A linked data terminology for copyright based on ontolex-lemon

Víctor RODRIGUEZ-DONCEL [1,a], Cristiana SANTOS [b], Pompeu CASANOVAS [b,c],
Asunción GÓMEZ-PÉREZ [a], Jorge GRACIA [a]

[a]*Ontology Engineering Group, Universidad Politécnica de Madrid*
[b]*Autonomous University of Barcelona-Institute of Law and Technology*
[c]*Royal Melbourne Institute of Technology-Centre for Applied Social Sciences*

Abstract. Ontolex-lemon is the de facto standard to represent lexica relative to ontologies and it can be used to encode term banks as RDF. A multi-lingual, multi-jurisdictional term bank of copyright-related concepts has been published as linked data based on the ontolex-lemon model. The terminology links information from WIPO (concepts and definitions), IATE (multilingual terms, usage notes) and other sources as Creative Commons (multilingual definitions) or DBpedia (general concepts). The terms have been hierarchically arranged, spanning multiple languages and targeting different jurisdictions. The term bank has been published as a TBX dump file and is publicly accessible as linked data. The term bank has been used to annotate common licenses in the RDFLicense dataset.

Introduction

Legal translations, namely the *translations of texts within the field of law*, are among the most difficult types of translations. The legal system referred by the source text may be different from the legal system referred by the target text, and the translation of the parts with a specific legal significance must be particularly precise at ensuring the correspondence of concepts at both sides. The mistranslation of a clause in a contract can lead to lawsuits or loss of money.

A *term bank* (also known as *term base* or more informally as *terminology*) is a database of concepts and terminological data related to a particular field. Terminologies help keeping translations consistent and help choosing the most adequate term when precision is required. Further, the *localization* of legal texts require of specialized terminologies where the exact concept in a legal system must be invoked.

The work presented in this paper describes a terminology created in a half-automated process, where terms and their definitions have been extracted and integrated from different lexical sources and mapped in a supervised process.

The resulting terminology has been published[2] in the TBX format –ISO 30042 [1]– which is the standard for the exchange of terminologies; and it has also been

---

[1] Corresponding author: vrodriguez@delicias.fi.upm.es
[2] The copyright terminology is online at: http://cosasbuenas.es/s/copyrighttermbank

published in Resource Description Format (RDF)[3], according to the schema described by Cimiano et al. [4]. The RDF version is especially suitable for establishing links with other resources (like DBpedia[4]) and with other terminologies. IATE[5], the inter-institutional database of the European Union (EU), has been taken as the external reference for some of the extracted terms.

Plain texts can be annotated, making reference to concepts or terms in a term bank. This work also presents the text of a license that has been annotated with the terms in the copyright terminology here presented.

The use of a terminology of legal terms found in licenses is not exhausted with the mere translation or localization. Once in a digital format, it can alleviate the task of identifying the key elements in new licenses as in [5] or can help the study of comparative law.

The paper is organized as follows. Section 1 describes the motivation for having a term bank of copyright-related terms published as linked data. Details on the followed methodology and publication are given in Section 2; Section 3 provides the related work and finally Section 4 contains the conclusions.

1. Motivation: legal term banks as linked data

The representation of copyright and related rights constitutes a part of legal knowledge currently at the limelight of European policy. Progress has been made in delivering copyright-related actions identified in the Digital Agenda[6], the Intellectual Property Strategy[7] and in the "Licences for Europe"[8]. Moreover, it is worth to consider the "Draft Report of the European Parliament"[9,10] towards an harmonisation of copyright across Europe and updating it to meet current challenges, which safeguards fun-damental rights and enables the offer of innovative online services in the EU. Endeavours like the terminology presented in this paper pave the way towards such harmonisation.

The complexity of the regulatory system in this field, together with the variety of the corpus of copyright (patchwork of international and European sources, such as the

---

[3] http://www.w3.org/RDF/

[4] http://dbpedia.org/

[5] http://iate.europa.eu/

[6] Communication on content in Digital Single Market (COM(2012) 789 final).

[7] In order to modernise the EU copyright legislative framework, "*A Single Market for Intellectual Property Rights*" (COM(2011) 287 final) was announced, which proposed series of measures to promote an efficient copyright framework for the Digital Single Market that include short and long-term key policy actions in various areas: patents, trademarks, geographical indications, multi-territorial copyright licensing, digital libraries, IPR violations, and IPR enforcement by customs.

[8] As a premise for a cultural policy and from a structured stakeholder dialogue, industry-led solutions were put forward by stakeholders as a contribution to improve the availability of copyright-protected content online in the EU. Available at http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52012DC0789

[9] Draft Report on the implementation of Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society (2014/2256(INI)), 14/01/2015. A legislative proposal on copyright reform is expected for September this year 2015.

[10] In particular, the Report calls for the harmonization of copyright terms and exceptions across Europe, new exceptions for emerging use cases like audio-visual quotation, e-lending and text and data mining, as well as the adoption of an open norm to "*allow for the adaptation to unanticipated new forms of cultural expression*".

Berne Convention for the Protection of Literary and Artistic Works, the *WIPO Copyright Treaty*, the *Directive 2001/29/EC*[11] (Copyright Directive), amongst other correlated sources[12]), poses difficulties to search, retrieve and understand the legal information in this domain. Moreover, in a pluralistic legal order [14] the "*[EU] legislation is drafted in several languages and [...] the different language versions are all equally authentic. An interpretation of a provision of [EU] law thus involves a comparison of the different language versions*"[13], in accordance with the principle of linguistic equality[14], which entails a "full multilingualism" [15]. Settled case-law refers that '*the need for a uniform interpretation of [EU] regulations makes it impossible for the text of a provision to be considered in isolation but requires, on the contrary, that it should be interpreted and applied in the light of the versions existing in the other official languages* […] [A]*ll the language versions must, (...) be recognised as having the same weight* ".[15]

However, due to the factors that act as constraints in particular judgments, "*limited multilingualism*" seems a more realistic approach [24]. Besides, identifiable hindrances prevent cross-border access to legal information:

⸺ Disclosure of open data makes difficult to retrieve relevant and useful information due to its overload and oversupply (large assortments of data);

⸺ Legal documents are published as plain text without hyperlinks to the official legal resources, averting navigation and reasoning among documents; national and EU websites are sometimes poorly interconnected or they use different identification systems;

⸺ Data is not always published in machine readable formats like XML or RDF for Linked Open Data, but in heterogeneous, non standard formats;

⸺ Ambiguity and polysemy of legal terms [6]: the terminological misalignment and the conceptual misalignment [9] between the terminology used at the EU level from that of the national level, even when implementing EU directives [7];

⸺ Context-specificity of legal terms: the meaning of terms is related to the context of the legislation defining it (several context-specific definitions of legal terms with a common thread)[16];

⸺ Cultural-specificity of legal terms: the meaning of terms is related to the context of the legal and political culture to which these terms belong to (think not only of the

classical distinction between Common and Civil law countries, but of cultures with ideographic languages such as Chinese and Japanese) [8].

The need for cross-border multilingual access to legislation is required for legal practitioners, such as judges, lawyers, translators, legal drafters and scholars, but also to other decision-makers, amongst enterprises, public administrations and citizens, subject to regulatory compliance (even outside their own area of expertise and also jurisdiction), in order to: (i) exploit legal (open) data and therefore produce new innovative services for the legal information provision market; (ii) to predict the impact of implementing the EU legislature in each member state by enriching [7] structurally the documents (with navigable references along legal texts) and semantically (with concepts from ontologies and annotations); (iii) enhance information retrieval, automatic translation, automated reasoning; (iv) ensure the principle of legal certainty; (v) possibly strengthening the textual (or literal interpretation), and teleological interpretation upon which the European Court of Justice (ECJ)'s reasoning primarily rests.

Our work integrates the Linked Open Legal Data [16] momentum that illustrates "*the accessibility and semantic interoperability of legal sources*" [17]. Some of the advantages for rendering multi-lingual, multi-jurisdictional legal term banks published as linked data are:

— Clear separation and identification of concepts and terms, as data fits a formalized model and every resource is identifiable in a permanent manner.
— Easy browsing from a term in one language to an equivalent term in another language, although this makes only full sense when a preferred term is specified [18].
— Easy browsing among general terms and the jurisdiction-specific terms, as concepts can be hierarchically organized. This clarity helps towards the harmonization of copyright terms in the EU, an explicit goal in the EU copyright roadmap [19].
— Easy comparative analysis, as multiple sources are provided.
— Improved discovery and unequivocal identification of concepts and corresponding terms at both European and national levels.
— Better organization of conceptual domain knowledge and its availability of interrelated data sets on the Web in standard formats.

2. Linked resources and methodology

2.1. Publication format

In order to build the present linked term bank, several resources have been considered.

---

[17]European Council, Draft Strategy on European e-justice 2014–2018, 2013 (2013/C 376/06).

[18] For example, IATE defines *preferred term* as: "*a term which should be used instead of any other (equally correct) synonym(s) present, for harmonisation purposes*"

[19] For a explicit mention, see the "Draft Report on the implementation of Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society (2014/2256(INI))"

a) **WIPO** (World Intellectual Property Organization) publishes documents which include glossaries on copyright-related terms[20]. WIPO is an especially authoritative source as the custodian of the treaties on copyright signed by almost every country. Given their almost-universal validity, the definitions provided by WIPO are attributed thus to the most general concepts.

b) IATE (Inter-Active Terminology for Europe) is the current EU's inter-institutional terminology resource database, created from several preexisting databases like EURODICAUTOM (Commission), TIS (Council) and EUTERPE (Parliament), among others. IATE is managed by representatives from different institutions including the authoritative entities like the ECJ or the Translation Centre for the Bodies of the European Union. IATE contains more than 8 million terms in all official 24 EU languages. It has been recently published as a linked data resource [4].

c) **Creative commons licenses** are text documents published along with the referred work, and usually symbolized by icons, summaries or hyperlink references. Creative Commons licenses, massively adopted by the internet culture, have been published in versions tweaked for up to 60 different jurisdictions and different languages. These licenses commence with the definition of the key terms, which typically address the ones used in the target jurisdiction. From version 4.0[21], Creative Commons aimed at a neutral text, capable of fitting every legal system. Consequently, these definitions have been added to the general concepts and not to the jurisdiction-specific concepts.

d) Other resources. Finally, the term bank can be linked to other linked resources to make it a highly connected linked data resource. In particular, the term bank of copyright-related resources has been linked to DBpedia, the linked data version of Wikipedia, and Lexvo.org[22], a dataset of entities about language. The use of these resources was possible as they had been published under open licensing modalities.

The methodology followed to create the term bank has been the following:

1. **Collection of top concepts**. Key copyright-related concepts have been extracted along with a general definition from the WIPO glossaries.
2. **Mapping to IATE**. The linked data version of IATE version was systematically queried in search of direct matches. From the different sources of IATE, the legal one was preferred over others when more than one term matched. The resulting links were verified and completed manually by inspecting the official IATE place[23].
3. **Addition Creative Commons terms**. Over 100 creative commons terms have been defined, including the different versions, different jurisdiction ports and different languages. These resources are well classified in the RDFLicense dataset[24] [3], which also provides the links between license identifiers and legal texts. Creative Commons issued versions of the same license adapted to different jurisdictions before their version 4.0. Definitions from version 4.0 were added to the general concepts. The publication style of Creative Common licenses favors its automatic parsing and the formatting codes can be easily removed.

---

[20] http://www.wipo.int/edocs/pubdocs/en/copyright/891/wipo_pub_891.pdf

[21] "*version 4.0 international license [...] is the most up-to-date version of our licenses, drafted after broad consultation with our global network of affiliates, and it has been written to be internationally valid.*", http://creativecommons.org/version4

[22] http://www.lexvo.org/

[23] http://iate.europa.eu/

[24] http://rdflicense.appspot.com/

The publication of the dataset was made according to the linked data publication guidelines[25] and those specific for term bases [4].

2.2. The copyright term bank

The information in term banks is usually arranged as depicted in Figure 1, following the principle of strict separation between abstract concepts and the terms referring to them. More than one term is possible for the same concept, even in the same language.
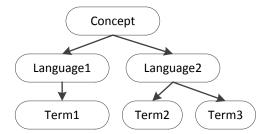

Figure 1. Concepts and terms in terminology databases.

Concepts are typically accompanied by definitions, whereas terms are sometimes provided with additional information like the source, reliability, domain, additional notes, comments and the context of use.

In order to build our term bank, the structure has been extended to tackle the multi-jurisdiction information that is provided, and jurisdiction-specific concepts have been arranged as subspecies of general concepts (Figure 2).
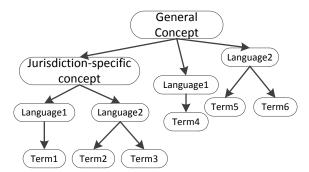

Figure 2. Concepts in a legal terminology database.

2.3. Publication format

TermBaseeXchange (TBX) is the industry XML standard language used to represent terminology data, sometimes used as native format, sometimes as interchange format. It is published by ISO as standard ISO30042 and by the Localization Industry

---

[25]http://www.w3.org/TR/ld-bp/

Standards Association (LISA). The following excerpt shows an example of the published terminology in TBX:

```
<termEntry id="Derivativework (ES)>
 <langSetxml:lang="es">
  <tig>
   <term>obra derivada</term>
   <termNote type="termType">fullForm</termNote>
   <descrip type="reliabilityCode">3</descrip>
  </tig>
 </langSet>
 <langSetxml:lang="ca">
  <tig>
   <term>obra derivada</term>
   <termNote type="termType">fullForm</termNote>
   <descrip type="reliabilityCode">3</descrip>
  </tig>
 </langSet>
</termEntry>
```

Code excerpt 1. Fragment of a term bank in the industry-standard TBX format (ISO30042)

For an advanced format where linking to other resources is made more straightforward, the RDF data structure as in [4] has been chosen. In order to represent the linguistic information, we have adopted the *ontolex-lemon* model [2][23], whose representative schema is shown in Figure 3. OntoLex is based on the ISO Lexical Markup Framework (LMF) and is an extension of the lemon model (LExicon Model for ONtologies). The specification of *ontolex-lemon* is currently under finalization by the W3C Ontolex Community Group[26] but it is already a *de facto* standard to represent ontology lexica.

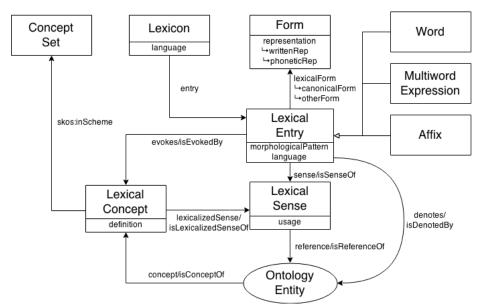[26] https://www.w3.org/community/ontolex/wiki/Final_Model_Specification

Figure 3. The Ontolex-lemon model. Boxes denote OWL classes. The upper part of the boxes contains the class name and the lower part contains the name of datatype properties. Black arrows denote object properties, the white arrow denotes derivation and the symbol ↳ denotes a subproperty relationship.

The example in the excerpt that follows shows two concepts: the universal concept of "derivative work" (lines 7-13) and the concept of "derivative work" in particular in the Spanish jurisdiction (lines 14-20). "Derivative work" is a general concept (`skos:Concept`) that can be linked to the corresponding IATE concept (74645) and even to a DBpedia resource ("Derivative_work"). "Derivative work (ES)" is an abstract concept enshrined in 5 terms in 5 languages (Spanish, Catalan, Galician, Basque, Aranese) for which the creative commons licenses have a translation of the Spanish port. One of these terms is shown in lines 21-25, "obra derivada" in Galician language. The convention of using uppercase for denoting the country code of a jurisdiction has been used, as well as using lowercase to denote the language code.

```
01 @prefix rdfs:<http://www.w3.org/2000/01/rdf-schema#> .
02 @prefix skos:<http://www.w3.org/2004/02/skos/core#> .
03 @prefix tbx:<http://tbx2rdf.lider-project.eu/tbx#> .
04 @prefix ontolex:<http://www.w3.org/ns/lemon/ontolex#> .
05 @prefix dct:<http://purl.org/dc/terms/> .
06 @prefix ctb:< http://tbx2rdf.lider-project.eu/converter/resource/cc/> .

07 ctb:derivative_work
08   a skos:Concept;
09   rdfs:label "derivative work";
10   skos:definition "a new work that translates or transforms one or more
 original copyrighted pre-existing works"@en;
11   dct:source "WIPO";
12   owl:sameAs <http://dbpedia.org/resource/Derivative_work> ;
13   skos:closeMatch <http://tbx2rdf.lider-project.eu/data/iate/IATE-74645> .

14 ctb:derivative_work (ES)
15   a skos:Concept;
16   rdfs:label "derivative work (ES)";
```

```
17    cc:jurisdiction <http://dbpedia.org/resource/Spain> ;
18    skos:narrower ctb:derivative_work ;
19    ontolex:isDenotedBy ctb:obra_derivada_gl, ctb:lan_eratorri_eu, ctb:%C3%B2b
ra_derivada_oci, ctb:obras_derivadas_es, ctb:obra_derivada_ca ;
20    skos:definition "e. Consideraranse obras derivadas aquelas obras creadas a
 partir da licenciada, como por exemplo: as traducións e adaptacións; as revisi
óns, actualizacións e anotacións; os compendios, resumos e extractos; os arranx
os musicais e, en xeral, calquera transformación dunha obra literaria, artístic
a ou científica. Para evitar a dúbida, se a obra consiste nunha composición mus
ical ou gravación de sons, a sincronización temporal da obra cunha imaxe en mov
emento (synching) será considerada como unha obra derivada para os efectos dest
a licenza."@gl .

21    ctb:obra_derivada_gl
22     a ontolex:LexicalEntry;
23     ontolex:lexicalForm [ontolex:writtenRep "obra derivada"@gl ] ;
24     dct:source <http://creativecommons.org/licenses/by/3.0/es/legalcode.gl>;
25     tbx:reliabilityCode "3" .
```

Code excerpt 2. RDF Turtle serialization of one general concept (Derivative work), its derived concept for the Spanish legislation (Derivative_work (ES)) and one of its lexical entries ("obra derivada") in the galician language. To improve the legibility, the chars '%20' in the namespaced URIs have been replaced by a blankspace. Equivalently, parentheses have been introduced.

3. Qualified translations

The above RDF representation based on *lemon* supports the modeling of copyright and related rights from a multilingual perspective. In this way, translations among different lexical representations of terms, expressed in different natural languages, can be inferred by traversing the RDF graph through their common ontolex:LexicalSense. For instance one can translate "obra derivada" from Galician into Spanish by pivoting on their common sense [27] in the above example. However, the meta-operational relationship between legal reference and coreference has to be worked out.

However, this method does not account for the specific type of linguistic translation that is taken place (e.g., literal translation, cultural equivalence, etc.). There exist, however, an extension of the *lemon* model, the so called *lemon translation module* [28], that reifies the translation relation and allows associating additional information to it, such as type of translation, confidence degree, provenance, and even the directionality of the relation [23]. This module has been integrated in the new Ontolex-lemon model as part of the new *vartrans* module.

In the case of the copyright term bank, using such mechanism to represent translations allows distinguishing between term descriptions that are a literal translation one from the other (for instance "obra derivada" in the previous example) from other situations in which the translated description has been adapted to the cultural or jurisdictional specificities of the target language or legal system. This might be beneficial for future semantic-aware applications in the legal domain. For instance, when legal terminology has to be compared across language, it can be done within the same jurisdictional domain, thus being a *literal translations* acceptable, or across

---

[27] http://tbx2rdf.lider-project.eu/converter/resource/cc/derivative%20work%20%28ES%29

[28] http://purl.org/net/translation

jurisdictions, in which *legal equivalents* (rather than literal translations) have to be found.

The use of the *vartrans* module is exemplified in the following code excerpt. The lexical entries have two senses, which related by means of the reference to a common concept. The translation is reified and can be qualified as `trdcat:directEquivalent` or similar.

```
01 @prefix vartrans: <http://www.w3.org/ns/lemon/vartrans#> .
02 @prefix trcat: <http://purl.org/net/translation-categories#>

03 ctb:derivative work (ES)
04   a skos:Concept;
05   rdfs:label "derivative work"@en ;
06   cc:jurisdiction <http://dbpedia.org/resource/Spain> .

07 ctb:lan eratorri eu
08  a ontolex:LexicalEntry;
09  ontolex:lexicalForm [ontolex:writtenRep "lan eratorri"@eu ] ;
10  ontolex:sense <http://example.org/sense_1>  .

11 ctb:obra%20derivada gl
12   a ontolex:LexicalEntry;
13   ontolex:lexicalForm [ontolex:writtenRep "obra derivada"@gl ];
14   ontolex:sense <http://example.org/sense_2>  .

15 <http://example.org/sense_1> ontolex:reference ctb:derivative_work_(ES) .
16 <http://example.org/sense_2> ontolex:reference ctb:derivate_work_(ES) .
17 <http://example.org/sense_1-sense_2-trans> a vartrans:Translation ;
18   vartrans:relates <http://example.org/sense_1> ;
19   vartrans:relates <http://example.org/sense_2> ;
20   vartrans:category trcat:directEquivalent .
```

Code excerpt 3. Example of use of the ontolex-lemon vartrans module. Prefixes from Code excerpt 1 also apply.

4. Related Work

In the literature, different methods exist for approaching the multilingual complexity of European law, for example controlled vocabularies, implemented in terminology database (such as IATE run by all the main EU Institutions that we have resort to in our work), thesauri (as EUROVOC), semantic lexicons or lightweight ontologies(as WordNet, EuroWordNet and, in the legal domain, JurWordNet) that we evoke here. EuroVoc Thesaurus [29] is the most important multilingual, multidisciplinary standardized thesaurus created by the EU, covering the activities of the EU. EuroVoc is managed by the Publications Office, which moved forward to ontology-based thesaurus management and semantic web technologies conformant to W3C recommendations as well as latest trends in thesaurus standards. However, EuroVoc represents a wide-coverage and faceted thesaurus built specifically for processing the

---

[29] http://eurovoc.europa.eu/drupal/

documentary information of the EU institutions: the legal terminology is quite poor and limited to the legal fields belonging to the competence of EU.

The CELLAR repository provides semantic indexing, advanced search and data retrieval for multilingual resources to the information system of the Publications Office of the European Union information system. Resources and their Functional Requirements for Bibliographic Records (FRBR) embrace both the web of data perspective and the library or "bibliographic" data perspective [22]. Its new ontology development assumes that "the FRBR classes are collectors of resource metadata at their specific taxonomy level", thus, allowing a direct constant access to the FRBR levels [22, p. 35]. This represents certainly an improvement over the existing model, as it enhances the accessibility of the OP multilingual documents. However, its scope is also limited to the vocabulary of EU documents.

The Legal Taxonomy Syllabus [6] is a multilevel, multilingual ontology that takes a comparative law perspective to the modeling of legal terms and concepts from EU Directives, helping to increase European terminological consistency. Syllabus is an open-access database linking European terms with national transposition law and also linking terms horizontally (i.e., between national legal orders).

LexALP [20] uses a technique defined for general lexical databases to achieve cross language interoperability between languages of the Alpine Convention. This multilingual legal information system combines three main components, i) a terminology data base, ii) a multilingual corpus, and iii) the relative bibliographic database. In this way the manually revised, elaborated and validated (harmonised) quadrilingual information on the legal terminology (i.e. complete terminological entries) will be closely interacting with a facility to dynamically search for additional contexts in a relevant set of legal texts in all languages and for all main legal systems involved.

The multilingual lexical database version of WordNet, EuroWordNet [19], compounds wordnets expressing lexica of 8 European languages. The wordnets are structured in terms of synsets (sets of synonymous words). Each synset in the monolingual WordNets is linked to the others by cross-lingual equivalence relations to the English synsets recorded by the Inter-Lingual-Index (ILI). The database can be used for monolingual and cross-lingual information retrieval. The LOIS [19] database is compatible with the EuroWordNet architecture, and forms an extension of the EWN semantic coverage into the legal domain. Within this framework, LOIS contributes to the creation of a European Legal WordNet.

5. Conclusions

We have perceived a particular European policy deference towards rendering the copyright and related rights domain more accurate. We have framed some of the advantages for yielding a multi-lingual, multi-jurisdictional legal term bank published as linked data in this domain. Therefore our work presents an effort to achieve a technical and semantic interoperability among linguistic domain concepts.

However, creating a term bank of legal terms is a time-consuming task where expertise in the law of different countries is needed and even domain-specific terminologies require a considerable effort. Legal terminologies, legal concepts and legal knowledge are not synonymous.

Several problems might be raised: (i) ISO standards secure the exchange of terminologies but do not manage the *legal value* of such terminologies, (ii) as said, Version 4.0 of Creative Commons aims at a neutral text, capable of fitting every legal system, but nothing prevents legal operators (e.g. judges) to offer different interpretations of general concepts at the jurisdiction-specific level; (iii) the term "bank" related to other linked resources such as *DBpedia* or *Lexvo.org* entails a more careful examination of this kind of relationship, as the *valence* (i.e. the number of edges incident to the vertex) of resources might not be equivalent, and in fact they are not.

Defining `owl:sameAs` relationships requires a more careful examination of the functional entrenchment of legal sources [21], a more extended comparative work of both legal and cultural systems [22] and and a closer attention to limited multi-lingualism in EU and national Courts [24]. This work will be completed in the next future with the annotations of a complete dataset of existing licenses.

References

[1] ISO 30042:2008 (2008). Systems to manage terminology, knowledge and content – TermBase eXchange.

[2] J. McCrae, G. Aguado-de Cea, P. Buitelaar, P. Cimiano, T. Declerck, A. Gómez-Pérez, J. Gracia, L. Hollink, E. Montiel-Ponsoda, D. Spohr, and T. Wunner, "Interchanging lexical resources on the semantic web," Language Resources and Evaluation, vol. 46, pp. 701-719, 2012.

[3] Rodriguez-Doncel, V., Villata, S., and Gomez-Perez, A. (2014). A dataset of RDF licenses. In Hoekstra, R., editor, Proceedings of the 27th International Conference on Legal Knowledge and Information System, pp.–189.

[4] Cimiano, P., McCrae, J., Rodriguez-Doncel, V., Gornostay, A. Gomez-Perez and B. Simoneit. Linked Terminology: Applying Linked Data Principles to Terminological Resources, in Proc. of the 4th Biennial Conference on Electronic Lexicography (to appear 2015)

[5] Cabrio, E., Aprosio, A. P., and Villata, S. (2014). These are your rights: A natural language processing approach to automated RDF licenses generation. In The Semantic Web: Trends and Challenges, pp. 255–269. Springer

[6] G. Ajani, G. Boella, L. Lesmo, A. Mazzei, and P. Rossi. Terminological and ontological analysis of european directives: multilinguism in law. In 11th Int. Conf. on Artificial Intelligence and Law (ICAIL), pp. 43–48, 2007.

[7] G. Ajani; G. Boella; M. Martin; A. Mazzei; D. P. Radicioni; P. Rossi, Legal Taxonomy Syllabus 2.0 , in "3rd Workshop on Legal Ontologies and Artificial Intelligence Techniques joint with 2nd W. on Semantic Processing of Legal Texts", June 8th, 2009, Barcelona, Spain.

[8] Makoto, N., Ogawa, Y. and Toyama, K. "Extraction of Legal Definitions and Their Explanations with Accessible Citations." P. Casanovas, U. Pagallo, M. Palmirani, G. Sartor, *AI Approaches to the Complexity of Legal Systems, AICOL-V,* . Springer Berlin Heidelberg, 2014. 157-171.

[9] L. Lesmo, G. Boella, A. Mazzei (2006); Multilingual Conceptual Dictionaries Based on Ontologies: Analytical Tools and Case Studies. In Proc. Of V Legislative XML Workshop, pp. 1-14, EPAP

[10] J. McCrae, D. Spohr, P. Cimiano (2011). Linking lexical resources and ontologies on the semantic web with lemon. In Proc. of the 8th extended semantic web conference, pp. 245-259, G. Antoniou et al. (Eds.), Vol. Part I. Springer-Verlag

[11] V. Lyding, E. Chiocchetti, G. Sérasset, and F. Brunet-Manquat(2006). The LexALP information system: term bank and corpus for multilingual legal terminology consolidated. In Proc. of the Workshop on Multilingual Language Resources and Interoperability (MLRI '06). Association for Computational Linguistics, Stroudsburg, PA, USA, 25-3

[12] Kea European Affairs (2010) Study concerning Multi-territory licensing for the online distribution of audiovisual works in the European Union. Final Report Prepared for the European Commission, DG Information Society and Media.

[13] Francesconi, E.; Spinosa, P.; Tiscornia, D. A linguistic-ontological support for multilingual legislative drafting: the DALOS Project, P. Casanovas et al. (Ed.) LOAIT June 4th, 2007, Stanford University, Stanford, CA, USA. CEUR-WS 321, 2008: 103-11.

[14] Poiares Pessoa Maduro, L.M., Interpreting European Law: Judicial Adjudication in a Context of Constitutional Pluralism, European Journal of Legal Studies, 2007, 1, 2.

[15] D. Hanf and E. Muir, Le droit de l'Unioneuropéenne et le multilinguisme, in D. Hanf, E. Muir et K. Malacek (eds), *Langue et construction européenne* (Cahiers du Collège d'Europe, Bruxelles, 2010) at 23.

[16] Casellas, N., Linked Legal Data: A SKOS Vocabulary for the Code of Federal Regulations', available at: http://www.semantic-web-journal.net/system/files/swj311_2.pdf

[17] Sartor, G., Legislative Information and the Web, Legislative XML for the Semantic Web, Principles, Models, Standards for Document Management, Sartor, G., Palmirani, M., Francesconi, E., Biasiotti, M.A. (Eds.), Law, Governance and Technology Series, Springer, 4, 2011.

[18] Kemp, R., Legal aspects of managing Big Data, Computer Law & Security Review, Volume 30, Issue 5, Pages 463-614 (October 2014).

[19] Peters W., Sagri M.T., Tiscornia D., (2007), The structuring of legal knowledge in LOIS, Artificial Intelligence andLaw, 15:117-135.

[20] Lyding, V., et al., The LexALP Information System: Term Bank and Corpus for Multilingual Legal Terminology Consolidated, Proceedings of the Workshop on Multilingual Language Resources and Interoperability, pages 25–31, Sydney, July 2006.

[21] Casanovas, P., Casellas, N., and Vallbé, J.J.. "Empirically Grounded Developments of Legal Ontologies: A Socio-Legal Perspective." P. Casanovas, G. Sartor, M.A. Biasiotti, M. Fernández-Barrera (Eds) *Approaches to Legal Ontologies*. Springer Netherlands, 2011. 49-67.

[22] Francesconi, E., Küster, M.W., Gratz, P. and Thelen, S.. "The Ontology-Based Approach of the Publications Office of the EU for Document Accessibility and Open Data Services." In *Electronic Government and the Information Systems Perspective*, pp. 29-39. Springer International Publishing, 2015.

[23] J. Gracia, E. Montiel-Ponsoda, D. Vila-Suero, and G. Aguado-de Cea, "Enabling language resources to expose translations as linked data on the web," in Proc. of 9th Language Resources and Evaluation Conference (LREC'14), Reykjavik (Iceland). European Language Resources Association (ELRA), May 2014, pp. 409-413

[24] M, Derlén, *Multilingual Interpretation of European Union Law*, Alphen aan Den Rijn, Kluwer Law International, 2009