Knowledge of the Law in the Big Data Age
G. Peruginelli and S. Faro (Eds.)

doi:10.3233/FAIA190019

Language Resources as Linked Data for the Legal Domain

Patricia MARTÍN-CHOZAS,
Elena MONTIEL-PONSODA and Víctor RODRÍGUEZ-DONCEL
Ontology Engineering Group, Universidad Politécnica de Madrid (Spain)

Abstract. This Chapter describes a four-stage methodology to generate Linguistic Linked Data for the legal domain: identification, creation, transformation (to RDF) and linking. The goal of this process is to enhance the presence of legal language resources in the Linguistic Linked Open Data cloud. Since this Chapter is framed within the H2020 LYNX project, aimed at creating a Legal Knowledge Graph, a parallel objective is to employ the resources generated as a linguistic foundation to annotate, classify and translate the legal resources represented in this graph.

Keywords. legal language resources, terminology extraction, linked data, legal knowledge graph

1. Introduction

Originally, *language resources* have been considered as works that collect any type of linguistic information. For the purposes of this work, we define language resources as *pieces of data containing linguistic information in machine readable forms*. There are several types of language resources depending on their format and the type of information represented: glossaries and terminologies (specialised terms), lexical databases (linguistic knowledge for computers), dictionaries (general terms), thesauri (hierarchical controlled vocabularies), etc. Many general dictionaries are available online, such as Merriam Webster[1] and Oxford Dictionary[2]; other terminological resources containing specialised knowledge can also be found on the Internet, such as TermSciences[3] and UNterm[4]. However, language resources for the legal domain are not that present in the Web, since they tend to be owned by legal publishers, thus, not accessible and sometimes published in obsolete and proprietary formats. Moreover, *legal jargon* is intricate and the meaning of terms varies as the legal framework changes. Updating non-machine readable legal glossaries is a time-consuming and difficult task to accomplish. On the other hand, a good understanding of legal terminology is essential to comprehend legal documentation, which also tends to be outdated.

To soften the mentioned hindrances regarding legal terminology and legal documentation, the LYNX project aims at creating a *Legal Knowledge Graph (LKG)*, that is in-

[1] https://www.merriam-webster.com/.
[2] https://www.oxforddictionaries.com/.
[3] http://www.termsciences.fr/termsciences/?lang=en.
[4] https://unterm.un.org/UNTERM/portal/welcome.

terlinking public and private legal resources, metadata, standards and general open data from the legal domain. The idea is to offer access to updated multilingual and multi-jurisdictional legal information. For that purpose, a steady open-access legal language foundation is required.

Linked Data [1] is a particularly convenient form to create such a language cloud, since it is intended to publish interlinked machine-readable data in open-source and non-proprietary formats. In fact, the *Linguistic Linked Open Data (LLOD)*[5] cloud gathers language resources published according to the Linked Data Principles [2], following the RDF set of W3C specifications. Again, within this cloud, legal knowledge is underrepresented. The objective of this contribution is to create the *Linguistic Legal Linked Open Data* (LLLOD) cloud that starts covering the legal knowledge gap. This new legal cloud will set the language foundations to annotate, classify and translate the legal documents taking part of the LKG.

2. Related Work

This Section is divided into three parts: a survey of language resources published as Linked Data, an analysis of those that belong to the legal domain and a summary of the most common models to represent linguistic information.

2.1. Language Resources in RDF

The two main efforts to store and share linguistic information are WordNet [3], a lexical database in English, and BabelNet, an extensive multilingual encyclopedic knowledge base [4]. However, apart from these two knowledge bases, there are many other projects devoted to publishing language resources as per the Semantic Web standards.

One of the most relevant works in this field is the conversion of *IATE* into RDF, the terminological database of the European Union, originally built in TBX (TermBase eXchange format), but unofficially ported to RDF. In this transformation, the *SKOS* vocabulary was used to model term entries with the `skos:concept` property and term relations with `skos:broader` and `skos:narrower` properties. Likewise, *Ontolex* model was used to represent lexical information and term translations thanks to the `ontolex:reference` property [5].

Terminoteca RDF gathers two sets of resources: *Terminesp*, a multilingual terminological database developed by the Spanish Association for Terminology[6]; and terminological glossaries from the *Terminología Oberta* service of the Catalan Terminological Centre[7] (*TERMCAT*). Since the datasets contained in Terminoteca RDF are multilingual, the `vartrans` module of *Ontolex* was a substantial part of the data modeling stage, resulting in a multilingual repository of linked terminologies[8] [6]; [7].

A similar work was made for the conversion into RDF of the bilingual dictionaries used in the *Apertium*[9], free-open-source machine translation system, supported by the Spanish Government and several Spanish universities [8]. The dictionaries are now published as Linked Data following the *lemon* model as part of the *LLOD* cloud [9].

[5]http://linguistic-lod.org/llod-cloud.
[6]http://www.aeter.org/.
[7]http://www.termcat.cat/en.
[8]http://linguistic.linkeddata.es/terminoteca/.
[9]https://www.apertium.org.

Additionally, one of the biggest resources shaping the Linguistic Linked Open Data cloud is *AGROVOC*, a multilingual thesaurus modeled with SKOS-XL and composed by more than 35,000 terms in 29 languages [10]. It has been aligned with other thesauri in the *LLOD*, such as *GEMET* for environment, *TheSoz* for social sciences and *STW* for economics by using the `skos:exactMatch` proprety.

Many other projects are focused on publishing linguistic Linked Data, but these are some of the most representative for this work as they present similar content, models, linked resources and languages.

2.2. Legal Language Resources in RDF

Although the presence of legal language resources in the Web of Data is scarce, some projects have been devoted to improve their representation in the cloud.

One of these projects resulted in the generation of a termbank of multilingual and multi-jurisdictional legal data by linking relevant datasets in the domain such as *IATE*, *Creative Common licenses*, *World Intellectual Property Organization (WIPO) documents*, *DBpedia*[10] and *Lexvo*[11] [11].

Another apposite project in this area is *Eunomos*, a legal knowledge management system based on legislative XML and ontologies [12]. For the extraction and modeling of legal concepts, the system relies on the *Legal Taxonomy Syllabus* ontology, for terminology management of the European Directives [13].

One of the most significant resources taking part in this contribution, *EuroVoc*, has also been represented as Linked Data, following the SKOS vocabulary [14]. This multilingual and multidisciplinary thesaurus created and maintained by the Publications Office of the European Commission is now linked with other sound resources at European level, such as the *UNESCO* and the *GEMET* thesauri. EuroVoc is also available through a *SPARQL endpoint*[12], developed by PoolParty[13].

The Publications Office has also developed the *CELLAR repository* to publish part of the bibliographic resources of the European Union gathered in the *EUR-Lex portal*[14]. Such resources are semantically described by an ontology providing open access, long term preservation, indexing and retrieval services [15]. They continue enhancing semantic interoperability by linking multilingual terminologies to build a *Public Multilingual Knowledge Management Infrastructure* within the *ISA2* project [16].

2.3. Models to Represent Linguistic Information

Some of the resources referenced in Section 2.2 and those that are part of the Linguistic Linked Open Data cloud have been represented following different models, depending on the nature of each resource (structure, content, objectives, etc.). Some of the most common models to represent linguistic information are briefly listed as follows:

- *Lemon*, the *Lexicon Model for Ontologies*, is intended to represent lexical information of a given term, such as the sense, form, abbreviation, etc.[17]

[10] https://wiki.dbpedia.org/.
[11] http://www.lexvo.org/.
[12] https://lynx.poolparty.biz/PoolParty/sparql/Eurovoc4.3.
[13] https://www.poolparty.biz/.
[14] https://eur-lex.europa.eu/.

- *Ontolex* is the evolution of *lemon*. Supported by the W3C Ontology-Lexica Community Group[15], it allows the representation of relations amongst senses, forms and translations [5].
- *LIR*, the *Linguistic Information Repository*, was intended for ontology localisation, offering access to multilingual data [18].
- *Lexinfo* associates additional linguistic information to elements in an ontology [19].
- *SKOS*, the *Simple Knowledge Organization System*, structures thesauri and taxonomies, easing the creation of hierarchical relations between terms. It is widely used within the Semantic Web since it can be combined with formal representation languages, such as the *Web Ontology Language* (OWL) [20].

Choosing the most appropriate vocabulary is decisive for the representation of resources on the Web of Data. However, this task is only one of the steps of a reliable *methodology* that should be followed in order to *publish sound resources as per the Linked Data paradigm*. Such methodology stresses the importance of preprocessing the data, choosing a sound URI naming strategy, selecting the right technology for RDF generation and reliably linking with other datasets in the cloud [21].

3. Motivation

The aforementioned methodology also emphasizes the importance of reusing resources. When searching for datasets to reuse in this work, *a lack of legal language resources* has been identified, specifically in the three subdomains that are of our interest in LYNX project: labour law, data protection and industrial standards. To help represent legal domain in the Web of Data, *new linguistic resources need to be created* from scratch by extracting terms from legal corpora.

While the advantages of publishing Linked Data are increasingly gaining attention, the greatest part of the available language resources nowadays are *not in machine readable formats yet*. Furthermore, although governments and public institutions are publishing legislation on the Internet, they barely apply open format standards and generate the most part of the documents in PDF. This practice causes huge hindrances when updating, sharing and reusing resources. For this reason, it is required to *transform these resources into RDF* so they can be interlinked to provide a more efficient access to multilingual and multi-jurisdictional legal information.

It is worth mentioning that many of the linguistic portals and repositories consulted throughout this work were not updated or maintained any longer. *Documentation of resources* is key to store and share up-to-date knowledge. Part of this contribution has also focused on the creation of a data portal to keep track of all resources handled in this work and specially those shaping the first approach of the *LLOD* cloud.

4. Contribution

The four stages in which this methodology is divided cover the gaps mentioned in Section 3 and are structured as follows:

- *Identification of existing language resources* that could be reused in this project.

[15] https://www.w3.org/community/ontolex/.

- *Generation of new of language resources* by the extraction of terms from legal corpora provided by LYNX partners.
- *Conversion into RDF* of several language resources identified in the first stage and others created afterwards.
- *Linking the resulting terminological resources* from the previous stages with other existing datasets in the LLOD.

4.1. Identification of Existing Resources

Three different search strategies have been explored with the aim of spotting potentially useful resources:

- Identification through general web search
- Identification through literature
- Identification through specialised portals (ELRC-SHARE[16], Retele[17], CLARIN[18] and the OLAC Language Resource Catalog[19], amongst others).

Besides AGROVOC, IATE, EuroVoc, STW or GEMET, mentioned in Section 2, several significant language resources have also been identified. The *German Labour Law Thesaurus*, for instance, covers different areas of labour law and it is published as Linked Data. *JuriVoc* is a multilingual thesaurus containing juridical terms hierarchically structured. Likewise, the *TERMCAT* institute published glossaries from the labour law domain in XML, a very convenient format to be transformed into RDF, since they present similar structures.

Therefore, Table 1 gathers the first set of available language resources that could be reused in this project. It is worth mentioning that they present different formats and that not all of them contain legal information but cover many adjacent domains. In addition, some resources from the general domain have also been gathered since they present other interesting features (updates, links with other datasets, etc.).

4.2. Creation of New Resources

The approach followed here can be divided into three different stages:

- Evaluation of term extraction tools.
- Extraction of terms from legal corpora.
- Evaluation of extracted terms.

For the creation of resources, the first task consisted in identifying available *Automatic Term Extraction (ATE)* tools to be tested, in order to choose the one that met the needs of this work. Nine ATE tools have been evaluated: Translated.net[20], VocabGrabber[21], TermSuite[22], TermoStat Web[23], SketchEngine[24], Fivefilters[25], Termine[26], Pootle[27]

[16] https://www.elrc-share.eu/.
[17] http://catalogo.retele.linkeddata.es/.
[18] https://www.clarin.eu/.
[19] http://www.language-archives.org/.
[20] https://labs.translated.net/terminology-extraction/.
[21] https://www.visualthesaurus.com/vocabgrabber/.
[22] http://termsuite.github.io/.
[23] http://termostat.ling.umontreal.ca.
[24] https://www.sketchengine.co.uk/.
[25] http://fivefilters.org/term-extraction/.
[26] http://www.nactem.ac.uk/software/termine.
[27] https://pootle.translatehouse.org/.

and TBXTools[28]. Eight evaluation criteria were considered, including availability of the tool, file formats, type of extracted terms and additional services. After several extraction tests per tool and based on the quality of the results as assessed by expert terminologists, *SketchEngine* was the tool selected to generate new glossaries.

Table 1. Set of available language resources identified

ID	Name	Description	Language
iate	IATE	EU terminological database	EU languages
eurovoc	Eurovoc	EU multilingual thesaurus	EU languages
eur-lex	EUR-Lex	EU legal corpora portal	EU languages
conneticut-legal-glossary	Connecticut Legal Glossary	Bilingual Legal Glossary	en, es
unesco-thesaurus	UNESCO Thesaurus	Multilingual multidisciplinary thesaurus	en, es, fr, ru
library-of-congress	Library of Congress	Legal corpora portal	en
imf	International Monetary Fund	Economic multilingual terminology	en, de, es
eugo-glossary	EUGO Glossary	Business monolingual dictionary	es
cdisc-glossary	CDISC Glossary	Clinical monolingual	en
stw	STW Thesaurus for Economics	Economic monolingual thesaurus	en
edp	European Data Portal	EU datasets	EU languages
inspire	INSPIRE Glossary (EU)	General terms and definitions in English	en
saij	SAIJ Thesaurus	Controlled list of legal terms	es
calathe	CaLaThe	Cadastral vocabulary	en
gemet	GEMET	General multilingual thesauri	en, de, es, it
informea	InforMEA Glossary (UNESCO)	Monolingual glossary on environmental law	en
copyright-termbank	Copyright Termbank	Multi-lingual term bank of copyright-related terms	en, es, fr, pt
gllt	German labour law thesaurus	Thesaurus with labour law terms	de
jurivoc	Jurivoc	Juridical terms from Switzerland	de, it, fr
TERMCAT	TERMCAT	Terms from several fields including law	ca, en, es, de, fr, it
termcoord	Termcoord	Glossaries from EU institutions and bodies	EU languages
agrovoc	Agrovoc	Controlled general vocabulary	29 languages

Terms were extracted from three different corpora provided by LYNX partners, each one representing one area of law: a set of collective agreements from the *labour law domain*, a set of regulations from the *data protection* domain and a set of decisions from the *industrial standards* domain.

As a result, the tool returned a list of 200 candidate terms, single and multi-word, per set of documents. The lists were evaluated by professional terminologists to analyse the quality of the result. Such evaluation has been performed by verifying the terms against well-known terminological databases widely used by language professionals: IATE[29], Linguee[30] and also BabelNet[31].

These checks showed that several candidates were not correctly identified by the tool and part of those that were correctly extracted are not relevant for the legal domain. Therefore, Table 2 gathers the amount of 'clean terms' after the evaluation of each new glossary.

The glossaries have been organised in XLS files with eight columns per entry. Each column represents an attribute of the term (URI, definition, usage note, etc.) that will be represented in the conversion stage as an RDF property.

[28] https://sourceforge.net/projects/tbxtools/.
[29] https://iate.europa.eu/home.
[30] https://www.linguee.es/.
[31] https://babelnet.org/.

Table 2. Term lists after the evaluation stage

Resulting term lists		
Labour Law Glossary (ES)	Data Protection Glossary (EN)	Industrial Standards Glossary (EN)
102 terms	98 terms	109 terms

4.3. Conversion into RDF

The objective of this stage is to transform into RDF the glossaries created in the previous Section and some of the resources identified in Section 4.1. As stated in the mentioned Section, TERMCAT platform gathers glossaries in XML from the labour law domain. Since the LYNX project handles data from this domain, two of them have been selected to be reused, converted and linked in this work. Consequently, five glossaries have been converted into RDF: three new resources created by term extraction from legal corpora (two in English and one in Spanish) and two existing TERMCAT glossaries (one in English and one in Spanish).

Due to the nature, goal and format of the glossaries, SKOS was the vocabulary selected to transform the resources, since it is an intuitive model able to represent all the term attributes contained in the glossaries.

Therefore, Figure 1 exemplifies the representation of a term entry and its attributes as they are structured in the glossaries.

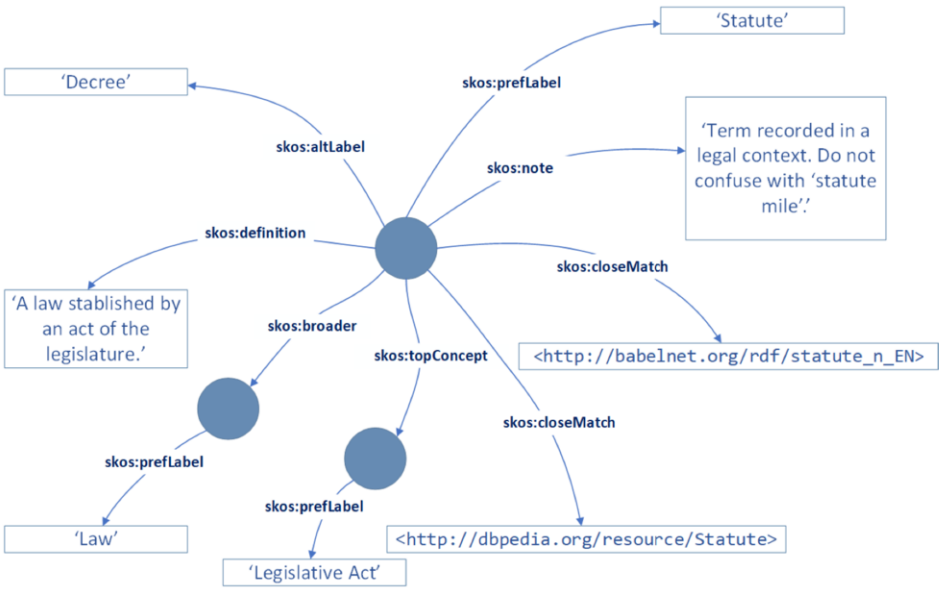

Figure 1. Representation of a term entry with SKOS

The URI of the term is represented by the `skos:concept`. It has been generated by following the URI naming strategy of related work in RDF generation such as the conversion of TERMCAT files [6] and the Apertium Bilingual Dictionaries [8]. These approaches use URL of the server were the resource is located, the part of speech of each

term and the ISO 639-1 language code to build each identifier. An example of a term entry in the glossaries handled here is the following:

http://linguistic.linkeddata.es/terminoteca/lynx/statute-n-en

Finally, the metadata of each glossary (author, creation date, title and description of the resource, etc.) were modeled with the DublinCore[32] ontology.

4.4. Linking Step

Once the glossaries have been represented in RDF, the next step is to generate hyperlinks that connect them with other knowledge bases and linguistic linked resources that are already part of the Linguistic Linked Open Data cloud. These links between resources are exceptionally helpful to share information and enrich the glossaries with context, translations, related terms, usage notes, etc.

Both conversion and linking processes have been performed with OpenRefine[33]. The linking service of OpenRefine can be executed by using either a SPARQL endpoint or an RDF dump of the knowledge base. In this case, the SPARQL endpoint option has been applied since the three involved knowledge bases offer this kind of access: *DBpedia*[34], *EuroVoc*[35] and *BabelNet*[36].

Table 3 shows the results of the linking experiments:

Table 3. Results of the linking tests

Glossaries	BabelNet		EuroVoc		DBpedia		Total terms
Labour Law Glossary ES	47	46%	-	-	8	7.84%	102
Data Protection Glossary EN	70	71%	13	13.27%	59	60.20%	98
Industrial Standards EN	37	33.94%	-	-	70	64.22%	109
Termcat Glossary ES	-	-	97	13.18%	61	8.29%	736
Termcat Glossary EN	-	-	104	13.90%	118	15.78%	748

These results show that the number of links generated is highly dependent on the content of the glossaries and on the type of knowledge base to which they have been linked with. For instance, the lowest percentages appear when linking labour law glossaries with DBpedia, while more technical domains are well represented.

Figure 2 represents the result of this work as the first approach of the Linguistic Legal Linked Open Data cloud (LLLOD).

All the language resources appearing in Figure 2 are duly documented in LYNX Data Portal[37] and all of them can be openly accessed. The number of datasets of the portal is constantly increasing as the LYNX project progresses.

[32] http://dublincore.org/.
[33] http://openrefine.org/.
[34] https://wiki.dbpedia.org/.
[35] http://eurovoc.europa.eu/.
[36] https://babelnet.org/.
[37] http://data.lynx-project.eu/.

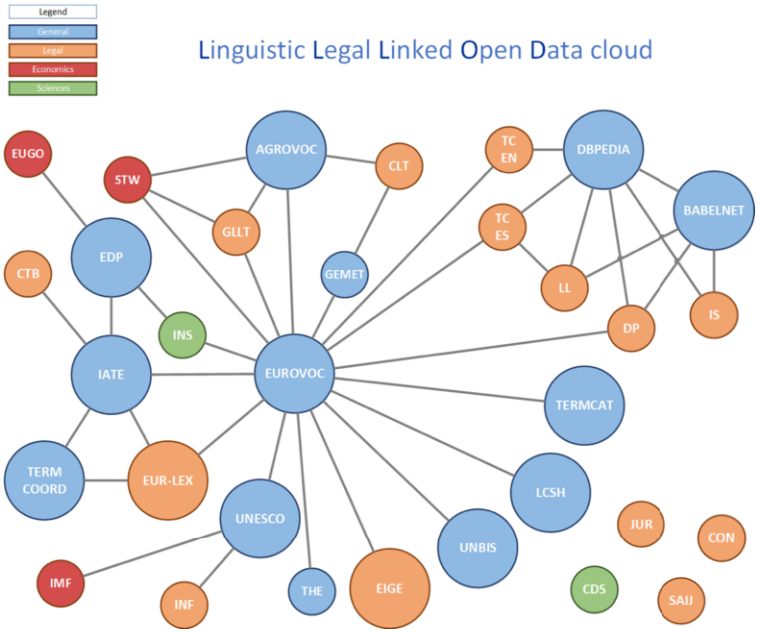

Figure 2. First approach of the Linguistic Legal Linked Open Data cloud

5. Conclusions and Future Work

Legal domain undergoes a lack of language resources in structured and open formats. From the whole set of resources identified, the 39% corresponds to archived resources that might be relevant for the domain but published in obsolete formats or not supported any longer. The next steps on this matter will be focused on publishing the resources identified in the first stage of this methodology as Linked Data to enrich the LLOD.

During the creation of new resources, it was noticed that performance of ATE tools for the legal domain is still limited. These experiments showed that a 40% of the candidate terms were not correctly identified, which means that a huge amount of manual work is still required, unaffordable in large projects. More research on this field is also required in order improve the accuracy of ATE tools on legal corpora and the automation of the whole extraction process.

On the other hand, the current glossaries generated contain monolingual information represented with SKOS vocabulary. In order to add value to these resources, term translations and additional information will also be included. Therefore, other RDF models need to be considered for the representation of this type of relations (e.g. Ontolex `vartrans` module) [22].

Likewise, as mentioned in Section 4, linking tools present several drawbacks that also involve a huge amount of manual work. Consequently, other tools for publishing Linked Data, such as VocBench[38] and Silk[39], will also be tested.

[38] http://vocbench.uniroma2.it/.
[39] http://silkframework.org/.

Acknowledgements

This work has received funding from the EU's Horizon 2020 Research and Innovation programme through the contract LYNX (grant agreement no.: 780602).

References

[1] Berners-Lee, T. (2006). *Design Issues*, https://www.w3.org/DesignIssues/LinkedData.html.
[2] Bizer, C., Heath, T. & Berners-Lee, T. (2009). Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems, 5*(3), 1-22.
[3] Miller, G. A. (1995). WordNet: A Lexical Database for English. *Communications of the ACM, 38*(11), 39-41.
[4] Navigli, R. & Ponzetto, S. P. (2012). BabelNet: The Automatic Construction, Evaluation and Application of a Wide-coverage Multilingual Semantic Network. *Artificial Intelligence, 193*, 217-250.
[5] Cimiano, P., McCrae, J. P., Rodríguez-Doncel, V. et al. (2015). Linked Terminologies: Applying Linked Data Principles to Terminological Resources. In *Proceedings of the eLex 2015 Conference*, 504-517.
[6] Bosque-Gil, J., Montiel-Ponsoda, E., Jorge, G. & Aguado-de-Cea, G. (2016). Terminoteca RDF: A Gathering Point for Multilingual Terminologies in Spain. In Erdman Thomsen, H., Pareja-Lora, A. & Nistrup Madsen, B. (Eds.). *Term Bases and Linguistic Linked Open Data. Proceedings of the 12th International Conference on Terminology and Knowledge Engineering*, 136.
[7] Bosque-Gil, J., Gracia, J. & Gómez-Pérez, A. (2016). Linked Data in Lexicography. *Kernerman Dictionary News, 24*, 19-24.
[8] Gracia, J., Villegas, M., Gómez-Pérez, A. & Bel, N. (2017). The Apertium Bilingual Dictionaries on the Web of Data. *Semantic Web, 9*(2), 231-240.
[9] McCrae, J., Spohr, D. & Cimiano, P. (2011). Linking Lexical Resources and Ontologies on the Semantic Web with Lemon. In Antoniou, G., Grobelnik, M., Simperl et al. (Eds.). *The Semantic Web: Research and Applications. Proceedings of the 8th Extended Semantic Web Conference*. Springer, 245-259.
[10] Caracciolo, C., Stellato, A., Morshed, A. et al. (2013). The AGROVOC Linked Dataset. *Semantic Web, 4*(3), 341-348.
[11] Rodríguez-Doncel, V., Santos, C., Casanovas, P. & Gómez-Pérez, A. (2015). A Linked Term Bank of Copyright-related Terms. In Rotolo, A. (Ed.). *Legal Knowledge and Information Systems. Proceedings of the 28th Jurix Conference*. IOS Press, 91-100.
[12] Boella, G., Humphreys, L., Martin, M., Rossi, P., van der Torre, L. & Violato, A. (2012). Eunomos, a Legal Document and Knowledge Management System for Regulatory Compliance. In De Marco, M., Te'eni, D., Albano, V. & Za, S. (Eds.). *Information Systems: Crossroads for Organization, Management, Accounting and Engineering*. Springer, 571-578.
[13] Ajani, G., Lesmo, L., Boella, G., Mazzei, A. & Rossi, P. (2007). Terminological and Ontological Analysis of European Directives: Multilinguism in Law. In Gardner, A. & Winkels, R. (Eds.). *Proceedings of the 11th International Conference on Artificial Intelligence and Law*. ACM, 43-48.
[14] Díez, L. A., Pérez-León, B., Martínez-González, M. & Blanco, D. J. (2010). Propuesta de representación del tesauro Eurovoc en SKOS para su integración en sistemas de información jurídica. *Scire: representación y organización del conocimiento, 16*(2), 47-51.
[15] Francesconi, E., Küster, M. W., Gratz, P. & Thelen, S. (2015). The Ontology-based Approach of the Publications Office of the EU for Document Accessibility and Open Data Services. In Kö, A. & Francesconi, E. (Eds.). *International Conference on Electronic Government and the Information Systems Perspective*. Springer, 29-39.
[16] Schmitz P., Francesconi E., Hajlaoui N. & Batouche B. (2017). Towards a Public Multilingual Knowledge Management Infrastructure for the European Digital Single Market. In *1st Workshop on the OntoLex Model (OntoLex-2017), Shared Task on Translation Inference Across Dictionaries & Challenges for Wordnets, Proceedings of the LDK 2017 Workshops*, Vol-1899, p. 33-42.
[17] McCrae, J., Aguado-de-Cea, G., Buitelaar, P. et al. (2012). Interchanging Lexical Resources on the Semantic Web. *Language Resources and Evaluation, 46*(4), 701-719.
[18] Montiel-Ponsoda, E., de Cea, G. A., Gómez-Pérez, A. & Peters, W. (2011). Enriching Ontologies with Multilingual Information. *Natural Language Engineering, 17*(3), 283-309.

[19] Cimiano, P., Buitelaar, P., McCrae, J. & Sintek, M. (2011). LexInfo: A Declarative Model for the Lexicon-ontology Interface. *Web Semantics: Science, Services and Agents on the World Wide Web, 9*(1), 29-51.

[20] Miles, A. & Bechhofer, S. (2009). *SKOS Simple Knowledge Organization System Reference*. W3C recommendation, 18.

[21] Vila-Suero, D., Gómez-Pérez, A., Montiel-Ponsoda, E., Gracia, J. & Aguado-de-Cea, G. (2014). Publishing Linked Data on the Web: The Multilingual Dimension. In Buitelaar, P. & Cimiano, P. (Eds.). *Towards the Multilingual Semantic Web*. Springer, 101-117.

[22] Bosque-Gil, J., Gracia, J., Aguado-de-Cea, G. & Montiel-Ponsoda, E. (2015). Applying the Ontolex Model to a Multilingual Terminological Resource. In Gandon, F., Guéret, C. & Villata, S. et al. (Eds.). *The Semantic Web: ESWC 2015 Satellite Events. Proceedings of the European Semantic Web Conference*. Springer, 283-294.