

Interlinking Legal Data

Erwin Filtz, Sabrina Kirrane, Axel Polleres

Vienna University of Economics and Business, Austria
{firstname.lastname}@wu.ac.at

Abstract. In recent years, the European Union has been working towards harmonizing legislation thus allowing for easier cross-border access to, exchange and reuse of legal information. This initiative is supported via standardization activities such as the European Law Identifier (ELI) and the European Case Law Identifier (ECLI), which provide technical specifications for web identifiers and vocabularies that can be used to describe metadata pertaining to legal documents. Unfortunately, to date said initiative has only been partially adopted by EU member states, possibly due to the manual effort involved in curating the metadata. As a first step towards streamlining this process, we propose a cross-jurisdictional legal framework that demonstrates how legal information stored in national databases can be linked at a European level using Natural Language Processing together with external knowledgebases to automatically populate the knowledge base.

1 Introduction

Globalization has increased the number of cross-border cross-jurisdictional activities, bringing with it the need for alignment of and improved accessibility to multilingual legal data across Europe. Therefore, one of the key goals of the European Union (EU) is to harmonize laws and to improve access to legal information. From an information access perspective, standardization activities such as the European Law Identifier (ELI)¹ and the European Case Law Identifier (ECLI)² aim to improve accessibility and integration of legal data by proposing standard Web identifiers and metadata schemas for legislation and court proceedings respectively. From a legislation perspective, we also see a push towards this harmonization in terms of EU-wide *regulations* (laws to be enforced by all member states) and *directives* (which are transformed into national laws). However, there are still a large number of national laws that are not governed by the EU centrally. Most member states record legal information (i.e., legislation and court proceedings in their respective national language) in heterogeneous national legal databases that are usually accessible via Web search interfaces or application programming interfaces.

Although the ELI and ECLI are also relevant for national legislation and court cases, these standards have not yet received widespread adoption, possibly due to the incurring costs in terms of either manual effort involved or changes required to existing systems in order to sustainably implement and maintain these relatively new standards.

We strongly believe that ELI and ECLI do not necessarily need to be built into the existing legal production processes, but instead can be re-engineered through semantic

¹ [http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52011XG0429\(01\)&from=EN](http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52011XG0429(01)&from=EN)

² [http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52012XG1026\(01\)&from=EN](http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52012XG1026(01)&from=EN)

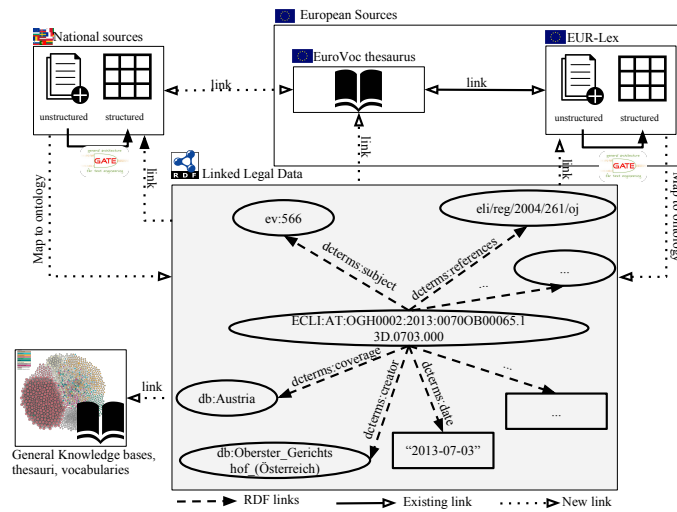


Fig. 1. Framework components to be integrated into a Linked Legal Data Knowledge Graph

extraction and mapping rules, which exploit knowledge about the legal document production process and document structures. Moreover, we argue that the ELI and ECLI standards themselves and their associated metadata-guidelines serve as an excellent basis for a trans-national European legal knowledge graph (KG). Possible use case applications of such a knowledge graph are widespread (c.f. [3], including: (1) supporting the comparative analyses of court decisions and different legal interpretations of legislations; (2) enabling the analyses of the evolution of legislation and jurisdiction; (3) interlinking legal knowledge with other data (such as online discussions, news, etc.).

Previous work is either presenting an idea [3], focusing on representing legal information based on the Akoma Ntoso XML format [2], hence missing linkability required for a legal KG, or solving very specific problems like an ECLI parser for the automatic extraction of legal links and making them available in a machine-readable format [1]. Previous work can be seen as a starting point, but is not sufficient for the creation of a legal KG at the moment.

2 A Linked Legal Data Framework

Fig. 1 illustrates our proposed framework to overcome existing problems in relation to the accessibility of legal data across Europe and includes the primary data source components: the (1) *European EUR-Lex database* containing legal documents issued by the EU using a classification scheme from the *EuroVoc thesaurus*; (2) *national databases* containing information about national laws and court decisions; and (3) *general knowledge bases*, such as DBpedia and Wikidata that also contain information about legal concepts and aspects (which can help us to create links to the outside), but also to enrich for instance links to EuroVoc keywords, which is commonly used to annotate EU legal

documents. The basic concepts of the linked legal KG are legislation identifiers (ELIs), case identifiers (ECLIs) and their properties. The source components in Fig. 1 are connected in three different ways: (i) information is imported into the KG from source components (dotted arrows); (ii) the KG contains backlinks to these sources (dotted arrows); and (iii) existing links already exist between some sources (solid arrows).

The EU institutions (e.g., Council of the European Union, European Commission, etc...) routinely publish and update freely accessible legal documents in the *EUR-Lex*³ database, maintained by the European Publications Office (OP)⁴. This database contains metadata-enhanced legal documents in each of the official languages of the EU member states, such as the authentic Official Journal of the European Union, EU treaties, regulations, directives and EU case law, dating back to 1951⁵.

Each EU member state has its own national legal database, which is used to store legal information, usually in the national language(s). Information is often displayed in HTML and/or available for download as PDF, however a few countries also provide access their national legal databases via an API. For instance, Austria provides an API to access legal documents and associated metadata that comply with ECLI in a JSON serialization. While, Germany⁶ offers documents and metadata in XML, Finland⁷ goes as far as offering legal information as linked data in JSON-LD and via a SPARQL endpoint.

Legal documents present in both the European and national databases often contain concepts for which supplementary information is available in external databases, such as Wikidata⁸ and DBpedia⁹ as well as thesauri like STW¹⁰. This external information could be used to enhance legal documents with additional information and increase the interlinking with other datasets.

3 A Linked Legal Data Knowledge Graph Population

We are using the proposed ECLI and ELI ontologies as a foundation for our legal KG to build upon. The information being included in the KG might be contained in the metadata, the legal document text or can be inferred from the datasource. For space limitations we focus on a high-level description of the mappings, more information about the used methodology and NLP pipeline to be included in the actual poster.

Direct and Configuration Mappings. Certain information contained in the metadata provided by the national databases could be directly linked to the corresponding properties in the ECLI / ELI ontologies without the additional data extraction steps. Configuration files could be used for properties not contained in the metadata, but remaining the same for an entire corpus. Given that legal documents in a country are typically issued in the official language, the language property can be globally set for the corpus of a country.

³ <http://eur-lex.europa.eu>

⁴ <http://publications.europa.eu/>

⁵ <http://eur-lex.europa.eu/content/welcome/about.html>

⁶ <http://www.rechtsprechung-im-internet.de>

⁷ <http://data.finlex.fi/en/main>

⁸ <http://www.wikidata.org>

⁹ <http://wiki.dbpedia.org>

¹⁰ <http://zbw.eu/stw/version/9.0/about.en.html>

Court	EuroVoc	Individual			Combined	
		STW	Wikidata	DBpedia	Wikidata DBpedia	DBpedia Wikidata
OGH	0.30	0.25	0.38	0.43	0.45	0.48
VfGH	0.25	0.16	0.34	0.33	0.39	0.40

Table 1. Results of mapping keywords to EuroVoc descriptors

Indirect Mappings. Missing information requires preprocessing steps such as natural language processing (NLP) techniques or information from external knowledge bases for the mapping to the appropriate ECLI property, e.g. the ECLI properties *dcterms:subject*, *dcterms:description* allow the user to map information about the field of law and descriptive elements. Keywords provided in a national database in natural text must be mapped to the corresponding EuroVoc descriptor to enable multilingual search of legal information. Preliminary results shown in Table 1 for 500 supreme (OGH, 40 distinct keywords) and constitutional court (VfGH, 411 distinct keywords) decisions show the share of keywords that can be mapped directly to EuroVoc or using (combinations) of external knowledgebases and thesauri for translations and the increase of mappings when using external sources. Domain-specific thesauri, document classification systems based on the EuroVoc scheme and NLP techniques could be used to improve the low numbers and increase the share of keywords that can be mapped to an EuroVoc descriptor.

4 Summary and Future Work

We proposed a cross-jurisdictional legal framework demonstrating how legal information stored in national databases can be linked at a European level. The proposed legal KG uses a lightweight ontology based upon the ELI and ECLI specifications and their metadata guidelines as a starting point. For future work we plan to improve the precision and recall by applying different mapping strategies.

Acknowledgments. Funded by the Austrian Federal Ministry of Transport, Innovation and Technology (BMVIT) DALICC project <https://www.dalicc.net>.

References

1. Agnoloni, T., Bacci, L., Peruginelli, G., van Opijnen, M., van den Oever, J., Palmirani, M., Cervone, L., Bujor, O., Lecuona, A.A., García, A.B., Caro, L.D., Siragusa, G.: Linking european case law: BO-ECLI parser, an open framework for the automatic extraction of legal links. In: Legal Knowledge and Information Systems - JURIX 2017: The Thirtieth Annual Conference
2. Boella, G., Caro, L.D., Graziadei, M., Cupi, L., Salaroglio, C.E., Humphreys, L., Konstantinov, H., Marko, K., Robaldo, L., Ruffini, C., Simov, K.I., Violato, A., Stroetmann, V.: Linking legal open data: breaking the accessibility and language barrier in european legislation and case law. In: 15th International Conference on Artificial Intelligence and Law, ICAIL 2015
3. Montiel-Ponsoda, E., Rodríguez-Doncel, V., Gracia, J.: Building the legal knowledge graph for smart compliance services in multilingual europe. In: 1st Workshop on Technologies for Regulatory Compliance (co-located with JURIX 2017) (2017)