notation*, used for 12 of the 20 ontologies. Other specific properties have been used such as

Table 2. Retrieving codes from ontologies for reconciling the alignments

Ontology	Code source	Example
SNMI **MTHMSTFRE** **MSTDE** **STY FR** **STY EN** **MDRFRE** **MEDDRA** **Who-art (Who)** **ICD10** **MESH** **MSHFRE** **MEDLINEPLUS EN**	The internal code is set to the *skos:notation* property	This concept of MSTDE ontology has as an internal code the value "**MT200025"**. <http://purl.bioontology.org/ontology/MTHMST/MT200025> a owl:Class ; skos:prefLabel """Gastric angioectasia (diagnosis)"""@eng ; **skos:notation """MT200025"""**^^xsd:string ;
ICF	The internal code is set to the *icd:icdCode* property	The internal code is "**s1208**". <owl:Class rdf:ID="s1208"> **<icd:icdCode** rdf:datatype="http://www.w3.org/2001/XMLSchema#string" >**s1208**</icd:icdCode>
SNMIFRE **CIF** **WHO-ARTFRE** **CIM-10** **CISP-2**	Internal code extracted form URI of concepts. Extraction of all concepts were done by filtering on the *skos:prefLabel* property	From this URI http://chu-rouen.fr/cismef/SNOMED_int.#M-40030 of SNMIFRE concept, we extract the internal code "**M-40030"**
MEDLINEPLUS Fr	Concepts have no internal code, we used the CUI for alignment, filtering on *skos:altLabel* property because it's set to CUI. We have used a regular expression to detect cases where this property is set to labels and not to CUI.	<rdf:Description rdf:about="http://chu-rouen.fr/cismef/MedlinePlus#T351"> <skos:prefLabel xml:lang="fr">douleur</skos:prefLabel> <skos:altLabel xml:lang="fr">C0008031</skos:altLabel> <skos:altLabel xml:lang="fr">C0030193</skos:altLabel> <skos:altLabel xml:lang="fr">C0030231</skos:altLabel> </rdf:Description>
ICPC2P	The internal code is set to the icpc2p:*icpccode* property	<http://purl.bioontology.org/ontology/ICPC2P/ICPCCODE> """**A01**"""^^xsd:string ;

Table 3. Summary of results

French ontology	**Concepts**	**English ontology**	**Concepts**	**Mapped concepts**	**Mapping %**	**Mappings number**	**Properties (skos ; gold)**
STY	133	STY	133	133	100%	133	exactMatch; freeTranslation
MDRFRE	66378	MEDDRA	66378	66378	100%	66378	exactMatch; freeTranslation
CIF	1495	ICF	1495	1495	100%	1495	exactMatch; freeTranslation
MTHMSTFRE	1700	MSTDE	1699	1700	100%	1700	exactMatch; freeTranslation
MSHFRE	26142	MeSH	252242	26220	99.79%	26220	exactMatch; freeTranslation
WHO-ARTFRE	3482	WHO	1724	3482	100%	3482	broadMatch ; translation
CISP2	745	ICPC2P	7537	665	70%	5063	narrowMatch ; translation
MEDLINEPLUS	795	MEDLINEPLUS	2113	771	97%	1520	closeMatch; translation
CIM-10	19853	ICD10	12318	19813	99%	19813	exactMatch; freeTranslation 62% broadMatch; translation 37%
SNMIFRE	106266	SNMI	109150	102093	96%	102093	exactMatch; freeTranslation

icpc2p:icpccode for ICPC2P or *icd:icdCode* for ICF. In the five cases where a code property did not exist, we extracted the code

from the URI of classes (those were always provided as the files were extracted from BioPortal). In these cases, in order to retrieve only classes corresponding to a concept in the ontology, we filtered and retained only the classes that have a property *skos:prefLabel*. Technically, we have developed a script (function) to extract the code used for each alignment. Eventually, more treatments were necessary such as elimination of the type attached to the value or use of regular expression to isolate the exact code string. Table 2 summarizes the property used to retrieve the mapping code with examples for each ontology.

5.3 Saving Data

To store the extracted data in step (5.2), we have created a table into a relational database for each ontology. Each triplet retrieved adds a new record in the appropriate table. For example, the triplet (http://purl.bioontology.org/ontology/MSHFRE/D001542; skos:notation; D001542) extracted from the French version of MeSH generates the record shown in Table 4 with the three columns:

- Id: a sequential number that identifies each record in the table
- Code: a string that contains the code previously extracted and which may be an internal code of the concepts in the ontology, the CUI or any other relevant mapping criteria.
- URI: globally identifies the concept. This URI is either assigned by the ontology designers or created by the BioPortals.

Table 4. Example of content stored in a relational table.

Id	Code	URI
1	D001542	http://purl.bioontology.org/ontology/MSHFRE/D001542

There is no unique constraint on the relational table for the two fields "Code" and "URI". This is justified by the fact that for a given URI it is possible to have multiple codes, and for a given code we can have multiple URI. Even if this situation should not occur in theory, it actually does happen in practice. Especially with properties such as CUI that are not specific to the ontology but had been added when the ontology was integrated with other ones in the UMLS Metathesaurus. For instance, this is the case with the French version of the MEDLINEPLUS ontology, which contains 442 concepts that have more than one CUI. For example the concept *minéraux* (http://chu-rouen.fr/cismef/MedlinePlus#T4298) has nine different CUI. Consequently we have nine triplets that have the same concept (same URI) but different codes (different CUI) which involves the creation of nine records that have all the same URI. We also encountered cases where the code targets multiple URIs within the same ontology (ICPC2P, MedlinePlus FR, and CIM10). We will address these cases in more details in the Results section.

5.4 Reconciliation of Mappings

At this stage, we do a "join" clause on the "Code" field between the two corresponding tables. As we mentioned in the previous paragraph, the code used in the "join" is not necessarily unique within the same ontology. Consequently, the number of couples (fr URI, en URI) resulted from the "join" can be more than the number of concepts of one of the ontologies (or both) (see Figure 3). For example, the CISP2 ontology has generated 5063 mapping couples whereas it has only 745 concepts. The multiple assignment of a code to several resources generates also duplicated couples that we eliminated (using the SQL keyword "distinct" in the join query) because it represents only redundant information (see Figure 4).

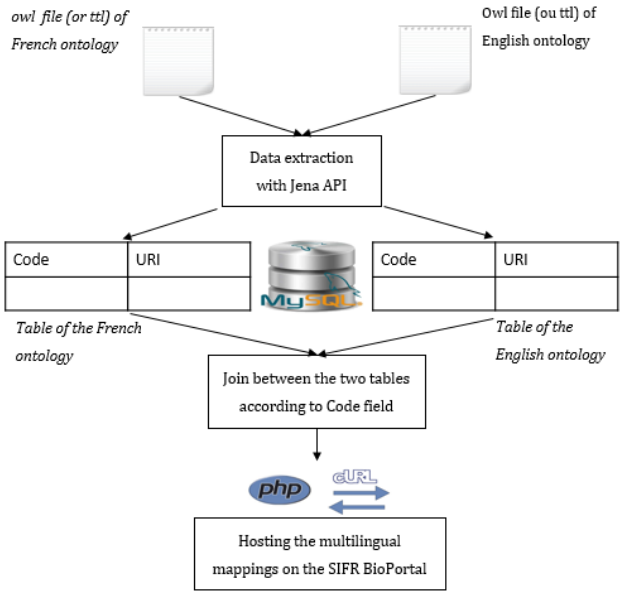

Figure 2. Multilingual mapping Methodology.

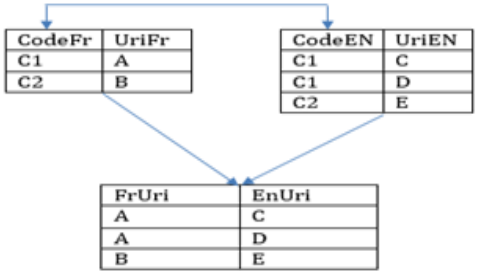

Figure 3. Case of several mappings for one URI (concept)

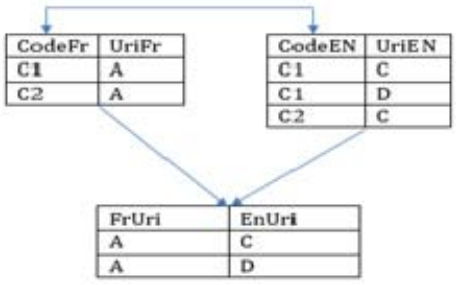

Figure 4. Case of mapping repetition, the third mapping (A, C) was deleted.

5.5 Mapping Property Selection and Loading in SIFR BioPortal

This final step allows representing and storing produced mappings in a formal and permanent way within the SIFR BioPortal. The particularity of our multilingual mappings, compared to other BioPortal mappings resides in the couple of properties by which they were tagged (GOLD translation property and SKOS alignment property) as follows:

5.5.1 skos:exactMatch/gold:freeTranslation

We used these properties when the French concept has the same meaning as the English concept. This is usually the case when the mapping is based on full equality between the internal code of the French concept and the internal code of the English concept. In that case, there are no more mappings generated than the minimum number of classes between the two ontologies being mapped. Fortunately, it is the most frequent case.

5.5.2 skos:broadMatch/gold:translation

We used these properties when the source concept is more precise than the target concept. This situation occurs when the translated ontology was modified (e.g., new sub-concepts more precise). For example, the French concept *agression par un moyen non précisé établissement collectif* (http://chu-rouen.fr/cismef/CIM-10#**Y09**.1) from CIM-10 which has internal code **Y09.1**, do not have an English concept having the same internal code in ICD10. However, we can map it to the concept *Assault by unspecified means* which has internal code **Y09** (http://purl.bioontology.org/ontology/ICD10/**Y09**) and tag this mapping with these properties.

5.5.3 skos:narrowMatch/gold:translation

We used this couple of properties when the target concept is more precise than the source concept. This situation occurs when the original ontology has been modified after the translation into French. We had this case in the alignment of CISP2 to ICPC2P through the icpc-code (internal code). Concepts of ICPC2P have the icpc-code following by another code (term-code), which is not the case for the French ontology CISP2 that only uses original icpc-code. Consequently, one French concept is mapped to several English concepts, which have the same icpc-code. For example, the concept labelled *tumeur bénigne* having **B75** as a code, has been mapped to eight English concepts more precise, three among them are (*benign neoplasm of the blood*, **B75**001), (*benign neoplasm of the lymphatics*, **B75**002), (*benign neoplasm of the spleen*, **B75**003). Hence, the need to use such properties.

5.5.4 skos:closeMatch/gold:translation

In the absence of an internal code, we had to use less precise identifiers such as the CUI for ontologies coming from UMLS. CUIs are identifiers at the Metathesaurus level, and not in the sources ontologies. Therefore, it is not a direct translation of the concept from one language to another but rather concepts that mean the same thing as they were affected to the same CUI.

The Table 5 summarizes semantic properties used to tag multilingual mappings between two concepts in our study. The first column indicates the type of translation equivalence as it was identified by Chen and Chen [35], the second one describes possible relationships between concept and the last one indicates the semantic properties used to tag our multilingual mappings. For more information about how multilingual mappings are represented and stored on SIFR BioPortal see section 6.9.

6. RESULTS

Our aim is to provide multilingual alignment for French versions of ontologies. We express our results as a percentage of the French ontology concepts for which, we were able to provide at least one translation mapping (see Table3).

Table 5. Summary of semantic properties used to describe mapping between two concepts

Translation equivalence	Relation between two concepts	Used properties
exact	Two equivalent concepts	skos:exactMatch gold:freeTranslation
inexact	Two concepts very similar but not equivalent	skos:closeMatch gold:Translation
partial	Target concept is less precise	skos:broadMatch gold:translation
	Target concept is more precise	skos:narrowMatch gold:translation

6.1 STY/STY, MDRFRE/MEDDRA, CIF/ICF

For these three couples of ontologies, we got a full alignment, one to one for all concepts (percentage of 100%) thanks to the internal code of concepts.

6.2 MTHMSTFRE/MSTDE

Among the 1700 concepts of the French version, only 2 concepts have not been mapped since their codes do not exist in the English version. These two concepts have as an internal code MT180041 and MT180042. In the English version, there is one concept that has not been mapped. Its internal code is set to "nocode". However, this concept has two CUI which are those assigned to the unmapped French concepts. Thus, the mapping of these two French concepts was done according to their CUI, which allowed us to get a percentage alignment of 100%. We estimate that this "nocode" case is an error in the integration of MSTDE in UMLS and we will share this with NLM.

6.3 MSHFRE/MESH

The number of concepts in the English version (252242 concept) is ten times greater than the number of concepts in the French version (26142 concept), because the French version contains only the MeSH descriptors without the other additional concepts[11]. Our mappings cover almost all French concepts with a percentage of 99.79%. Only 55 concepts of the French version have not been aligned because their codes do not exist in the English one such as D020185 (*Acides benzoïques*), D002134 (*Protéine de liaison au calcium dépendant de la vitamine D*), D006587 (*Acides hexanoïques*). Moreover, even trying to align these 55 concepts using CUI, we have not found their CUI in the English MeSH. When CUI are used in MSHFRE but they do not exist in MESH, we think they are probably mistakes that were made by translators or problems which appeared during the integration of the new translation into UMLS. Indeed, the two versions should be perfectly aligned since they both come from UMLS 2015AA. To correct this problem we will communicate the 55 non-mapped concepts to INSERM and NLM.

6.4 WHO-ARTFRE/WHO-ART

In the English version WHO-ART, the internal code of concepts can be retrieved through the *skos:notation* property. But in the French version, this code cannot be found in any property of WHO-ARTFRE; we had to extract it from URIs of concepts. Indeed, this code is located at the end of the URI of each concept.

[11] See http://mesh.inserm.fr/mesh for more information about MeSH translation.

For example, the code of the concept with URI http://chu-rouen.fr/cismef/WHO-ART#**1545**_PT is **1545**.

We have noticed that the French version has undergone some customization. Indeed, a code of the English version can reference several French sub-concepts that have the same code but suffixed to differentiate them. For example, the code 1723 references four concepts that have the following codes: 1723_IT0, 1723_IT1, 1723_IT2, 1723_PT. Therefore, the number of French concepts is greater than the number of English concepts (3320 vs 1724). The French version is more detailed, their concepts are more precise than those of the English version, so we used the two properties (*skos:broadMatch* ; *gold:translation*) to describe mappings between these ontologies. Finally, all the French concepts were mapped.

6.5 CISP2/ICPC2P

The French version CISP2 contains 745 concepts while the number of generated mappings was 5141. This is explained by the fact that the English version which has been modified in this time after the French translation. English concepts have been customized to generate new more detailed concepts (sub-concepts of the original ones: 7354 concepts). Therefore, an icpc-code of the French version was mapped with one or several English concepts that have the same icpc-code but differentiated through another code called "term code".

For example, the code **A01** is affected to a single concept in the French version (*Douleur générale/de sites multiples*; **A01**) while the English version contains four more precise concepts as follows: (*generalised aches*; **A01**001), (*generalised pain*; **A01**003), (*body pain*; **A01**004), (*chronic pain*; **A01**005). Consequently, a single concept of CISP2 may generate several couples of mapping, one for each English concept that has the same icpc-code. For this reason, we used the properties (*skos:narrowMatch; gold:translation*) to describe the mappings. 59 of the French ontology concepts have not been aligned with the icpc-code such as: (*Autre analyse de laboratoire*; 38), (*Conseil thérap/écoute/psychothérapie*; 58), (*Examen microbiologique/immunologique*; 33). We figured out these concepts do not have an icpc-code as the rest of the concepts that consists of a letter followed by two digits. In addition, these concepts have no CUI property as well. It seems they have been added in the translation, or removed from the English version. For this reason we reached only 70% of mappings for CISP2 and we will communicate our results to the translators.

6.6 MEDLINEPLUS FR/EN

We had to use the CUI property to align MEDLINEPLUS as its concepts have no other internal code[12]. The French version of MedlinePlus contains 795 concepts. Each concept has one or more CUI value (442 concepts have more than one) which gives 1686 distinct couples (concept, CUI). The English version contains 1986 distinct concepts and each concept has a single CUI value. Indeed, the URI of each English concept is suffixed by the CUI assigned to it, for example: C0003803 is the CUI of the URI http://purl.bioontology.org/ontology/MEDLINEPLUS/C0003803. It is surprising to note that there are CUIs that do not exist in the English version but are assigned to concepts of the French version. However, even if we ignore the concepts with these CUI (147 concept), the number of couples (concept, CUI) remain greater than the number of concept (1520 couples vs 795 concepts). 123 concepts of these 147 have other CUIs belonging to the English version but the remaining 24 concepts have no CUI belonging to the English version. So these concepts do not exist or no longer in the English version. Therefore, 24 French concepts are not mapped and we obtain a percentage alignment of 90% in terms of aligned couples and 97% in terms of aligned concepts. We have tried to refine the study, for eight among concepts that do not appear in the English version at all, we applied the following procedure:

1. Search the preferred label in the French ontology;
2. Translate manually the term, using the terminology portal TermSciences (www.termsciences.fr) or another lexical resource (e.g., BabelNet or even simple Google translation), into English;
3. Search, in the English version, the obtained translated term and if the English corresponding concept exists, note its CUI.

Table 6. Correspondences between unmapped French concepts and English concepts

CUI used in FR	Preferred Label	CUI used in EN	Proffered Label
C0156543	Avortement	C0392535	Abortion
C2362506	Fitness et exercice	C1456706	Fitness and Exercise
C0021311	Infections	C3714514	Infections
C1456593	santé mentale et comportement	C1832070	mental health and behavior
C1456620	vivre avec le SIDA	C2963182	Living with HIV/AIDS
C1456571	nutrition des nourrissons et des bébés	/	"nutrition of infants and babies" not found
C2362562	sécurité du patient	C1113679	patient safety
C0002808	Anatomie	C0700276	Anatomy

As we can see in the Table 6, in seven cases over eight, we found the English concept, which corresponds to the French concept but with a different CUI. These results make us think that these 24 unmapped concepts are mistakes in the CUI choice during the translation process. We intend to communicate these concepts to the translators in order to detect possible errors and possibly update their translation.

6.7 CIM-10/ICD10

CIM-10 contains 19853 concepts while its English version, ICD10, contains 12318 concepts. Here again, we figured out that the French version has undergone some customizations; it was enriched with more detailed concepts resulting from specialization of the original concepts. A "join" clause according to the internal codes of concepts between the two ontologies

[12] Note that this is a parsing error from the NCBO BioPortal. The UMLS does offer a code property for MEDLINEPLUS. We have notified the NCBO of this error and will re-execute our mappings reconciliation once the problem will be fixed. However, even with this code property MEDLINEPLUS will generated multiple mappings.

generated mapping percentage of 62% (12 308 concepts were mapped). We observed that there are six chapters in the French version, CIM-10, that do not have the same internal code as their English counterparts such as chapter (*autres maladies infectieuses*; **B99**) in CIM-10, while in ICD10 the same chapter is (*Other infectious diseases*; **B99-B99**). These chapters have the characteristic to contain only one entry. We had to treat them manually since the join according the code field did not work. All of the previous mappings (automatic and manual) were tagged with properties (*skos:exactMatch; gold:freeTranslation*).

As for the concepts generated by specializations (which codes do not exist in the English version), we extracted the code of their direct unique parent concept (the first 3 digits of their internal code) and mapped them with the correspondent English parent concepts tagging them with the properties (*skos:broadMatch ; gold:translation*). For example, all the French concepts (*Agression par d'autres moyens précisés /domicile* ; **Y08.0**), (*Agression par d'autres moyens précisés/ établissement collectif* ; **Y08.1**), (*Agression par d'autres moyens précisés /lieu de sport et d'athlétisme* ; **Y08.3**) were mapped with the English concept (*Assault by other specified means* ; **Y08**). By following this process, we reduced the number of unmapped concepts from 7545 to 40 concepts, which gives 99% of mapped French concepts.

6.8 SNMIFRE/SNMI

The French version SNMIFRE has 106266 concepts, while the English version contains 109150 concepts; there is a difference of 2884 concepts. Using the internal code, 102093 French concepts have been mapped (96% of the French ontology). However, there remained 4173 concepts of the French version without mapping. We tried then to use the CUI property, but those 4173 concepts are part of a set of 9510 French concepts that do not have this property (whereas all concepts of the English version does have a CUI). We have not found another relevant field to use for mapping the remaining 4173 concepts.

6.9 Multilingual Mappings Hosted on SIFR BioPortal

All alignments produced in our study are hosted on the SIFR BioPortal with a script that uses SIFR BioPortal REST web service API (http://data.bioportal.lirmm.fr/documentation). As a result, for all ontologies processed during this work, when browsing a concept (see Figure 6), we can see in the "Class Mappings" tab the multilingual alignments classified as "Interportal mappings" with a flag to indicate that it is a linguistic mapping to English, we can also observe the properties used. The aligned concept link allows the user to switch from the SIFR BioPortal to the target concept in the NCBO BioPortal. Like all the content of the SIFR BioPortal, in addition to the graphical interface, these multilingual mappings are also available directly via the REST web service API and a SPARQL endpoint which makes them part of the web of data; easily readable and reusable by any semantic web applications.

7. DISCUSSION

In this work, we propose an approach to formally represent semantic links between translated ontologies and their original ones. Particularly, we focused on French ontologies hosted within the SIFR BioPortal and their English counterparts hosted within the NCBO BioPortal.


```
{"creator":"http://data.stageportal.lirmm.fr/users/AAmina" ,
 "source_contact_info":"a_annane@esi.dz",
 "relation" : ["http://www.w3.org/2004/02/skos/core#exactMatch",
               "http://purl.org/linguistics/gold/freeTranslation"],
 "source":"REST",
 "source_name":"Reconciliation of multilingual mapping",
 "comment" : "Multilingual mapping",
 "classes" : { "http://chu-rouen.fr/cismef/SNOMED_int.#A-01020" : "SNMIFRE",
          "http://purl.bioontology.org/ontology/SNMI/A-01020" : "ncbo:SNMI"}}
```


Figure 5. An example of a multilingual (JSON-LD format) mapping stored on SIFR BioPortal[13]

Our approach consists in reconciling and representing these links as multilingual mappings using semantic web properties. However, this work should not be confused with multilingual mapping extraction that consists in aligning two different ontologies, which have no relationship with each other and which are not in the same language. Indeed, in most of our cases, we have used internal codes to reconcile links. Hence, the semantic link between the translated concept and its origin existed implicitly through the internal code despite difficulties we have met in certain cases. Our mission was to restore these links and represent them in a formal way and publicly made them available where the ontologies actually reside. However, our approach to represent and store the mappings can be used to represent any kind of mappings either reconciled or extracted assuming the relevant semantic properties will be used. In our case, we have chosen SKOS and GOLD properties. They are complementary, especially in the linguistic aspect. Indeed, the *gold:translation* does not represent the difference between the narrow translation, broad translation or close translation (see Table 5), but combining with SKOS properties we have the exact description. For example the couple (*skos:narrowMatch ; gold:translation*) describes inexact translation of type *narrow*.

We could also have tagged the best mappings with the *owl:sameAs* property because in theory the concept is exactly the same, and their logical entailment should be equivalent. However, we did not want to take the risk to assign such a property without experimentally verifying that no other inconsistencies will show up. We therefore left it to future users the choice of considering those mappings as *owl:sameAs* when materializing the triples e.g., (Cls1 tag Cls2) out of BioPortal's mapping repository..

It is necessary to evaluate the result of an alignment process [19, 24] to be able to use them. However, since we did a reconciliation of mappings, we have restored links between concepts based on internal code of concepts and not on a terminological, structural or semantic measures [36]. Consequently, our approach gives automatically reliable and verifiable results. Indeed, 92% of the produced mappings are the result of total equivalence of concepts' internal code and 7% of partial equivalence (the internal code of the French concept is included in the internal code English concept or the reverse). We do acknowledge that the remaining 1% of MEDLINEPLUS mappings had to be verified because of the multiple affectation of CUI to a given concept in the French version, which is not the case in the English version (see section 6.4). For example, the concept *santé au travail pour les professionnels de santé* has two CUIs (C1456673, C0206333), therefore, it was mapped with two English concepts (Blood-Borne Pathogens, C0206333) and (*Occupational Health for Healthcare*

[13]In BioPortal, the mappings have more properties to describe their provenance (not represented here).

Providers, C1456673).Whereas in this case, only the second target concept is correct. What causes the error was the wrong affectations of CUIs and our work should help the translator of the ontology to fix them. It is important to note that even if the community produces less and less "monolingual" ontologies and that designers are opting increasingly for "multilingual ontologies"; we cannot assume that ontology translation will not happen anymore. Indeed, regardless of the richness of an ontology in terms of language (2, 3 even 10 languages), it would never cover all languages. Translated ontologies remain then an ineluctable solution to ensure their exploitation in other languages that are not supported in native version. We hope this study will convince ontology translators about the importance of reusing the same identifiers when creating a translated version. Eventually, the best situation is to follow the semantic web principles and actually reuse the exact same URI, when available, rather than creating a new one. Furthermore, in the process of creating multilingual ontologies, there is still a challenge of going further than the simple use of the *xmllang* tags and move to using lexical standards models such as Lemon.

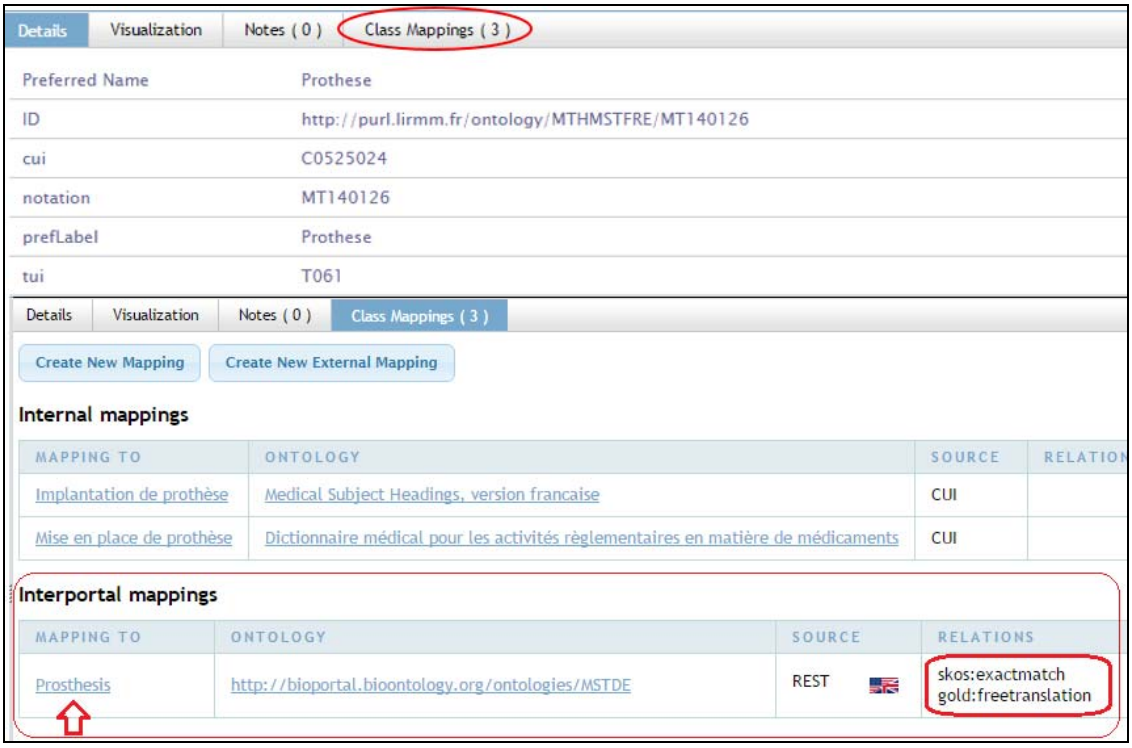

Figure 6. Example of a multilingual mapping for the concept *Prothèse* in MTHMSTFRE within the SIFR BioPortal. Direct link is: http://stageportal.lirmm.fr/ontologies/MTHMSTFRE/?p=classes&conceptid=http%3A%2F%2Fpurl.lirmm.fr%2Fontology%2FMTHMSTFRE%2FMT140126&jump_to_nav=true#mappings For each French ontology treated in this work, we can consult on the "class mapping" section "Interportal mapping". We can see the semantic properties used to tag the multilingual mapping, and switch to the English concept hosted on NCBO BioPortal.

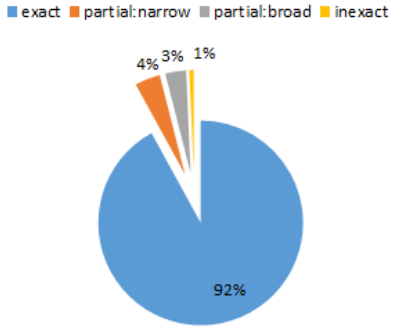

Figure 7. Distribution of multilingual mappings per type

The multilingual mapping links produced in this study can have several applications including the integration of biomedical data of different languages, and multilingual semantic search and indexing. In the continuation of the SIFR project, these links will be integrated into the French version of the NCBO Annotator [16] which will expand direct annotations with French ontologies to new annotations with (i) their corresponding English ontologies, (ii) other English-only ontologies mapped one another inside the NCBO BioPortal. In addition, our mappings will also be a good corpus for automatic translation of biomedical ontologies i.e., they can help translators themselves to translate more ontologies. Although this is not mostly the case for the current 20 selected ontologies, we have to assume both the original and the translated ones could be modified in the future[14]. Therefore, it is necessary to implement a strategy to keep multilingual mappings up to date [37]. Currently, we run the script again when a new version is available; we remove all the old multilingual mappings to store the new ones. This processing will be done automatically once the

[14] This is for instance the case with MeSH and MSHFRE that are modified every year respectively by NLM and INSERM.

script of multilingual mapping reconciliation will be integrated directly into the SIFR BioPortal.

8. CONCLUSION AND FUTURE WORK

Biomedical ontologies play a key role in the semantic interoperability of data. They have often been created in a particular language(s) and then translated to other language(s) in order to reuse them. Indeed, the use of ontologies for annotation, search and indexing of data is strongly linked to the syntactic correspondence between ontologies languages and the data languages. Hence, there is a need of multilingual management to allow the exploitation of knowledge formalized in the ontologies in other languages in addition to their original language(s).

In this work, we performed alignment reconciliations; we have restored formal semantic links between ten translated French ontologies and their English counterparts using semantic properties of SKOS and GOLD vocabularies. Finally, all these mappings (228k) are stored on the SIFR BioPortal platform (http://bioportal.lirmm.fr/mappings) and they are available to the scientific community as linked open data through a SPARQL endpoint and also as a web service API that returns JSON-LD format. In a near future, we also plan to process LOINC (Logical Observation Identifiers Names and Codes) that was recently made available in French in UMLS.

To accomplish this work, we had to treat each pair of ontology apart with its specificities especially in the choice of alignment property and how to recover it. Refinements were needed when translated ontologies did not follow exactly the content of the original ontology (in English).

Through this study, we have found some anomalies in certain pairs of ontologies which we intend to communicate to the translators in order to review them and eventually correct them. The current listing of the anomalies and the concepts that are concerned is available as well as our reconciliation scripts and the data used at: https://github.com/sifrproject/multilingual_mappings.

This work represents a part of the SIFR project aiming to efficiently manage multilingualism in a repository of biomedical ontologies such as NCBO BioPortal. As future work, will use these mappings for the development of a process to infer ontologies translations automatically based on multilingual ontologies [1], different dictionaries and Metathesaurus like UMLS, BabelNet, etc. We will also work on the valorization of these mappings in services such as indexing, annotation and semantic search. Another interesting work to achieve will consist in materializing some of the reified mappings (exact mappings) into *owl:sameAs* direct mappings and use a reasoner to check possible inconsistencies in the whole repository of interconnected ontologies

ACKNOWLEDGEMENTS

This work was achieved during a LIRMM-ESI collaboration within the SIFR project funded in part by the French National Research Agency (grant ANR-12-JS02-01001), the NUMEV Labex (grant ANR-10-LABX-20), the Computational Biology Institute of Montpellier (grant ANR-11-BINF-0002) as well as by University of Montpellier and the CNRS. We thanks the NCBO (Stanford University) and the CISMeF group (Rouen CHU Hospital) for the availability of the ontologies used in this work.

REFERENCES

[1] Névéol, A., et al. Language resources for French in the biomedical domain. in 9th International Conference on Language Resources and Evaluation (LREC'14). Reykjavik, Iceland. 2014.

[2] Noy, N., et al., BioPortal: ontologies and integrated data resources at the click of a mouse Nucleic Acids Research, 2009. 1(4).

[3] Jonquet, C., V. Emonet, and M.A. Musen. Roadmap for a multilingual BioPortal. in MSW4'15: 4th Workshop on the Multilingual Semantic Web. 2015. Portoroz, Slovenia.

[4] Bodenreider, O., The unified medical language system (UMLS): integrating biomedical terminology. Nucleic Acids Research, 2004. 32(suppl 1): p. D267-D270.

[5] Bollegala, D., G. Kontonatsios, and S. Ananiadou, A Cross-lingual Similarity Measure for Detecting Biomedical Term Translations. PLoS ONE, 2015. 10(6).

[6] Darmoni, S., et al. VUMeF: extending the French involvement in the UMLS Metathesaurus. in AMIA Annual Symposium Proceedings. 2003. American Medical Informatics Association.

[7] Zweigenbaum, P., et al. Towards a unified medical lexicon for French. in Medical Informatics in Europe (MIE). 2003.

[8] Grosjean, J., et al. Multi-terminology cross-lingual model to create the Health Terminology/Ontology Portal. in American Medical Informatics Association Annual Symposium. 2012.

[9] Meilicke, C., R. García-Castro, and F. Freitas, MultiFarm: A benchmark for multilingual ontology matching. Web Semantics: Science, Services and Agents on the World Wide Web, 2012. 15: p. 62-68.

[10] Fu, B., R. Brennan, and D. O'Sullivan. Multilingual ontology mapping: Challenges and a proposed framework. in Workshop on Matching and Meaning-Automated Development, Evolution and Interpretation of Ontologies. 2009.

[11] Deléger, L., M. Merkel, and P. Zweigenbaum, Translating medical terminologies through word alignment in parallel text corpora. Journal of Biomedical Informatics, 2009. 42(4): p. 692-701.

[12] Buitelaar, P., et al., Towards linguistically grounded ontologies, in The semantic web: research and applications2009, Springer. p. 111-125.

[13] Gracia, J., Cross-lingual ontology matching as a challenge for the Multilingual Semantic Web, in Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik2012.

[14] Fu, B., R. Brennan, and D. O'Sullivan. Cross-Lingual Ontology Mapping and Its Use on the Multilingual Semantic Web. in Multilingual Semantic Web. 2010. Raleigh, North Carolina, USA.

[15] Köhler, S., et al., The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. Nucleic Acids Research, 2013: p. gkt1026.

[16] Jonquet, C., N.H. Shah, and M.A. Musen, The Open Biomedical Annotator in American Medical Informatics Association Symposium on Translational BioInformatics, AMIA-TBI'09 2009. p. 56-60.

[17] Jonquet, C., V. Emonet, and M.A. Musen. Roadmap for a multilingual BioPortal. 4 th Workshop on the Multilingual Semantic Web 2015.

[18] Buitelaar, P. and P. Cimiano, Towards the Multilingual Semantic Web2014: Springer.

[19] Trojahn, C., et al., State-of-the-Art in Multilingual and Cross-Lingual Ontology Matching, in Towards the Multilingual Semantic Web, P. Buitelaar and P. Cimiano, Editors. 2014, Springer Berlin Heidelberg. p. 119-135.

[20] Liang, A.C. and M. Sini, Mapping AGROVOC and the Chinese Agricultural Thesaurus: definitions, tools, procedures. New Review of Hypermedia and Multimedia, 2006. 12(1): p. 51-62.

[21] Spohr, D., L. Hollink, and P. Cimiano, A machine learning approach to multilingual and cross-lingual ontology matching, in The Semantic Web–ISWC 20112011, Springer. p. 665-680.

[22] Fu, B., R. Brennan, and D. O'Sullivan, A configurable translation-based cross-lingual ontology mapping system to adjust mapping outcomes. Web Semantics: Science, Services and Agents on the World Wide Web, 2012. 15: p. 15-36.

[23] Tigrine, A., Z. Bellahsene, and K. Todorov, Light-Weight Cross-Lingual Ontology Matching with LYAM++, in On the Move to Meaningful Internet Systems: OTM 2015 Conferences, C. Debruyne, et al., Editors. 2015, Springer International Publishing. p. 527-544.

[24] Euzenat, J. and P. Shvaiko, Ontology Matching2013: Springer Berlin Heidelberg.

[25] Farrar, S. and D.T. Langendoen, A linguistic ontology for the semantic web. Glot International, 2003. 7(3): p. 97-100.

[26] McCrae, J., D. Spohr, and P. Cimiano, Linking lexical resources and ontologies on the semantic web with lemon, in The semantic web: research and applications2011, Springer. p. 245-259.

[27] Cimiano, P., et al., LexInfo: A declarative model for the lexicon-ontology interface. Web Semantics: Science, Services and Agents on the World Wide Web, 2011. 9(1): p. 29-51.

[28] Montiel-Ponsoda, E., et al., Modelling multilinguality in ontologies. 2008.

[29] Gracia, J., et al. Enabling language resources to expose translations as linked data on the web. in Proc. of 9th Language Resources and Evaluation Conference (LREC 2014), Reykjavik (Iceland). 2014. Reykjavik, Iceland: European Language Resources Association.

[30] Bosque-Gil, J., et al., Applying the OntoLex Model to a Multilingual Terminological Resource, in The Semantic Web: ESWC 2015 Satellite Events, F. Gandon, et al., Editors. 2015, Springer International Publishing. p. 283-294.

[31] Gracia, J., et al., Challenges for the multilingual Web of Data. Web Semantics: Science, Services and Agents on the World Wide Web, 2012. 11: p. 63-71.

[32] Jonquet, C. and M. Musen, A. Gestion du multilinguisme dans un portail d'ontologies: étude de cas pour le NCBO BioPortal. in TOTh'14: Terminology and Ontology : Theories and applications Workshop. 2014. Bruxelles, Belgium.

[33] Noy, N., N. Griffith, and M. Musen. Collecting Community-Based Mappings in an Ontology Repository. in 7th International Semantic Web Conference, ISWC'08. 2008. Karlsruhe, Germany: Springer.

[34] Ghazvinian, A., et al. What Four Million Mappings Can Tell You about Two Hundred Ontologies. in 8th International Semantic Web Conference, ISWC'09. 2009. Washington DC, USA: Springer.

[35] S. jiun, C. and C. H. hua, Mapping multilingual lexical semantics for knowledge organization systems. The Electronic Library 2012. 30(2): p. 278–294.

[36] Shvaiko, P. and J. Euzenat, Ontology Matching: State of the Art and Future Challenges. Knowledge and Data Engineering, IEEE Transactions on, 2013. 25(1): p. 158-176.

[37] Hartung, M., T. Kirsten, and E. Rahm. Analyzing the evolution of life science ontologies and mappings. in Data Integration in the Life Sciences. 2008. Springer.