Artificial Intelligence and Law (2021) 29:485–539
https://doi.org/10.1007/s10506-021-09282-8

ORIGINAL RESEARCH

The linked legal data landscape: linking legal data across different countries

Erwin Filtz[1,2] · **Sabrina Kirrane**[1] · **Axel Polleres**[1,3]

Accepted: 6 January 2021 / Published online: 25 February 2021

Abstract

The European Union is working towards harmonizing legislation across Europe, in order to improve cross-border interchange of legal information. This goal is supported for instance via standards such as the European Law Identifier (ELI) and the European Case Law Identifier (ECLI), which provide technical specifications for Web identifiers and suggestions for vocabularies to be used to describe metadata pertaining to legal documents in a machine readable format. Notably, these ECLI and ELI metadata standards adhere to the RDF data format which forms the basis of Linked Data, and therefore have the potential to form a basis for a pan-European legal Knowledge Graph. Unfortunately, to date said specifications have only been partially adopted by EU member states. In this paper we describe a methodology to transform the existing legal information system used in Austria to such a legal knowledge graph covering different steps from modeling national specific aspects, to population, and finally the integration of legal data from other countries through linked data. We demonstrate the usefulness of this approach by exemplifying practical use cases from legal information search, which are not possible in an automated fashion so far.

Keywords Linked data · Legal knowledge graph · Legal ontology · Law identifier

✉ Erwin Filtz
erwin.filtz@wu.ac.at

Sabrina Kirrane
Sabrina.Kirrane@wu.ac.at

Axel Polleres
Axel.Polleres@wu.ac.at

1 Institute for Data, Process and Knowledge Management, Vienna University of Economics and Business, Vienna, Austria

2 Siemens AG Österreich, Vienna, Austria

3 Complexity Science Hub, Vienna, Austria

1 Introduction

The law can be seen as a framework that consists of a set of orders defining the rules that govern society. There rules are set by an authority (legislative branch, eg. parliament), enforced by another authority (executive branch, eg. law enforcement authorities) and are defended and interpreted by yet another authority (judicial branch, eg. courts). In order to enable citizens to comply with the law it must be made publicly available. In former times laws were posted on official bulletin boards. Nowadays, legal information systems publicly accessible via the web are used for this purpose. For instance, the Austrian legal information system *Rechtsinformationssystem des Bundes* (RIS)[1] provided by the *Federal Ministry for Digital and Economic Affairs* (BMDW)[2] is a central, publicly available, free of charge, web-accessible platform containing legal documents, such as legislations and court decisions, published by various Austrian authorities (e.g. legislative bodies on both a federal and a state level, courts and tribunals). In addition, jurisdictions have an official manner in which they publish legally binding amendments to existing laws or the abrogation of a law. These publications are usually called bulletins, law gazettes or have other specific names depending on the country.

Yet, despite having legal information publicly available, the documents contained in RIS (or, likewise, other national legal information systems) are not entirely linked with each other. That is, while legal professionals are able to infer links between legal documents and to understand cross-references within those documents by reading the text, the documents and the corresponding metadata are often stored in separate databases, making them hard to access—in particular for non-experts. The lack of integration often results in a tedious time-consuming legal information search process, for instance information may need to be retrieved from the judiciary database for the court decision, and the federal law database for legal provisions. This problem gets even worse when legal documents from other jurisdictions are involved, such as legislative acts from the EU that influence national law, or in the case of cross-boarder cases.

Representing legal information as Linked Data such that legal documents are linked across databases could therefore be highly beneficial, as such linking could speed up the legal information search process significantly and make legal information more accessible, by enabling structured queries and automated aggregation of and navigation through legal information interlinked in a machine-readable manner. Semantic technologies and Linked Data principles have already proven their effectiveness when it comes to data integration, and thus it is not surprising that researchers from the legal domain have already shown interest in the technology (Casanovas et al. 2016). Based on the *Resource Description Framework (RDF)*,[3] a data model that can be used to link data in a standardized, machine-interpretable manner, these technologies allow for the interlinking of data and metadata, making it possible to answer questions that cannot be answered easily at present – due to missing links in legal documents, missing

[1] https://www.ris.bka.gv.at/.

[2] https://www.bmdw.gv.at/en.html.

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

integration of other available legal datasets (e.g. from other authorities not integrated in a legal information system or from other jurisdictions), etc.

The problem of tedious legal information search is obviously not unique to Austria. Other countries, governments and non-governmental initiatives, are also looking into linking legal data and enhancing their national legal information systems using semantic technologies. For instance, Finland provides access to legal information via the *Finlex Data Bank*,[4] which has a web-based search interface and also allows for parts of the legal data to be downloaded in RDF (Oksanen et al. 2019). Other countries, like Greece have set up programs[5] to increase transparency in the legal system and make it more accessible. However, additional steps are required in order to ensure that these separate national initiatives are interoperable. Towards this end, the European Union is working towards enuring better access and exchange of legal information across different countries. While each country is encouraged to set up or continue their own legal information systems—the EU proposes a common set of metadata for legislative and judiciary documents. The *European Legislation Identifier (ELI)*[6] and the *European Case Law Identifier (ECLI)*[7] are non-binding proposals by the EU Council[8] to foster the exchange of legal information by providing legal documents with a minimum set of metadata. In light of increasing globalization and harmonization activities within the European Union it is important that all member states not only adopt the proposed ELI and ECLI ontologies, but also provide national extensions and schemes where required. To this end, our work is guided by the following hypothesis:

> Interlinking national and international legal information from various sources and representing them as Linked Data in a Legal Knowledge Graph will enhance the legal information search process by extending querying possibilities that are not possible at the moment.

The above hypothesis leads to the following research questions:

(i) Can existing ontologies be combined and extended in order to construct a legal knowledge graph?
(ii) Which approaches are needed in order to automatically populate the legal knowledge graph?
(iii) Is it possible to enhance the legal inquiry and search process by linking legal knowledge graphs from other countries?

In order to answer the aforementioned research questions it is necessary to compare the existing ontologies and their properties with the national requirements to determine where extensions are required. Furthermore, the sources of the entities required for the legal knowledge graph population need to be extracted from the document text using state of the art methods. Linking legal data across borders with data from other

[4] https://www.finlex.fi/en/.

[5] https://diavgeia.gov.gr/.

[6] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52012XG1026(01).

[7] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52011XG0429(01).

[8] Body of the European Union composed of national ministers of each EU member state.

countries requires an analyses of the current situation regarding (linked) legal data in these countries. Towards this end, in this paper we make the following contributions:[9]

- We provide an overview of the knowledge graph construction process for our Legal Knowledge Graph (LKG), based on requirements derived from the Austrian legal system, and its current legal information system RIS;
- We propose several legal knowledge graph population methods and exemplify them using our Austrian use case scenario;
- We perform a comparison of rule based and deep-learning based approaches for the automatic extraction of legal entities from legal documents; and
- We provide a comparative analysis of the European legal knowledge graph landscape and identify key challenges and opportunities when it comes to integration across Europe.

The remainder of the paper is structured as follows: Sect. 2 presents the necessary background information on RDF and legal ontologies. The motivating use cases scenario and corresponding requirements used to guide our work are presented in Sect. 3. Our proposed legal knowledge graph construction and population process for Austrian legal data is presented in Sect. 4. Section 5 contains an overview of the current European legal knowledge graph landscape along with key challenges and opportunities when it comes to the integration of these different efforts. A critical discussion of different use case examples is provided in Sect. 6, followed by the discussion of related work in Sect. 7. Finally, Sect. 8 concludes the paper and discusses directions for future work.

2 Background

Knowledge Graphs (Hogan et al. 2020) are a trending topic, which is attracting increased interest in various domains: in order to organize and link information in a flexible manner, such knowledge graphs typically contain both factual and schematic (or, resp., ontological) information, in a flexible and extensible graph structure. Open standards and technologies to create, represent, interchange and process Knowledge Graphs origin from the Semantic Web and Data activities within the World Wide Web Consortium (W3C).[10] In this section we provide background information on respective standards and principles, such as the Resource Description Framework (RDF) and Linked Data, and discuss existing legal ontologies that serve as a basis to create our legal knowledge graph.

2.1 Semantic web and linked data

Legal information is typically represented as natural text with the information contained inside documents is not readily available in a machine readable format. When

[9] Additional material is available under: https://github.com/efiltz/legal-knowledge-graph.

[10] https://www.w3.org/2001/sw/.

```
1  PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2  PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
3  PREFIX eli: <http://data.europa.eu/eli/ontology#>
4  PREFIX xsd: <http://www.w3.org/2001/XMLSchema#date>
5  PREFIX frbroo: <http://iflastandards.info/ns/fr/frbr/frbroo/>
6  <http://data.europa.eu/eli/dir/2014/92/oj>
7      rdf:type
8          eli:LegalResource ;
9      eli:type_document
10         <http://publications.europa.eu/resource/authority/resource-type/DIR> ;
11     eli:date_publication
12         "2014-08-28"^^xsd:date .
13 <http://data.europa.eu/eli/ontology#LegalResource>
14     rdfs:subclassOf frbroo:F1_Work .
```

Listing 1 RDF snippet for EU Directive 2014/92/EU (serialized in Turtle)

it comes to machine-readability the *Resource Description Framework (RDF)*[11] can be used to make metadata statements about a particular resource (e.g. in our case a legal provision or a court decision) which is identified by a *Unique Resource Identifier (URI)*. Listing 1 shows an RDF snippet about the EU directive 2014/92/EU. In the first five lines URI prefixes used to appreviate *namespaces* are defined, such that for instance `eli:LegalResource` turns into http://data.europa.eu/eli/ontology#LegalResource (line 8). An overview of the used namespaces in this paper is presented in Listing 2. Web URIs are represented using the *Hypertext Transfer Protocol (HTTP)*.[12] Besides URIs also typed and untyped *Literals* are used in RDF to describe properties of a certain resource. While *untyped* literals are always interpreted as text strings, *typed* literals may have a datatype that tells us how to interpret the information, for instance whether a string is to be interpreted as a textual string (`xsd:string`) or as a date (`xsd:date`), as shown in the example for property `eli:date_publication` in line 12. An RDF statement consists of the three components *subject, predicate, object* and is called a *triple*, which may also be viewed as a directed typed link or edge between subjects and objects. The so connected RDF triples form a graph structure. A collection of triples describing schema and instance data is called *ontology*. Although RDF can be serialized in various formats (e.g. RDF/XML,[13] N-Triples[14]) in this paper we use the Terse RDF Triple Language (Turtle)[15] due to its simplicity and readability. Additional formats include *RDF in Attributes (RDFa)*,[16] which is used to embed RDF in HTML and XML documents, or JSON-LD.[17]

RDF Schema (RDFS)[18] and the *Web Ontology Language (OWL)*[19] are used to describe classes of and properties (relations) between resources. The core features

[11] https://www.w3.org/TR/rdf11-concepts/.

[12] https://tools.ietf.org/html/rfc2616.

[13] https://www.w3.org/TR/rdf-syntax-grammar/.

[14] https://www.w3.org/TR/n-triples/.

[15] https://www.w3.org/TR/turtle/.

[16] https://www.w3.org/TR/rdfa-primer/.

[17] https://www.w3.org/TR/json-ld11/.

[18] https://www.w3.org/TR/rdf-schema/.

[19] https://www.w3.org/TR/owl2-overview/.

```
prefix lkg: <https://data.wu.ac.at/legal/lkg#>
prefix av: <https://data.wu.ac.at/legal/austrovoc#>
prefix owl: <http://www.w3.org/2002/07/owl#>
prefix rdf: <http://www.w3.org/1999/02/22−rdf−syntax−ns#>
prefix rdfs: <http/www.w3.org/2000/01/rdf−schema#>
prefix dcterms: <http://purl.org/dc/terms/>
prefix skos: <http://www.w3.org/2004/02/skos/core#>
prefix xsd: <http://www.w3.org/2001/XMLSchema#date>
prefix cdm: <http://publications.europa.eu/ontology/cdm#>
prefix frbroo: <http://iflastandards.info/ns/fr/brbr/frbroo/>
prefix eli: <http://data.europa.eu/eli/ontology#>
prefix ev: <http://eurovoc.europa.eu/>
prefix gn: <http://sws.geonames.org/>
```

Listing 2 Namespaces used in examples throughout the paper (serialized in Turtle)

of RDFS are summarized in the ρdf subset (Muñoz et al. 2009), which contains properties to define simple taxonomies in terms of class (`rdfs:subClassOf`) and property (`rdfs:subPropertyOf`) hierarchies. In such a hierarchy, implicit superproperties between resources, as well as membership in the superclass from membership in the subclass can be inferred. Likewise, domain (`rdfs:domain`) and range (`rdfs:range`) restrictions can be used to infer the class membership of subjects or objects of particular properties as shown in Listing 1 line numbers 13 and 14 that the ELI class `eli:LegalResource` is a subclass of `frbroo:F1_Work`. OWL caters for the definition of more complex ontological axioms on classes and properties, which can be used for more complex reasoning.

The *SPARQL Protocol and RDF Query Language (SPARQL)*[20] is used to retrieve RDF data. SPARQL queries search for matches of user defined triples (*graph patterns*). A `SELECT` query allows users to define a graph pattern which must match the data and the variables to be returned. Basic graph patterns must match all results in order to be returned, whereas in an `OPTIONAL` query we can also define *optional patterns* that need not occur in all results and return an empty binding if not matched. With *alternative patterns* using `UNION` it is possible to define multiple graph patterns of which at least one must be fulfilled. The number of results can be reduced using a `FILTER` clause, which allows users to restrict results to literals that contain a particular string, or to apply comparison operators such as equals, greater than and so on, for instance as shown in Example 1. Long query result lists can be manipulated using *solution modifiers* such as `ORDER BY`, which sorts the results in an ascending or descending order based on the given variable, as well as `LIMIT` and `OFFSET` to restrict the number of results.

[20] https://www.w3.org/TR/sparql11-overview/.

Example 1 SPARQL Query: Which EU directives have been published in 2014?

```
PREFIX eli: <http://data.europa.eu/eli/ontology#>
PREFIX eu: <http://publications.europa.eu/resource/authority/resource-type/>
SELECT (?s as ?Directive)
WHERE {
  ?s eli:type_document eu:DIR .
  ?s eli:date_publication ?d .
  FILTER (year(?d) = 2014)
}
```

Directive
<http://data.europa.eu/eli/dir/2014/23/oj>
<http://data.europa.eu/eli/dir/2014/92/oj>
...

In order to make machine-readable data more accessible on the Web, Tim Berners-Lee (Berners-Lee 2006) proposed a set of *Linked Data Principles* for publishing data on the Web, which fundamentally rely on RDF:

1. Use URIs as names for things.
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI provide useful information using the standards RDF and SPARQL.
4. Include links to other URIs, so that they can discover more things.

The *things* mentioned in the first principle refer to resources. Identifying resources with HTTP URIs allows the consumers to retrieve additional information about these resources on the Web. Information about the resources stored in RDF allows them to be retrieved using SPARQL. The fourth rule stipulates that resources should be linked with other resources and shall allow users or agents to browse through different resources by following links.

2.2 Legal ontologies

We base our modeling on the ELI and ECLI ontologies which are specific to the legal domain, as well as the European EuroVoc thesaurus which is also available as an RDF vocabulary. Both the ELI and ECLI ontologies have been proposed in the form of conclusions of the Council of the European Union which consists of EU member states' ministers of the respective policy area. Conclusions are documents that express a political expression without the intent of legal effects. *EuroVoc* is a standardized thesaurus containing normative terminology used in the context of European administration and publications, not restricted to legislation alone. In addition to ELI and ECLI we also introduce the Common Data Model (CDM) which is used by the EU to model their legal data.

European Law Identifier. The European Law Identifier (ELI) (Council of the European Union 2012) serves as a common system to identify legislative documents and its metadata first proposed in 2011 and is followed by additional Council conclusions in 2017 (Council of the European Union 2017) acknowledging the efforts of the participating

Table 1 Mandatory properties of the ELI ontology

Property	Description
`eli:realizes`	Describes that a legal expression materializes a legal resource
`eli:embodies`	Describes that a format represents a legal expression
`eli:type_document`	Indicates the type of a legal resource
`eli:language`	The language in which a legal expression is written
`eli:title`	The title of a legal expression
`eli:format`	Resource format expressed as URI (e.g. HTML)

countries, introducing an ELI task force and clarifying the three pillars of the ELI system. The three pillars (Francart et al. 2018) the ELI is built on are: [(i)] to foster the assignment of unique identifiers for laws; [(ii)] to use a common ontology that provides a metadata standard; and [(iii)] to As for classes and properties in the ELI ontology, for instance, the EU is required to publish legal acts in various languages and therefore needs the ability to represent different language versions of the same legal act. The ELI ontology distinguishes between three classes of resources and six mandatory properties. As shown in Table 1, a `eli:LegalResource` is a distinct intellectual creation such as a legal act which is realized by a `eli:LegalExpression` and embodied in a specific `eli:Format`. Hence, a `eli:LegalExpression` has a `eli:title` and `eli:realizes` the base version in a particular language (`eli:language`) of a `eli:LegalResource` which is of a specific `eli:type_document`, for instance a directive. The `eli:LegalExpression` is published in a `eli:Format` which is the actual physical representation, whereas physical includes paper as well as electronic formats such as HTML or PDF.

The ELI (both in terms of identifier syntax and in terms of the usage of metadata properties) is modeled in different ways from country to country depending on the respective legal system. Notably, the Council conclusions defines all of the syntactic components of the ELI being optional, such that national requirements can be fulfilled and not all components need to be implemented in each national legal system. Additional information for the member states as well as reference files for the ELI ontology are provided in HTML,[21] XLSX[22] and OWL[23] format. The ELI follows the principles set forth in the Functional Requirements for Bibliographic Records[24] (FRBR) ontology (Publications Office of the European Union 2020b) but uses the object-oriented version of FRBR[25] for the ELI ontology (prefix `frbroo:`), for instance `eli:LegalResource` is a `rdfs:subClassOf frbroo:F1_Work` and `eli:LegalExpression` is a `rdfs:subClassOffrbroo:F22_Self-Contained_Expression`. The

[21] http://publications.europa.eu/resource/distribution/eli_documentation/html/doc_user_manual/eli_ontology.html.

[22] http://publications.europa.eu/resource/distribution/eli/xlsx/owl/eli_ontology.xlsx.

[23] http://publications.europa.eu/resource/distribution/eli/owl/owl/eli.owl.

[24] https://www.ifla.org/publications/functional-requirements-for-bibliographic-records.

[25] https://www.ifla.org/files/assets/cataloguing/FRBRoo/frbroo_v_2.4.pdf.

Table 2 Mandatory properties of the ECLI ontology

Property	Description
`dcterms:identifier`	The URL where the resource can be retrieved
`dcterms:isVersionOf`	Indicates that a resource is a version of another resource
`dcterms:creator`	Full name of deciding court
`dcterms:coverage`	Indicates the country in which the court or tribunal has its seat
`dcterms:date`	The date when a decision has been rendered
`dcterms:language`	The language in which this particular is written
`dcterms:publisher`	The organization that is responsible for the publication of the document
`dcterms:accessRights`	Defines who can access the resource, *public* or *private*
`dcterms:type`	Defines the type of the rendered decision

ELI syntax is very flexible and can be adjusted to national requirements by adding and removing individual components. The syntax of the ELI identifier is defined as the base URI followed by *eli* with the rest of the components being optional and separated by slashes, for instance the ELI for a EU directive such as `http://data.europa.eu/eli/dir/2014/92/oj` looks different from an Austrian legal provision `https://www.ris.bka.gv.at/eli/bgbl/1979/140/P28a/NOR40180997`.

European Case Law Identifier. The European Case Law Identifier (ECLI) (Council of the European Union 2011) has been created to introduce an identifier for *case law*, and to define a minimum set of metadata for judiciary documents (e.g. court decisions). The ECLI does not define any specific classes and uses the properties of the Dublin Core Metadata Initiative (DCMI)[26] ontology with the prefix `dcterms`. In contrast to the ELI there is no separate formal ontology specification provided by the EU, but rather only a recommendation of nine mandatory (listed in Table 2) and eight optional properties which should be used to describe metadata relating to the documents. Moreover, the ECLI conclusion makes particular suggestions for the use of the `dcterms` vocabulary, for instance that the object of `dcterms:coverage` should be used for the country (or more closely defined location) where the court is seated. Unfortunately, these suggestions are given without explicit ontological commitments or formal axioms, e.g. in terms of explicit range restrictions.

The syntax of the ECLI identifier is more restricted compared to the ELI as it consists of five components separated by a double colon, for instance `ECLI:AT:OGH0002:2016:0100OB00012.16M.1220.000` for a decision of the Austrian Supreme Court. The order of the components is fixed and starts with the abbreviation *ECLI* and is followed by a country code (or code of an international organization). The third component is the court code of the deciding court which is individually assigned by each participating country and the year of the decision. The last component is an unique ordinal number of the decision.

[26] https://dublincore.org/.

```
<http://eurovoc.europa.eu/2836>
  a skos:Concept;
  skos:broader
    ev:138;
  skos:prefLabel
    "Verbraucherschutz"@de, "consumer protection"@en .
<http://eurovoc.europa.eu/138>
    skos:prefLabel
        "Verbraucher"@de, "consumer"@en .
```

Listing 3 Example EuroVoc (serialized in Turtle)

EuroVoc. The *EuroVoc* thesaurus[27] is a multi-domain and multi-lingual thesaurus provided by the Publications Office of the European Union (OP) used to classify EU documents into categories for easier information search. It is based on the Simple Knowledge Organization System (SKOS),[28] a well-known standard[29] to represent information using RDF. The individual terms in the EuroVoc thesaurus are of type `skos:Concept` and a collection of concepts is aggregated in a `skos:ConceptScheme`. Concepts are linked using the properties `skos:narrower` and `skos:broader` to represent the hierarchical structure of terms and `skos:related` for associative relations. EuroVoc is organized in 21 domains, for instance *Law, Economics, Trade* and 127 microthesauri. In total, EuroVoc contains more than 6,000 concepts and each concept has one preferred term (`skos:prefLabel`) and (optional multiple) non-preferred terms (`skos:altLabel`), i.e. synonyms. All concepts are available in the languages of the 23 EU member states and in addition three languages of EU membership candidate countries. The concepts are arranged in a way to avoid polihierarchies except for the Geography domain. Listing 3 shows a snippet of concept `ev:2836` with its preferred labels in German (*Verbraucherschutz@de*) and English (*consumer protection@en*) having a `skos:broader` concept *ev:138* which is labeled *Verbraucher@de* and *consumer@en*.

Common Data Model. The Publications Office of the European Union (OP) uses the Common Data Model (CDM)[30] for their published resources which is based on FRBR (Francesconi et al. 2015; Publications Office of the European Union 2020a). The resources that can be accessed via the Eur-Lex SPARQL endpoint are represented using the CDM ontology rather than the ELI and ECLI ontology. An RDF dump of the Eur-Lex data using ELI, up until 2018, is available on the EU Open Data Portal.[31] The usage of the CDM ontology results in using a different identifier for the documents in the Eur-Lex database CELLAR, the repository of the EU Publications Office, instead of the ELI identifier. A mapping between CELLAR and ELI identifiers is however provided using the predicate `owl:sameAs`.

[27] https://op.europa.eu/s/n3kP.

[28] https://www.w3.org/2004/02/skos/.

[29] https://www.w3.org/2009/08/skos-reference/skos.html.

[30] https://op.europa.eu/en/web/eu-vocabularies/model/-/resource/dataset/cdm.

[31] http://data.europa.eu/88u/dataset/eli-european-legislation-identifier-eurlex.

3 Use case and requirements: a case for legal linked data in Austria

The work presented herein is based on a project commissioned by the Austrian Ministry for Digital and Economic Affairs (BMDW).[32] The goal of this project was to investigate how the current Austrian legal information system RIS could be improved in terms of searchability and accessibility by: (i) transforming the metadata from RIS into a legal knowledge graph; (ii) further enriched with information extracted from document texts stored within RIS; and (iii) automatically interlinking these legal documents. In the following we provide an overview of the Austrian legal information system and the challenges, requirements and scenarios addressed in the course of the project.

Austrian legal information system. The *Rechtsinformationssystem des Bundes (RIS)*[33] is the legal information system of the Republic of Austria. RIS serves as a single point of information from which legal documents issued by various authorities can be searched and accessed. In addition to the web interface, RIS also provides access to its data via a REST API[34] enabling users to access RIS data in JSON.[35] Through the web interface different backend databases – subdivided into different parts of the legislation – such as *Bundesrecht* (federal law), *Landesrecht* (state law of the nine Austrian states) or *Judikatur* (judiciary) and many more – can be accessed. Documents in RIS can be retrieved in different formats like HTML, XML, RTF and PDF. Although the RIS web interface gives the impression that it is a single database containing all legal information, it is in fact a collection of independent databases which are not currently connected nor interlinked underneath.

Use case. Currently the search process is mainly based on basic keyword search with the possibility to add filters to restrict the search space for instance to timeframes by setting dates. The objective of the project was threefold: (i) develop a legal information system that is capable of also representing related information, i.e. links to other legal documents referenced within a document, to classify documents based on a classification schema; (ii) to allow for enhanced search capabilities by making certain information contained in documents explicit, for instance linking entities mentioned in the documents to external knowledge bases such as Geonames or DBpedia; and (iii) to support cross-jurisdictional search requests by integrating legal data from other countries and the European Union. The end goal being to allow us to seamlessly get answers to complex search queries such as the following:

- Which documents are referenced in a specific court decision?
- Over which districts does a court have competent jurisdiction?
- What are the national transpositions of a specific EU directive?
- Which legal documents regulate a specific legal area searched with keywords in a foreign language?

[32] https://www.bmdw.gv.at/.

[33] https://www.ris.bka.gv.at/.

[34] https://data.bka.gv.at/ris/api/v2.5/.

[35] https://tools.ietf.org/html/rfc8259.

Challenges. Primary challenges in the context of the project and the use case in order to facilitate the answering of such complex questions in a more automated manner include the following:

Unstructured/missing information Information about legal documents can be contained in both structured metadata but also within unstructured text, for instance law references in court decisions are not contained in metadata. Further, some connections between documents are only implicitly available in the text and while these can be detected by a human reader, a machine would struggle with the same task. In addition, the mandatory and optional properties within the ELI and ECLI ontologies can only be partially constructed from the document metadata alone.

Data silos The Council identified the need to disseminate legal information and that the identification and exchange of legal information from national authorities supports access to legal information.[36] At the moment these legal information systems are still separate silos. Our objective is that Linking legal data first nationally across so far disconnected backend databases *and*, as a second step, across Europe will help to reduce the problem of data silos. It is worth noting that automatic extraction from and linkage of existing databases should avoid any need to maintain the same information at multiple places, while also allowing the data to be easily integrated with other sources.

Redundant data storage Considering that legal documents contain references to each other, the legal information search process typically involves the need to search across the different databases. At the moment, additional information that should be made available for full text search but is not part of the particular database is stored in an additional column. Still, this leads to redundant data storage and does not add any beneficial additional information except enabling search. Furthermore, this situation results in anomalies which must be considered on insert, update and deletion operations. Linked data helps to avoid these anomalies as it does not require to store the same information redundantly at multiple places and therefore provides more flexibility.

Requirements From the challenges outlined above we derive three core requirements. It must be possible to **extract** information that is missing in the metadata from the document text. We need to **integrate** legal data from various national and international data sources into a single knowledge base. **Normalization** by assigning unique identifiers instead of plain text references should be used to avoid redundancies and inconsistencies.

Legal Knowledge Graph Creation Methodology

The aforementioned legal ontologies and use case requirements serve as an input for the legal knowledge graph creation process, which is depicted in Fig. 1. In the first step we model the ontology to represent the Austrian legal system based on ELI and ECLI and create a national thesaurus *AustroVoc* for the representation of Austrian specific terms, not covered in existing terminologies such as EuroVoc. Since ELI and ECLI are only describing a minimum set of metadata in order to be applicable to all EU member states, we needed to create additional classes and properties for our legal knowledge

[36] 2011/C127/01, 2012/C325/02: Identification of needs.

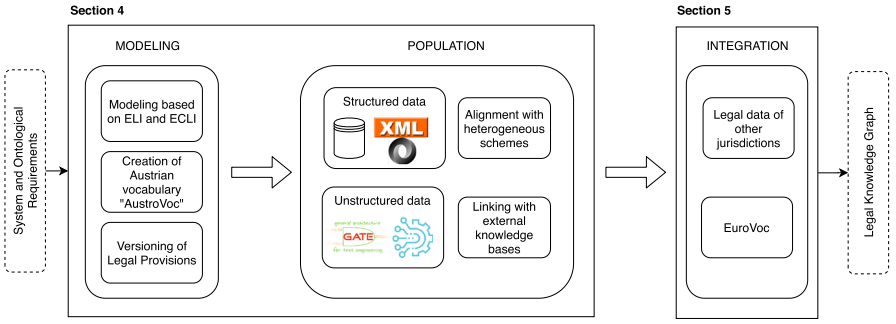

Fig. 1 Legal Knowledge Graph Creation Methodology

graph to reflect Austrian specific requirements. Content Ontology Design Patterns can help to create (legal) domain-specific ontologies, for instance already shown for the modeling of licensing (Rodríguez-Doncel et al. 2013) and consumer complaints (Santos et al. 2016), and provide building blocks to ensure reusability (Presutti and Gangemi 2008): in our particular case we can build on the already existing ELI and ECLI ontologies. However, on the one hand the existing ontologies are in parts not fine-grained enough and on the other hand legal documents and their metadata provide us with additional required information on the missing parts. Therefore, we extended these ontologies in a middle-out fashion, which seems appropriate in combining top-down and bottom-up approaches and helps us to keep an adequate level of detail (Uschold and Gruninger 1996). In the bottom-up phase we analyzed the available metadata and which additional, relevant data could be extracted from the Austrian legal documents (using Natural Language Processing techniques) in order to populate classes and properties that need to be added, keeping in mind our primary goal is interlinking of the documents, rather than describing the actual content of the documents. In the top-down phase we reused the existing ontologies and refined and extended classes, properties, as well as taxonomic terminologies/thesauri, where needed. This approach has also been described to be effective in the legal domain in a similar setting with existing legal ontologies that are extended based on underlying legal documents (Ghosh et al. 2016). Based on the resulting combined ontological schema, the resulting model has been populated with data from RIS and linked to external knowledge bases. Both steps are described in Sect. 4. In a final step, described in Sect. 5, we integrate external legal data from the European Union, the European thesaurus *EuroVoc* containing terms from different domains in the official languages of the EU member states and also legal data from selected other countries.

4 The Austrian legal knowledge graph

In this section we describe how we map explicit metadata information in **the Austrian legal information system** RIS as well as implicit information contained within the RIS documents to the ELI and ECLI ontological models introduced in Sect. 2.2. This mapping is used to form the foundations of our legal knowledge graph. Furthermore,

we introduce a national vocabulary *AustroVoc* which is mapped to EuroVoc where possible. Finally, the model is populated with data from RIS and linked with external knowledge bases.

4.1 Legal knowledge graph modeling

Given that our project was commissioned by the Austrian Ministry of Digital and Economic Affairs, who are interested in participating in the European linked legal data initiatives, we model our Austrian legal knowledge graph based on the ELI and ECLI ontologies. This decision was motivated by the fact that: (i) By doing so the ministry contribute towards the goals of ELI and ECLI as laid out in the Council conclusions for the introduction of ELI and ECLI; (ii) The EU is a supranational system that aims to provide easier access to and interlinking of legal information across Europe, which can only be successful if the various member states participate and use the same system; and (iii) It is possible to accommodate specific national requirements by extending the ELI and ECLI ontologies with classes and properties specific to the Austrian legal system, such that information contained in RIS for which ELI and ECLI do not provide properties can be represented. Such an approach is also common practice in other countries, for instance the Finnish Semantic Finlex Legislation Ontology or the Greek Nomothesia ontology.

When it comes to alternative modeling approaches, Francesconi et al. (2015) highlight the disadvantages of coupling resources with the corresponding FRBR classes stating that such a coupling leads to complex queries that are needed in order to retrieve metadata for all FRBR levels (resource, expression, etc...). Although the proposed alternative modeling reduces complexity it does so at the cost of interoperability, which is one of the core requirements underpinning our work. Considering, that linking is necessary to support the legal inquiry process across different jurisdictions, the proposed optimization needs to be built into the ELI and ECLI standards. The incorporation of the proposed optimization and others coming from the research community will be discussed later in Sect. 5.

4.1.1 Modeling the Austrian legal system based on ELI and ECLI

Since both ELI and ECLI are targeting a variety of different legal systems within the EU member states, they only provide two classes of legal documents, which we extended in order to represent specific legal document types used in Austria's national legal publication process, such as law gazettes and legal provisions. In our examples herein we exemplify our legal knowledge graph with a focus on federal law as well as jurisdiction by the justice branch, which includes decisions of the Supreme Court and lower courts. Figure 2 depicts our legal knowledge graph model with the specific classes we added colored gray. Nodes denote classes and edges properties connecting their respective domain and range classes.

Law Gazette The law gazette is used to publish new laws or any changes to existing laws, which happen in editorial instructions (e.g. *[...] in § X change amount Y to Z [...]*). We represent the law gazette with class `lkg:LawGazette`

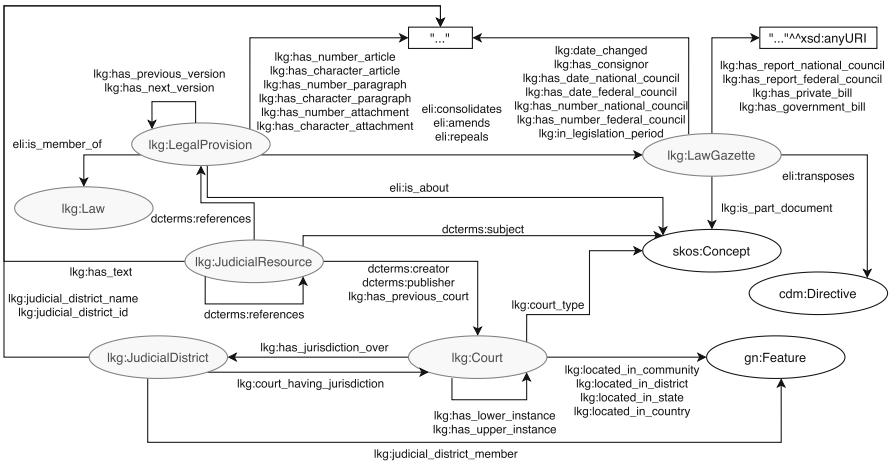

Fig. 2 Legal knowledge graph model

(subclass of `eli:LegalResource`).p We introduce new properties to provide background information about the legislative process which is a useful source to solve legal interpretation problems. These properties cover dates when law changes have been discussed in the councils (`lkg:has_date_national_council`, `lkg:has_report_national_council`) and links to the reports about the parliamentary discussion[37] which are available on the web (`lkg:has_report_national_council`, `lkg:has_report_federal_council`). These reports are useful in case there is a loophole in the law and the will of the parliament needs to be discovered. Bills initiate the legislative process and are linked using the properties `lkg:has_private_bill` and `lkg:has_government_bill`. The authority bringing in a bill is indicated with the property `lkg:has_consignor`. We use `lkg:is_part_document` to determine the type of the law gazette such as *constitutional law* or *order*. The legislation period in which a law gazette has been published is included for legal analysis and is indicated using the property `lkg:in_legislation_period`.

Legal Provision and Law. A `lkg:LegalProvision` (subclass of `eli:LegalResource`) is a resource containing the actual norm. In Austria each legal provision is an individual document with a *NOR* number as an unique technical identifier, for instance *NOR40180997* (see Listing 4) and a label used in legal practice, for instance *§ 28a KSchG* (Paragraph 28a of the Consumer Protection Law). Figure 3 shows the legal provisions *Artikel 2 B-VG* (Art. 2 of the Constitution) and *§ 28a KSchG*. A legal provision can be labeled *Artikel* (article) or *Paragraph* (paragraph) and is always seen in its entirety for modeling, irrespective of whether there is only one *Absatz*[38] (subsection) or multiple subsections.

[37] Publicly available at the Austrian parliament's website: https://www.parlament.gv.at/.

[38] The English translation of *Absatz* is *paragraph*, but we call the *Absatz* subsection to avoid confusion, as the word *Paragraph* in Austrian/German legal language rather refers to law articles.

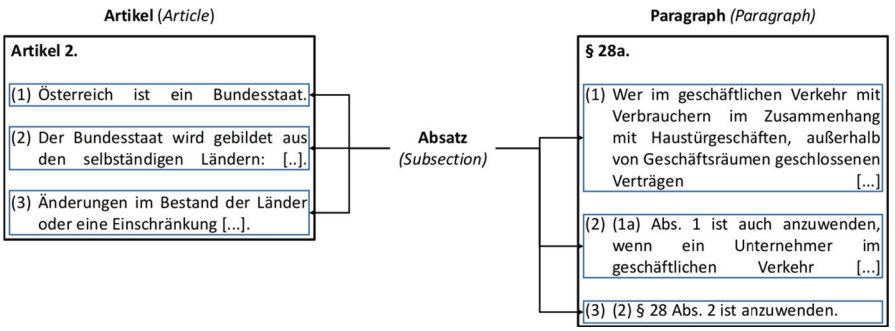

Fig. 3 Legal provision naming convention

```
<https://www.ris.bka.gv.at/eli/BGBl/1979/140/P28a/NOR40180997>
        lkg:has_number_paragraph
                28 ;
        lkg:has_character_paragraph
                "a" ;
        lkg:has_next_version
                ris:eli/BGBl/1979/140/P28a/NOR40192489 ;
        lkg:has_previous_version
                ris:eli/BGBl/1979/140/P28a/NOR40173437 .
```

Listing 4 Legal Provision §28a Consumer Protection Law (shortened, serialized in Turtle)

Listing 4 depicts an RDF snippet for legal provision *§ 28a KSchG* with the new properties we introduced in our extended `lkg:` ontology highlighted in red. Besides the *Artikel* and *Paragraph* there is also a *Anlage* (attachment) usually used for transitional provisions which combines both *Artikel* and *Paragraph*, for instance *Artikel 1 § 1*. We introduce new properties to model numbers as well as characters in the labels of legal provisions, for instance `lkg:has_number_paragraph` and `lkg:has_character_paragraph`. Analogously, for legal provisions named by article or attachment we use the properties `lkg:has_number_article`, `lkg:has_character_article` and `lkg:has_number_attachment`, `lkg:has_character_attachment` respectively. Two temporally subsequent legal provisions are linked with `lkg:has_next_version` and `lkg:has_previous_version`. We create the class `lkg:Law` because legal provisions can be a part of a *law book* which is a collection of legal provisions containing regulations about the same topic. The membership between a `lkg:LegalProvision` and `lkg:Law` is indicated with the ELI property `eli:is_member_of`.

Legal provisions are the basis for court decisions and it is therefore important to link a judgment with the correct version of a legal provision. The linking between judgments and legal provisions is achieved by following a *date-based linking approach* which links a judgment to the legal provision that is in force at the decision date because this will be the correct version most of the time. Furthermore, a specific version of a legal provision is always the sum of the initial version with all its amendments over time.

Judicial Resource. The class `lkg:JudicialResource` (subclass of `frbroo:F1_Work`) is used for judiciary documents which are modeled based

on the ECLI suggestions. We add the text of a court decision with the property `lkg:has_text`. The EU Publications Office (OP) provides *Named Authority Lists (NAL)* which are vocabularies to standardize the inter-institutional legal data exchange. Some of these NAL can be used by all countries, for instance the NALs for languages or countries, while other NAL are very EU-specific, for instance court-types which contain EU courts only and therefore cannot be used for national courts. We use these NALs for the ECLI properties that indicate in which country the deciding court is seated (`dcterms:coverage`), the language of the decision (`dcterms:language`) and the access rights (`dcterms:accessRights`). Properties populated with Austrian specific values, such as `dcterms:type`, `dcterms:publisher`, `lkg:previousCourt`, are linked with concepts contained in the *AustroVoc* thesaurus we created for this purpose.

Court and Judicial District A judgment in the judiciary branch is rendered by a `lkg:Court` of a specific type indicated with `lkg:court_type`. Furthermore courts are organized in a hierachical manner and have a higher instance indicated with the property `lkg:has_upper_instance` and a lower instance (`lkg:has_lower_instance`). A court is located in a community (`lkg:located_in_community`), district (`lkg:located_in_district`), state (`lkg:located_in_state`) and country (`lkg:located_in_country`). A district court also `lkg:has_jurisdiction_over` a `lkg:JudicialDistrict`.[39] Similarly, the property `lkg:court_having_jurisdiction` indicates the court having spatial competent jurisdiction. The competent jurisdiction is assigned to the lowest level of authorities, hence district courts. Since we know that a district court has competent jurisdiction over a particular area and that court has an upper instance we can also infer that a higher court has competent jurisdiction over all areas of all lower courts assigned to the higher court. To represent spatial information we us the publicly available database *Geonames*,[40] which provides identifiers and spatial information for locations in multiple languages as well as a small ontology (prefix `gn:`) describing these properties. Figure 4 illustrates the difference between political and judicial districts for the capital of Austria, Vienna which is divided into 23 political districts but only 12 judicial districts. The two political districts *Leopoldstadt* (`gn:2772614`) and *Brigittenau* (`gn:2781400`) are the members (`lkg:judicial_district_member`) of the single judicial district named (`lkg:judicial_district_name`) *Leopoldstadt*.

4.1.2 The Austrian vocabulary: AustroVoc

We propose a SKOS-based thesaurus *AustroVoc* containing Austrian specific terminology. ELI and ECLI encourage member states to create their own schema for the properties indicating a document type (`eli:type_document` and `dcterms:type`) and a document classification to describe the content or legal area of a document (`eli:is_about` and `dcterms:subject`). We create three different schemes for

[39] https://www.statistik.at/web_de/klassifikationen/regionale_gliederungen/gerichtsbezirke/index.html.

[40] https://www.geonames.org/.

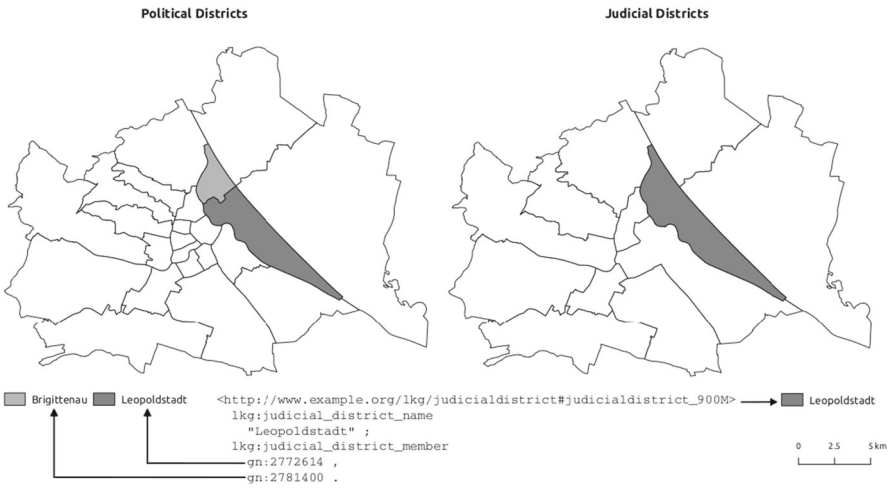

Fig. 4 Illustration of political and judicial districts for the Austrian capital Vienna

```
<https://www.ris.bka.gv.at/eli/BGBl/1979/140/P28a/NOR40180997>
    eli:is_about :bri2006 .
:bri2006 a skos:Concept;
    skos:broader :bri20;
    skos:prefLabel "Konsumentenschutz"@de ;
    rdfs:seeAlso ev:2836 .
```

Listing 5 Law index example (shortened, serialized in Turtle)

Gericht-typ (court type), *Bundesrechtindex* (law index) and *Resource-typ* (resource-type).

Gericht typ. The court types provided in the *Named Authority Lists (NAL)*[41] of the EU Publications Office cannot be used 'as is' since they only contain EU courts. That is why we create an additional *court-type* scheme which contains the different types of Austrian courts. We distinguish between public tribunals, for instance the Constitutional Court (`av:vfgh`), and ordinary courts, for instance the Supreme Court (`av:ogh`), which are responsible for different legal areas and are organized in a hierarchical way. Adding this information enables a search for judgments rendered by courts of a particular type and superior or subordinate courts and legal analysis.
Bundesrechtindex The *law index* is an index for Austrian federal law[42] provided by RIS organizing the law in a hierarchical manner. As shown in Listing 5 every legal provision is assigned to an entry in this index with the property `eli:is_about` which allows users to search for legal provisions belonging to a specific legal area, for instance §28a KSchG linked to the law index `av:bri2006`. We also use the law index to indicate the legal area of judgments dependent on the legal provisions

41 https://op.europa.eu/en/web/eu-vocabularies/at-dataset/-/resource/dataset/court-type.

42 https://www.ris.bka.gv.at/UI/Bund/Bundesnormen/IndexBundesrecht.aspx?TabbedMenuSelection=BundesrechtTab.

they are based on using `dcterms:subject`. Finally, where possible (for details, see Sect. 4.2.3 below) we link the national law index items with corresponding items to the European thesaurus *EuroVoc* using the property `rdfs:seeAlso` to enable a multi-lingual search across jurisdictions. For instance, the AustroVoc law index `av:bri2006` (*Konsumentenschutz@de*) is linked to the EuroVoc concept `ev:2836` (*Verbraucherschutz@de*).

Resource typ. As with the court-types mentioned above, the resource-types contained in the NAL[43] are EU specific and incomplete with regards to missing specific resources used and required in Austria. We again created our own schema for such specific *resource-types* in RIS. These mainly include different document types, for instance judiciary documents can be subdivided into *Entscheidungstext* (decision text) or a *Rechtssatz* (legal rule) which is a case summary from which general legal rules can be inferred. The properties used to indicate the document types are already available in ELI (`eli:type_document`) and ECLI (`dcterms:type`). These properties to indicate the document types are not to be confused with the property `rdf:type` that is used to indicate to which class a document belongs to, for instance judiciary documents are of type `lkg:JudicialResource` and legislative documents are of type `lkg:LawGazette` or `lkg:LegalProvision`.

4.2 Legal knowledge graph population

We describe different approaches to populate our legal knowledge graph with structured data from the RIS database. While some entities and their relationships can directly be extracted from structured metadata within RIS, for the population from unstructured (text) data we make use of Natural Language Processing (NLP) tools and techniques and provide a comparison of different (rule-based as well as machine-learning based) legal entity extraction approaches exemplified with a dataset of manually annotated court decisions.

4.2.1 Population from structured data

For the population from structured data we were provided with a dump of the relational RIS database which contains the metadata as well as the text of the legal documents contained in RIS. The database schema used does not satisfy the ELI or ECLI metadata requirements upfront. In addition, each RIS application is currently stored in a separate relational database.

Direct population A direct mapping [in analogy with the terminology used in R2ML (W3C Recommendation 2012)][44] of the legal knowledge by mapping attributes to URLs is possible where the required metadata is available. This is typically applicable to properties that have a literal as an object and preprocessing of the data is limited to a minimum, for instance transforming a date from datetime to

[43] https://op.europa.eu/en/web/eu-vocabularies/at-dataset/-/resource/dataset/resource-type.

[44] However as opposed to the strict definition in the R2RML standard, note that we speak herein also about direct mapping, when minor, straightforward syntactic literal transformations are applied.

date format, for instance for the properties `dcterms:date`, `dcterms:issued`, `eli:first_date_entry_in_force`, `eli:date_no_longer_in_force`, *eli:date_document* and `eli:date_publication` in ISO 8601[45] format (YYYY-MM-DD). Other properties that have a literal as their object, such as `eli:title`, `eli:title_short` and `eli:title_alternative`, are transformed without modification.

Indirect population This approach is used when there is data available in a structured format that cannot be directly fed into the legal knowledge graph, for instance in case of resource types represented as simple strings in the database which need to be mapped to/replaced with the AustroVoc vocabulary terms based on mappings between the input and the output data, or where linking requires additional lookups or conditionals. In more detail, RIS document types are indicated as strings or integers in the database but we created a concept scheme `av:resource-types` as suggested by the ELI and ECLI ontologies in AustroVoc. For instance, a legal provision of type *"BG"* (federal law) is replaced with the AustroVoc concept `av:leg_bg`, where the resource can be linked to its type using the properties `eli:type_document` for legislative documents and `dcterms:type` for judiciary documents. We proceed similarly when it comes to mapping the law index of legal provisions using the property `eli:is_about`. The law index item is also replaced with the corresponding `av:bundesrechtindex`. To assign judiciary documents a class we use the legal provisions mentioned in the text, look up the law index for each of the found legal provisions and assign the law index to the judiciary document in order to populate the `dcterms:subject` property for each judiciary document. Furthermore, references extracted from the document text are strings which need to be replaced with the actual URI of the referenced documents and linked using the `dcterms:references` and `eli:cited_by_case_law` properties.

Population by interlinking external sources Although the RIS database contains relevant legal information—for instance, legal provisions and court decisions—it does not provide additional structured background information that could also be interesting in terms of enhancing the legal search process by adding respective search attributes as well as enabling advanced analysis of the legal system. Such background information includes for instance spatio-temporal information about geographic entities or events mentioned in court decision, for instance the deciding courts or case relevant dates. Similar techniques for enhancing search by interlinking information from spatio-temporal knowledge graphs have already proven successful for Open Data search (Neumaier and Polleres 2019). As for geo-references, we enhance the court information with external data from Nominatim,[46] the search engine of OpenStreetMap (OSM),[47] and Geonames[48] from which we get an RDF dump we import in our legal knowledge graph. In order to get information about the Austrian courts we compile a list of court names and query Nominatim for address

[45] https://www.iso.org/iso-8601-date-and-time-format.html.

[46] https://nominatim.openstreetmap.org/.

[47] https://www.openstreetmap.org/.

[48] https://www.geonames.org/.

```
<https://data.wu.ac.at/legal/court#court_8>
    rdf:type
        lkg:Court ;
    rdfs:label
        "Bezirksgericht Leopoldstadt" ;
    lkg:cout_type
        av:bg ;
    lkg:located_in_community
        <http://sws.geonames.org/2772614/> ;
    lkg:located_in_country
        <http://sws.geonames.org/2782113/> ;
    lkg:located_in_district
        <http://sws.geonames.org/2761333/> ;
    lkg:located_in_state
        <http://sws.geonames.org/2761367/> ;
    rdfs:seeAlso
        <https://www.openstreetmap.org/relation/1651546> .
```

Listing 6 Example *Bezirksgericht Leopoldstadt* (shortened, serialized in Turtle)

information, for instance for *Bezirksgericht Leopoldstadt*.[49] The result has an entry *display_name* containing address information such as street, community, district, state and country. We extract this information and use Geonames in order to populate the properties `lkg:located_in_community`, `lkg:located_in_district`, `lkg:located_in_state` and `lkg:located_in_country` as shown in Listing 6, where the new information is highlighted in red. In addition we also include the OSM court information page using `rdfs:seeAlso` which allows users of the legal information system to retrieve location and contact information for the respective authorities.

4.2.2 Population from unstructured data

While some of the structured information contained in the RIS metadata is incomplete or not all attributes we are interested in are covered as metadata fields, some of this missing information can be extracted from the document text using Natural Language Processing (NLP) tools and techniques. Extracting entities from a text and classifying them into a set of classes (e.g. person, organization, etc...) is called Named Entity Recognition (NER) (Grishman and Sundheim 1996). In our case we extract legal entities, such as courts, legal provisions and law gazettes. For instance, court decisions contain references to other documents that are not available in the metadata, such as legal provisions and legal rules mentioned in the court decision text. We note though, that rather than structured hyperlinks, the references used in legal practice are oriented on the use by humans and therefore use simple textual labels such as *§ 28a KSchG* rather than URIs like `https://www.ris.bka.gv.at/eli/BGBl/1979/140/P28a/NOR40180997` to reference a legal provision. In order to transform such unstructured references to machine-readable links in our KG we therefore extract such textual entities to find corresponding ELI or ECLI identifiers of referenced documents, linking both documents

[49] https://nominatim.openstreetmap.org/search/BezirksgerichtLeopoldstadt?polygon_geojson=1&format=json&countrycode=AT&type=administrative.

with the properties `dcterms:references` (`lkg:JudicialResource -> lkg:LegalProvision`) and vice versa `eli:cited_by_case_law` (`lkg:LegalProvision -> lkg:JudicialResource`). Multiple approaches are available to extract information from document text, which could help us to link the documents with each other. We herein specifically compare a *rule-based approach* used in combination with *gazetteers* with more advanced approaches such as *conditional random fields* and *deep learning*. A comparative assessment of these orthogonal approaches helps to increase confidence in the extraction results in the legal domain.

Corpus. For a performance comparison between the different approaches we need an annotated training corpus of legal documents. To the best of our knowledge, there is no gold standard Austrian legal corpus available, thus we manually annotate 50 randomly selected decision texts from the Justice branch. The documents have are quite varied in length with an average of 11,669 tokens with ± 7,741.88 tokens standard deviation (SD), and 260.12 (± 262.71 SD) sentences. For the population of our knowledge graph we extract the following legal entities: *Case reference* is a reference to another decision text which is used to refer to decisions taken or arguments brought up in previous cases. In the corpus a document contains on average 33 (± 23 SD) case references. *Contributor* contains the names of the judges involved in a decision. The number of judges involved in a decision amounts 5 (± 2 SD) which is caused by the different compositions of the senates. *Court* is mentioned in the decision text to indicate the court taking the decision, but there are also courts in the appeal stages. courts are mentioned 15 (± 6 SD) times in a document. *Legal rule* is a summarizing statement of a ruling from which general rules are inferred and are often cited in decision texts to back up the decision. Legal rules are cited 23 (± 22 SD) times on average in the documents of the corpus. *Legal provision* is mentioned in the decision text and forms the legal basis on which the decision is grounded. Court decisions must be based on the law, it is therefore not surprising that 87 (± 72 SD) legal provisions are cited on average. *Law Gazette* is cited in cases where the court wants to refer to a specific version of law. A law gazette is usually cited together with a legal provision to indicate the specific version the court is referring to. Given the purpose of citing a law gazette in a court decision the number of citations is on average 4 (± 6 SD) per document. *Literature* is used to cite legal literature used to back up the decision. We also extract these references as they are with 50 (± 36 SD) citations on average an thus constitute a very important source. However, the literature is mostly (at least in Austria) only available against a paid subscription from various legal publishers.

Rule based approach. Given that legal documents follow a relatively regular structure and citation style we apply a rule-based approach for the information extraction using the Java Annotation Pattern Engine (JAPE) (Cunningham et al. 1999) which is part of the General Architecture for Text Engineering (GATE).[50] An example of how we can exploit the standardized citation style in legal documents is shown in Listing 7, which illustrates a (shortened) JAPE rule used to extract references to legal rules in a court decision. A JAPE rule has a left hand side where the rule is defined and a right hand side that defines what to do with the extracted information, with both sides

[50] https://gate.ac.uk/.

```
Input: Token
Rule: rs
(
    {Token.string == "RS"}
    {Token.kind == "number"}
):rs
-->
:rs.LegalRule = {legalrule = :rs@string}
```

Listing 7 Example snippet JAPE rule for the extraction of legal rules

separated with a -- >. After a tokenizer (splitting the text into its individual parts) has been applied, the JAPE rule takes a *Token* as an input and looks for the defined pattern in the *Rule* section. In this example a legal rule must start with a token with a string *RS* directly followed by a token of kind *number*. The returned result is the complete the legal rule string, for instance *RS0042781* which we can look up in the database in order to replace the literal text with its actual URI, thus generating a link between the two documents. Rules can easily be supported by gazetteers, which are lookup lists that are very suitable for static, recurring entities, hence entities that do not change frequently. We use gazetteers to assist with the detection of contributors (a list with most common names and academic degrees), courts, legal provision (a list with all law abbreviations) and literature (a list with the most common legal journals used in Austria). Note that for the rule based approach we included a score for a strict and a lenient evaluation. The strict evaluation of rules only counts occurrences as correct when the annotation of the rule matches the gold standard annotation exactly. Lenient results also count occurrences as correct when both annotations overlap with the rule (adding or omitting some words).

Conditional Random Fields An alternative, common approach to label textual sequence data using probabilistic models are Conditional Random Fields (CRF) (Lafferty et al. 2001). We use the implementation of the sklearn-crfsuite.[51] The features of a token, for instance position and casing, are used to calculate the probabilities of tokens following each other. In the legal domain CRF have already been used in the context of entity extraction tasks where it has shown good results [e.g. Dozier et al. (2010); Cardellino et al. (2017); Leitner et al. (2019)].

Deep learning approach For experiments involving embeddings and deep learning we use the Flair framework[52] which provides all the necessary functionality required for our evaluation and in addition also supports importing *pretrained German language models*, which we were hoping to boost the accuracy for our German legal document corpus. We compare the following language models: (i) *Flair*, which uses contextualized character level embeddings (Akbik et al. 2018) trained on a mixed corpus of web and Wikipedia documents; (ii) Language models using a transformer based architecture (Vaswani et al. 2017) provided by HuggingFace[53] (Wolf et al. 2019) known as *Bidirectional Encoder Representations from Transformers (BERT)* (Devlin et al.

[51] https://sklearn-crfsuite.readthedocs.io/en/latest/.

[52] https://github.com/flairNLP/flair.

[53] https://huggingface.co/.

Table 3 Evaluation results of legal entity extraction (*P=Precision, R=Recall, F=F-score. Best results highlighted in boldface.*)

			Case reference	Contributor	Court	Legal provision	Law gazette	Legal rule	Literature
Rule based	Rules strict	P	0.9782	0.7631	0.9892	0.8742	0.9150	1	0.6814
		R	0.9817	0.9406	0.9659	0.9074	0.9683	1	0.7865
		F	0.9799	0.8426	0.9774	0.8905	0.9409	**1**	0.7302
	Rules lenient	P	0.9806	0.7631	0.9919	0.8923	0.9200	1	0.8095
		R	0.9842	0.9406	0.9685	0.9262	0.9735	1	0.9343
		F	**0.9824**	0.8426	**0.9801**	0.9090	0.9460	**1**	0.8674
CRF	CRF	P	0.9868	0.9161	0.9852	0.9452	0.9638	0.9994	0.9145
		R	0.9710	0.9557	0.9416	0.9483	0.9364	1	0.8611
		F	0.9787	0.9328	0.9616	0.9459	**0.9473**	0.9997	**0.8866**
Deep Learning	Flair	P	0.9783	0.9187	0.9455	0.9324	0.9263	1	0.8596
		R	0.9800	0.9780	0.9486	0.9526	0.9245	1	0.8671
		F	0.9791	0.9435	0.9456	0.9414	0.9215	**1**	0.8629
	BERT	P	0.9687	0.9481	0.9557	0.9447	0.9546	0.9971	0.8497
		R	0.9738	0.9710	0.9762	0.9536	0.9336	1	0.8409
		F	0.9712	**0.9583**	0.9654	0.9489	0.9396	0.9986	0.8448
	DistilBert	P	0.9759	0.9316	0.9407	0.9446	0.9392	0.9979	0.8663
		R	0.9786	0.9878	0.9784	0.9600	0.9529	1	0.8604
		F	0.9772	0.9551	0.9586	**0.9521**	0.9437	0.9989	0.8626

2019) trained on German Wikipedia, German open legal data and news articles; and (iii) *DistilBERT* (Sanh et al. 2019) a faster and smaller version of BERT also trained on Wikipedia articles and web documents. DistilBERT uses a teacher-student setting to distill the knowledge from the teacher (the BERT model) to the student (DistilBERT model).

Evaluation For the evaluation of the individual results we measure *Precision (P)* as the share of relevant from the retrieved documents, *Recall (R)* as the share of retrieved documents to all documents that should be retrieved and *F-score (F)* as the harmonic mean of P and R (Manning et al. 2008). For our experiments we did not apply any preprocessing to the documents and apply a 5-fold cross-validation approach using a train/test/validation split of 80%/10%/10%. All models have been trained with default settings, in particular the deep learning models with a maximum of 150 epochs, starting learning rate of 0.1, patience 3 and an anneal factor of 0.5. The training stops automatically when the learning rate becomes too small.

Table 3 shows the results for the different legal entities, whereby approaches with the best F-scores are highlighted in boldface. Looking at the evaluation results we can see at first glance that there is no single clear best approach outperforming all other approaches on all legal entities. Furthermore, it can also be noted that the results of all extraction methods are comparable across all methods for the individual legal entities. In particular, the numbers show that rules perform well when the entities under

investigation are highly structured and always follow the same pattern, for instance case reference (e.g. *14Os108/20v)* and legal rule (e.g. *RS0042781*) which are very easy to recognize. Moreover, we use gazetteers to support rules with the extraction of the contributors. The rule looks for a degree (from a gazetteer) followed by a last name (from a gazetteer) within the head of the document. The inclusions of additional sources already decreases the performance of the rule based approach and automatic approaches perform better. When adding more variations and more complexity to the legal entities the performance of the rule-based and gazetteer supported approach deteriorates and machine learning based approaches perform better. The numbers of the legal provision, law gazette and literature show this effect. The citations of legal provisions can be simpler (e.g. *§ 41 ZPO* and more complex (e.g. *§§ 41, 43 Abs 2 erster Fall und § 50 ZPO*) which adds a lot of complexity to the rules and as a result makes the result much harder to create. The citations of the law gazettes changed over time by adding additional information (e.g. from *BGBl. 1969/207* to *BGBl. I Nr. 134/2015*). The most complex entity to extract is the literature as there are various types of literature (e.g. commentaries, books, articles,..) and citation styles. The higher complexity for literature is also reflected in the evaluation results. While the best F-scores for the other legal entities are somewhere in the 94% range, the F-score for literature is achieved by CRF with only 88%. The numbers also show that the gap between the rules and automatic approaches is bigger the more complex the rules (with gazetteer support) need to be. However, the gap between the individual approaches is very small. The F-scores of the three deep learning approaches (Flair, BERT, DistilBERT) are within 2% across all legal entities, thus we cannot nominate a clear winner in this segment. Also the difference across all approaches and legal entities falls within a range of 4%.

While the evaluation results show that the extraction approaches perform mostly equally well, we also should take into account the effort that is required to set up such a system for the extraction of legal entities. Rules can be easily and quickly created with only a few sample documents that cover the possible variations in which legal entities can appear. In addition, rules are easy to interpret and explain. The outcome of a rule is clear from the beginning, as a rule either matches a sequence of tokens or not. Gazetteers are suitable for entities that do not change frequently, for instance courts or names, but have a maintenance requirement and might need to be updated on a regular basis, otherwise rules using these gazetteers will start to fail over time. By contrast, approaches using (deep) machine learning promise to be more flexible and are also able to cover variations in patterns where a rule would fail. However, these approaches are less explainable and predictable, hence working with probabilities of the results and selecting the right algorithm for the right task is necessary.

In addition, we remark that it requires considerable effort to annotate documents required for training machine learning approaches as well as computational power and resources to perform both training and model fine-tuning. In our case, the experiments with our corpus of only 50 documents used the full capacity of our machine with 16GB of memory and requires a powerful GPU (a GTX 1080 Ti with 16GB memory) to perform the computations in a timely manner.

Summarizing the results shown by the experiments there is no clear best approach to extract legal entities from text. Thus the approach should be chosen based on the requirements, the available data from the legal information system acting as a data

source and human resources. We conclude in particular that rules, in combination with gazetteers, are a viable alternative and can keep up with state of the art NLP techniques using complex neural networks for the relatively well-structured texts in our domain, offering maintainability and explainability of extraction results.

4.2.3 Alignment of heterogeneous schemes

Last, but not least, our AustroVoc vocabulary, which is composed of terms specific to the Austrian legal system, contains for instance a law index which is very suited to be linked with related terms in EuroVoc, thereby, directly enabling a multi-lingual search (given that EuroVoc is available in multiple languages). As the main obstacle herein, legal language is diverse even within German speaking countries, plus EuroVoc contains "German" German whereas Austria often uses specific "Austrian" German terms, for instance we use the term *Konsumentenschutz* while EuroVoc contains the term *Verbraucherschutz* for "customer protection". Since we want to link the concepts of the Austrian law index with EuroVoc concepts, we adopt the approach described in Filtz et al. (2018). The simplest way to find a match is a direct lookup of the Austrian term in EuroVoc, if no match is found we also include external knowledge bases such as *DBpedia*,[54] *Wikidata*[55] and the *Standard Thesaurus Wirtschaft (STW)*[56] and search for additional language version of the term there. In case a match is found we can link the AustroVoc term with the corresponding EuroVoc term using the property `rdfs:seeAlso`, for instance we find a match from *Konsumentenschutz* to *Verbraucherschutz* and add the triple `av:bri2006 rdfs:seeAlso ev:2836` to AustroVoc as shown in Listing 5.

5 The European legal knowledge graph

Our final objective is to integrate the Austrian legal knowledge graph with other national legal knowledge graphs, which should enable interlinkage across different countries. We herein analyze the current situation regarding the provision of linked legal data as well as legal databases in other EU member states and perform a comparitive analysis. In addition to the legal information provided by governments, we also include a selection of non-governmental initiatives[57] and summarize challenges and opportunities we faced during this process.

5.1 Legal information provided by Governments

We include the EU member states without the United Kingdom and EU candidate countries in our analysis of whether and how they make legal information available

[54] https://wiki.dbpedia.org/.

[55] https://www.wikidata.org/.

[56] http://zbw.eu/stw/version/latest/about.

[57] We do not include commercial solutions.

in machine-readable form. We use the EU e-Justice portal[58] as a starting point for our research process, which includes overview pages on which EU member states can provide additional information about their implementation, for all EU member states for ELI[59] and ECLI.[60] While the country-specific ECLI information page contains all EU member states, the ELI information page only has information for 17 countries. Typically an explanation and examples are included as well as links to national legal databases. Some countries provide detailed information about their deployed ELI/ECLI structure while others do not provide any information or, respectively, only in the national language which needs to be translated using a translation service. When available, we followed the links provided, otherwise we used a search engine to manually find additional national legal databases and examples for legislative and judiciary documents (cf. Tables 10 and 11 (Appendix A.4) for links to databases and examples). In the first step we examine whether ELI/ECLI identifiers are visible in the document and in the second step we also scan the source code of the (HTML) document, searching the metadata for keywords such as *eli, ontology, dc, dcterms, creator* and *date*. We provide an overview of the properties used in the Appendix for ELI (Table 9) and ECLI (Table 7). Where we find metadata embedded in the document we parse the URL using EasyRdf[61] to automatically retrieve RDF triples per document. We also check whether countries use national Named Authority Lists (NALs), i.e. determine whether national information pages about the used NAL are provided. In addition to this search process on the national level we also queried the EU Open Data Portal[62] for national legal data. We also record per country the type of available search interfaces, available document formats, languages and availability of judiciary documents in the EU ECLI search engine.

Table 4 provides a comprehensive overview of the national ELI and ECLI implementation initiatives of the EU member states with a focus on the ELI/ ECLI implementation status. The columns *Implementation ELI* and *Implementation ECLI* describe the implementation status with *Identifier* referring to the situation where documents are given an ELI identifier and *Identifier/Metadata* indicates that the particular country also provides metadata for the documents. The general assumption is that all countries use the ELI ontology for legislative documents (and ECLI for judiciary documents respectively), but some countries provide national extensions in order to represent legal information based on national requirements. These additional ontology extensions are indicated in brackets, for instance Finland defined its own extensions for ELI in the *Semantic Finlex Legislation Ontology (SFL)*[63] and the *Semantic Finlex Case Law Ontology (SFCL)*[64] ontology for judiciary documents. Luxembourg also

[58] https://e-justice.europa.eu/.

[59] https://eur-lex.europa.eu/eli-register/implementation.html.

[60] https://e-justice.europa.eu/content_european_case_law_identifier_ecli-175-en.do?init=true.

[61] http://www.easyrdf.org/.

[62] https://data.europa.eu/euodp/en/home.

[63] http://data.finlex.fi/schema/sfl/.

[64] http://data.finlex.fi/schema/sfcl/.

Table 4 Linked legal data feature comparison of EU member states (*(+) indicates the usage of additional or other ontologies*)

Country	Implementation ELI	Implementation ECLI	Data Availability	Information ELI / ECLI / NAL	Thesaurus
Austria	Identifier	Identifier	–	✓ / – / –	✓
Belgium	Identifier	Identifier	–	✓ / ✓ / –	–
Bulgaria	–	Identifier	–	– / ✓ / –	–
Croatia	Identifier	Identifier	–	✓ / – / –	–
Cyprus	–	–	–	– / ✓ / –	–
Czech Republic	–	Identifier	–	– / ✓ / –	–
Denmark	Identifier/Metadata	–	RDF	✓ / ✓ / ✓	–
Estonia	–	Identifier		– / ✓ / –	✓
Finland	Identifier/Metadata (+)	Identifier/Metadata (+)	RDF	✓ / ✓ / ✓	✓
France	Identifier/Metadata	Identifier	RDFa	✓ / ✓ / –	–
Germany	–	Identifier/Metadata	–	– / ✓ / –	–
Greece	–	Identifier	–	–	–
Hungary	–	–	–	✓ / – / –	–
Ireland	Identifier/Metadata	–	RDFa, RDF	✓ / ✓ / –	–
Italy	Identifier/Metadata	Identifier	RDFa, RDF	✓ / ✓ / ✓	–
Latvia	– (+)	Identifier	–	– / ✓ / –	–
Lithuania	–	–	–	– / ✓ / –	✓
Luxembourg	Identifier/Metadata (+)	–	RDFa	✓ / – / ✓	–
Malta	–	Identifier	–	✓ / – / –	–
Netherlands	– (+)	Identifier/Metadata	RDFa, RDF	– / ✓ / –	–
Poland	–	–	–	–	–
Portugal	Identifier/Metadata	Identifier (+)	RDFa	✓ / – / –	–
Romania	–	Identifier	–	– / ✓ / –	–
Slovakia	–	Identifier	–	– / ✓ / –	✓
Slovenia	– (+)	Identifier	–	– / ✓ / –	–
Spain	Identifier/Metadata	Identifier	RDFa	✓ / – / ✓	–
Sweden	–	–	–	–	–

uses an additional ontology called *JOLUX*[65] in their *Casemates* project[66] incorporating the ELI ontology and extending it. Special cases are Latvia and Slovenia who do not participate in the ELI and therefore also do not assign ELI identifiers to their legislative documents but do provide a basic set of metadata (which is less than and different to ELI) using the *Open Graph Protocol (OGP)*.[67] Portugal assigns an ECLI identifier to judiciary documents, but uses OGP for the metadata. The Netherlands use for their legislative documents the *dcterms* and *Overheid* ontologies. We can see that 11 out of 27 countries implemented at least the first pillar of the ELI ontology (i.e. assigning an ELI identifier to the documents), hence giving an ELI identifier to

65 https://data.public.lu/en/datasets/r/53aa1301-2a42-465a-8803-c0cb5a3589e7.

66 http://www.legilux.lu/editorial/casemates.

67 https://ogp.me/.

legislative documents. Participation/Implementation is better in terms of ECLI, where 19 countries assign an ECLI identifier to judiciary documents, but the number of countries providing machine-readable metadata (i.e.,3) is lower compared to ELI (i.e.,9). Compared to a study conducted in 2017 (van Opijnen et al. 2017b) the participation in ECLI increased in the last years with additional seven countries now participating in ECLI with at least providing an ECLI identifier. The column *Data Availability* describes how the data is provided to the public with the majority of participating countries opting to use the RDFa format and embed the metadata in the source code of the document. Denmark, Finland, Ireland and Italy also allow users to download the data in RDF either from a national website or the European Open Data Portal. The Netherlands provide a web service[68] that can be used to download the data in RDF. We indicate whether information about the national implementation of ELI and ECLI as well as the usage of NAL is provided either using dedicated pages on the EU e-Justice portal or a national website. Some properties are very suitable for the usage of NAL, for instance `eli:language` or `eli:type_document`. An overview of the used NALs is provided in Appendix A.2, Table 8. We notice that there are more countries using NALs, however they do not all provide an information page. A thesaurus, such as EuroVoc or a national index of legal terms, is used by five countries as indicated in column *Thesaurus*.

We show the features of the EU member states' legal databases in Table 5. Central search interfaces are very convenient as users can find all the required information in the same place. However, as legal systems are typically divided into legislation and judiciary the information for both branches falls under the responsibility of different authorities and therefore provided at distinct places. The column *Central Interface* shows if there is a central interface available that enables users to access legislation as well as judiciary documents from different authorities even if they are stored in separated backend systems. The EU e-Justice portal contains an ECLI search engine[69] which enables users to search for ECLI identifiers and keywords in judiciary documents from multiple countries, but not all countries providing an ECLI identifier are also participating in the ECLI search engine. The *Search Interface* column indicates how the search process can be performed by users with the majority of countries providing a keyword-based search interface, which might be enhanced with additional filters, for instance to restrict dates to a certain time frame or select only special types of documents. Faceted search interfaces are implemented by a minority of countries only, *Both* means that one legal database provides a keyword-based search and the other legal database supports faceted search. We can also see that Finland and Luxembourg set up a public SPARQL endpoint which allows users to run structured queries on the data directly. The standard way to represent legal documents on the web is HTML as shown in column *Document Format*. While the content is displayed using HTML, the majority of legal information systems also allow users to download documents in PDF format. However, some countries provide documents in PDF only. A popular structured format is XML, supported by Austria, Estonia, Luxembourg and Spain. The EPUB format is only used in Spain. While it is clear that countries provide their

[68] https://linkeddata.overheid.nl/front/portal/services.

[69] https://e-justice.europa.eu/content_ecli_search_engine-430-en.do.

Table 5 Features of legal databases of EU member states (* *denotes a subset*)

Country	Central Interface	ECLI Search	Search Interface	Document Format	Languages
Austria	✓	–	Keyword	HTML, PDF, RTF, XML	DE, EN*
Belgium	–	✓	Keyword	HTML	FR, NL, DE
Bulgaria		✓	Keyword	HTML, PDF	BG
Croatia	–	✓	Keyword	HTML	HR
Cyprus	✓	–	Keyword	PDF	EL
Czech Republic	–	✓	Keyword	PDF	CZ
Denmark	–	–	Faceted	HTML, PDF	DK
Estonia	✓	✓	Keyword	HTML, PDF, TXT, XML	EE, EN*
Finland	✓	✓	Keyword, SPARQL	HTML	FI, SE
France	✓	✓	Keyword	HTML, PDF	FR, EN*, DE*, IT*, ES*
Germany	–	✓	Keyword	HTML	DE, EN*
Greece	–	✓	Keyword	PDF	EL
Hungary	–	–	Keyword	HTML	HU, EN*
Ireland	–	–	Keyword	HTML, PDF	EN
Italy	–	✓	Keyword	HTML	IT
Latvia	–	✓	Keyword	HTML, PDF	LV, EN*, RU*
Lithuania	–	–	Faceted	HTML, PDF	LT
Luxembourg	–	–	Faceted, SPARQL	HTML, PDF, XML, RDF	FR
Malta	–	–	Keyword	PDF	MT, EN
Netherlands	–	✓	Both	HTML, PDF, RDF	NL, FR, EN*
Poland	–	–	Keyword	PDF	PL
Portugal	✓	✓	Faceted	HTML, PDF	PT, EN*
Romania	–	–	Keyword	HTML	RO
Slovakia	✓	–	Keyword	HTML, PDF	SK
Slovenia	–	✓	Keyword	HTML, PDF, DOCX	SI, EN*
Spain	–	✓	Both	HTML, PDF, XML, EPUB	ES
Sweden	–	–	Keyword	HTML	SE

documents in their official language(s), Austria, Estonia, France, Germany, Hungary, Latvia, Netherlands, and Slovenia publish a subset of their documents, mainly the documents considered to be most important such as the constitution or the civil code, also in English.

5.2 Non-governmental initiatives

Besides linked legal data initiatives driven by governments there are also efforts by academia and industry in this direction often conducted in collaboration with and funded by governments. We are particularly interested in non-governmental initiatives working with ELI and ECLI providing a linked legal data framework or focusing on special legal areas.

Table 6 shows an overview of several non-governmental initiatives across Europe based on the information provided by the project websites, publications or namespaces used in RDF data retrieved via a SPARQL endpoint. The column *Project* shows the title of the project. We classify the projects as indicated in column *Type* into the classes *Linking* which means that this project aims to link legal data with other legal other data or external knowledge bases and *Extraction* means that the project is focusing on the extraction of specific information contained in legal documents. Column *Using ELI / ECLI* indicates whether a the project uses ELI, ECLI or both. In cases where the project results in extensions to the ELI and ECLI ontologies the name of these extensions is listed in column *Extension ELI / ECLI*. In cases where data is made available for download the format is shown in column *Data Availability*. Column *Thesaurus* indicates whether the European thesaurus EuroVoc or other thesauri (e.g. a national thesaurus) is used. When the data used in the project is linked with other external data such as DBpedia or Geonames this is indicated in column *Open Data Linking*. The column *SPARQL* shows whether a SPARQL endpoint is available to retrieve the data from that project.

The *Legal Knowledge Graph* project that aims to integrate legal data from disparate legal databases into a knowledge graph is described in Sect. 4. The *Semantic Finlex Project*[70] (Oksanen et al. 2019) carried out by the University of Aalto is, similar to our Austrian research project, based on the national legal database of Finland which contains legislative and judiciary documents, and transforms the data into linked legal data based on the ELI and ECLI ontologies. The results of this Finnish project are also visible in Table 4 as they are available to the public via the official Finlex website,[71] as well as via a SPARQL endpoint.[72] Finlex extends the ELI with the Semantic Finlex Legislation Ontology[73] (SFL) and ECLI with the Semantic Finlex Case Law ontology[74] (SFCL). The greek project *Nomothesia* (Chalkidis et al. 2017) by the University of Athens focuses on legislation only and is based on legal documents published in PDF format which are transformed into linked legal data based on the ELI which is

[70] https://seco.cs.aalto.fi/projects/lawlod/.

[71] https://data.finlex.fi/.

[72] https://www.ldf.fi/sparql-services.html.

[73] http://data.finlex.fi/schema/sfl/.

[74] http://data.finlex.fi/schema/sfcl/.

Table 6 Non-governmental initiatives using ELI and ECLI

Project	Type	Using ELI / ECLI	Extension ELI / ECLI	Data Availability	Thesaurus	Open Data Linking	SPARQL
Legal Knowledge Graph	Linking	✓ / ✓	LKG / LKG	RDF	EuroVoc, Other	✓	✓
Semantic Finlex	Linking	✓ / ✓	SFL / SFCL	RDF	EuroVoc, Other	✓	✓
Nomothesia	Linking	✓ / –	Nomothesia / –	RDF	–	✓	✓
EUCases	Linking	✓ / ✓	–	–	EuroVoc, Other	–	–
Lynx	Linking	✓ / –	Lynx-LKG	RDF	EuroVoc, Other	✓	✓
GDPRtEXT	Linking	✓ / –	GDPRtEXT	RDF	–	–	✓
Linkoln	Extraction	✓ / –	–	–	–	–	–
BO-ECLI	Extraction	– / ✓	–	–	–	–	–

incorporated in the Nomothesia ontology.[75] The data produced by the Nomothesia project is available for download as well as via a SPARQL endpoint[76] and includes DBpedia as an external knowledge base, for instance to link persons that are mentioned in legal acts. In the *EUCases* project (Boella et al. 2015) a first effort effort was made trying to link national and EU legislation and case law, which is no longer accessible because a login is required and there is no response to email requests.[77] This project also includes a proposal to link legal documents with the EuroVoc thesaurus and incorporates the Legal Taxonomy Syllabus (LTS) (Ajani et al. 2007). The EU funded *Lynx* project[78] aims at creating a legal knowledge graph with a special focus on compliance (Montiel-Ponsoda et al. 2017). This project includes Spanish legislation and jurisdiction as well as documents from selected countries and extends ELI and ECLI with the Lynx-LKG ontology.[79] The Lynx data can also be accessed via a SPARQL endpoint.[80] A legal domain-specific work is *GDPRtEXT*[81] extending the ELI to provide the General Data Protection Regulation (GDPR)[82] as a linked data resource together with a taxonomy of GDPR terms using SKOS (Pandit et al. 2018). The linked legal data version of the GDPR extends the ELI ontology with the GDPRtEXT ontology. The data and the ontology are available for download[83] and can be accessed via a SPARQL endpoint.[84] The Italian *Linkoln project* focuses on the automatic extraction of references from legal documents of the Italian Senate and is also able to extract ELI references (Bacci et al. 2019). The EU funded *BO-ECLI project*[85] running from 2015 to 2017 focused on the ECLI and investigated the implementation of the ECLI in selected countries resulting in a proposal of a new version of the ECLI due to discovered drawbacks (van Opijnen et al. 2017a).

6 Use case revisited

With an Austrian legal knowledge graph in place and a more complete picture of other similar international initiatives, we are now able to assess the potential benefits of linked legal knowledge both nationally and internationally. For instance, in terms of providing enhanced capabilities in terms of legal analyses, or in enabling us to answer complex search queries that would entail tedious manual research otherwise. Yet, we still herein have only made initial steps towards an EU wide linked legal data graph, wherefore we also discuss additional required steps and a respective roadmap.

75 http://legislation.di.uoa.gr/data/ontology.

76 http://legislation.di.uoa.gr/endpoint.

77 http://www.eucases.eu.

78 http://www.lynx-project.eu/.

79 http://lynx-project.eu/doc/lkg/.

80 http://sparql.lynx-project.eu/.

81 https://openscience.adaptcentre.ie/projects/GDPRtEXT/.

82 https://eur-lex.europa.eu/eli/reg/2016/679/oj.

83 https://old.datahub.io/dataset/gdprtext.

84 http://openscience.adaptcentre.ie/sparql.

85 https://bo-ecli.eu/.

6.1 Benefits of an integrated legal knowledge graph

Let us revisit the questions from Sect. 3: indeed, we can demonstrate the benefits of an integrated legal knowledge graph by underpinning them with example SPARQL queries providing answers to such questions.

– *Which documents are referenced in a specific court decision?*

Court decisions are based on the law and therefore reference legal provisions but also other court decisions and legal rulings. Users nowadays typically need to query the respective database, e.g. the law database for legal provisions, and manually search the referenced document in order to get the content. In a knowledge graph we can combine several involved steps into a single query that returns a court decision with all referenced documents, their texts, plus types of the documents. This leads to a more **efficient legal information search process**. To enable such a query we need to extract the referenced documents from the court decision and replace them with the respective URIs as well as a schema of document types. Example 2 shows the convenience of such a query when a lawyer is interested in a particular court decision and gets all referenced documents with their text and sorted by their types as result.

Example 2 SPARQL Query: Which documents are referenced in the Supreme Court decision with case number 10Ob12/16m?

```
SELECT DISTINCT ?Reference ?Text ?Type
WHERE {
  ?justiz rdfs:label "10Ob12/16m" .
  ?justiz dcterms:references ?ref .
  {
    ?ref rdf:type lkg:LegalProvision ;
         rdfs:label ?Reference ;
         eli:is_realized_by ?realization .
    ?realization lkg:has_text ?Text .
    ?ref eli:type_document ?type_document .
    ?type_document skos:prefLabel ?type .
  } UNION {
    ?ref rdf:type lkg:JudicialResource ;
         dcterms:type av:jud_rs ;
         rdfs:label ?Reference ;
         lkg:has_text ?Text .
    av:jud_rs skos:prefLabel ?type .
  } UNION {
     ?ref rdf:type lkg:JudicialResource ;
         dcterms:type av:jud_te ;
         rdfs:label ?Reference ;
         lkg:has_text ?Text .
     av:jud_te skos:prefLabel ?type .
  }
    FILTER (lang(?type) = 'de')
}
ORDER BY ?type
```

Reference	Text	Type
"§ 500 ZPO"	"§ 500. (1) Das Urteil oder der Beschluß [...]"	"Bundesgesetz"
"§ 28a KSchG"	"§ 28a. (1) Wer im geschäftlichen Verkehr [...]"	"Bundesgesetz"
"4OB89/88"	"Ein Veröffentlichungsbegehren im Sinne [...]"	"Rechtssatz"
...	...	...

– *Over which districts does a court have competent jurisdiction?*

Legal databases are typically domain-specific and focus on legal matters only without additional contextual references that would be useful to be included for scoping search (such as explicit spatio-temporal references). For instance, a lawyer has a client who is facing a lawsuit regarding a property. Therefore, the lawyer needs to know which court has spatial competent jurisdiction, in order to find related cases in a regional context. At the moment this information is not made explicit in the legal information system and the lawyer would need to look through various websites of the authorities to find out about the regionally competent jurisdiction. This problem can be addressed by integrating external data in our legal knowledge graph and leads to **enriched information content and better user experience**. Since, in our knowelgedge graph, we have readily linked the information about the Austrian courts and the judicial districts from the respective authorities with a geospatial hierarchy, also taking the court hierarchy into account, we can easily provide such information again by a straightforward SPARQL query. As shown in Example 3 the lawyer is now able to query the court having competent jurisdiction, just by providing the name of a community.

Example 3 SPARQL Query: Which court has competent spatial jurisdiction for the market town *Krieglach*?

```
select ?court where {
    ?geo gn:name "Krieglach" .
    ?jd lkg:judicial_district_member ?geo ;
        lkg:court_having_jurisdiction ?c .
    ?c rdfs:label ?court
}
```

Court
"Bezirksgericht Mürzzuschlag"

– *What are the national transpositions of a specific EU directive?*

Legal systems differ across countries but still we need to consider legal information from other countries from time to time, especially in a European context with the EU's harmonization activities through issuing commmon regulations, but also directives, which need to be transposed into national legislation. For companies wanting to expand their businesses abroad it is necessary to know the legal situation and standards in these foreign countries. So far, a lawyer needs to search for the legal information system of the other country and find out how a particular directive, that is relevant for the company, has been transposed.[86] Also, the Eur-Lex search interface is not always helpful here, because it does not provide the transposed texts. Integrating legal data across countries in a legal knowledge graph thus would enable **cross-jurisdictional search of legal information**. In our example, we demonstrate how this can be achieved, across countries that follow the proposed ELI and ECLI standards for legal data (cf. Sect. 5). As shown in Example 4 the company lawyer is able to find the concrete national transpositions of a given directive with the actual transposed texts, across national legislations, again with a single query.

[86] Further tedious search would be needed to find out about and compare respective jurisdictions across countries.

Example 4 SPARQL Query: What are the national transpositions of EU directive 2014/92/EU (with links to the resp. documents)?

```
select ?country ?title ?document where {
    VALUES ?format {
        <http://www.iana.org/assignments/media-types/text/html>
        <http://www.iana.org/assignments/media-types/application/html> }
    ?n ?p <http://data.europa.eu/eli/dir/2014/92/oj> ;
           eli:relevant_for ?c ;
           eli:is_realized_by ?r .
    ?r eli:title ?title ;
       eli:is_embodied_by ?document .
    ?document eli:format ?format .
    ?c skos:prefLabel ?country .
   FILTER (lang(?country) = 'en')
}
```

Country	Title	Document
"Ireland"	"European Union (Payment Accounts) Regulations 2016."	Document 1
"Austria"	"Bundesgesetz, mit dem ein Bundesgesetz über [...]"	Document 2
"Austria"	"Verordnung der Finanzmarktaufsichtsbehörde (FMA) über [...]"	Document 3
...	...	

Document 1: http://www.irishstatutebook.ie/eli/2016/si/482/made/en/html
Document 2: https://www.ris.bka.gv.at/Dokumente/BgblAuth/BGBLA_2016_I_35/BGBLA_2016_I_35.html
Document 3: https://www.ris.bka.gv.at/Dokumente/BgblAuth/BGBLA_2018_II_60/BGBLA_2018_II_60.html

Further integrating and harmonizing existing legal knowledge graphs across countries, as discussed in Sect. 5 would further enable comparison of the respective jurisdiction for a particular directive, across countries.

– *Which legal documents regulate a specific legal area searched with keywords in a foreign language?*

Legal systems are not only different in their structure but legal documents are typically penned in the official language(s) of a country, which puts an additional language barrier in the legal information search process. Additional sources such as the EuroVoc thesaurus, ideally aligned with national thesauri, which contain terms in multiple languages to the legal knowledge graph enables **multi-lingual search of legal information**. Linking legal documents with concepts instead of language-specific labels allows users to search in their language for documents written in another language. For instance, a lawyer is researching in a lawsuit covering another country and wants to know which legal provisions cover a specific legal area and is able to search in his language as shown in Example 5. Different languages are a barrier and supporting multi-lingual search is a step towards improved, more transparent access to legal information.

Example 5 SPARQL Query: Which documents belong to the category consumer protection searched by an Italian?

```
select ?law ?legalprovision ?document where {
    ?ev skos:prefLabel "protezione del consumatore"@it .
    ?austrovoc rdfs:seeAlso ?ev .
    ?lp eli:is_about ?austrovoc ;
        eli:jurisdiction <http://publications.europa.eu/resource/authority/country/AUT> ;
        eli:in_force eli:InForce–inForce ;
        eli:is_realized_by ?le ;
        lkg:has_number_paragraph ?number ;
        rdfs:label ?legalprovision .
    ?le eli:title_alternative ?law ;
        eli:is_embodied_by ?document .
    ?document eli:format <http://www.iana.org/assignments/media–types/application/html>
}
ORDER BY ASC(?law) ASC(?number)
```

Law	Legal Provision	Document
"KSchG"	"§ 1 KSchG"	Document 1
"KSchG"	"§ 42 KSchG"	Document 2
"VKrG"	"§ 1 VKrG"	Document 3
...	...	

Document 1: https://www.ris.bka.gv.at/Dokumente/Bundesnormen/NOR12041200/NOR12041200.html
Document 2: https://www.ris.bka.gv.at/Dokumente/Bundesnormen/NOR40050352/NOR40050352.html
Document 3: https://www.ris.bka.gv.at/Dokumente/Bundesnormen/NOR40117826/NOR40117826.html

6.2 Roadmap towards a linked legal knowledge graph

The current situation towards a truly interconnected legal knowledge graph on a European level looks promising, with many good starting points, but some challenges lie ahead to be addressed. On the one hand, providers of legal information, typically governments, would need to help to ease the access to law and support non-governmental initiatives to provide and obtain legal information. On the other hand, these providers are confronted with resource restrictions and other priorities, which slows down this process. We discuss some of the related challenges in the following.

Licensing and access policies. The publication of and access to legal information might be hindered by licensing and access policies, or lack thereof. Open (government) data is a goal of the European Union as laid out in the PSI-Directive,[87] which stipulates that documents from the public sector should be made available free of charge in machine-readable and open formats which also includes possibilities for a mass download. The PSI directive goes hand in hand with the *8 Open Government Data Principles*[88] to provide data in a machine-readable, license free, complete and accessible format in a timely manner. Following open government data publication methodologies such as COMSODE (Kucera et al. 2015) helps governments to set up respective publication strategies. The terms and conditions should be communicated in a clear manner and data provided ideally under a permissive license which also allows private initiatives to use the data for their business model by providing additional services, e.g. build on the data and restrict access to certain parts of the knowledge graph such as linked legal commentaries.

[87] EU 2019/1024http://data.europa.eu/eli/dir/2019/1024/oj.

[88] https://public.resource.org/8_principles.html.

Support of linked legal data initiatives. Our analysis of the legal landscape (cf. Sect. 5) shows that documents are provided in various formats with structured formats being the minority. The problem of having documents in an unstructured format as a starting point [e.g. Chalkidis et al. (2017)] might slow down the process of the providing linked legal data. It is therefore desirable that legal documents are provided in a structured format from the very beginning in order to enable the transition to and participation in an EU-wide linked legal data ecosystem. Hence, following the Linked Data Principles together with using appropriate linked data formats such as JSON-LD (W3C JSON for Linking Data Community Group 2012) or RDF serializations or XML standards for legal documents, such as Akoma-Ntoso[89] enables easy access to the data for linked legal data initiatives. The EU can help member states in activities towards the provision of linked legal data by providing detailed guidelines on how to use the proposed ELI and ECLI standards or software tools supporting the transition. Furthermore, the provision of dedicated vocabularies in addition to the existing named authority lists and EuroVoc thesaurus, which do not really fit the requirements of member states, are beneficial as it reduces the barrier of participating in ELI and ECLI.

We emphasize here, that despite the resulting documents are typically plain text documents, in many countries—including Austria—the legal document preparation process is regulated by clearly defined processes where, as opposed to extracting unambiguous metadata on hindsight only—such metadata and linked data creation could and should be directly included into these processes. Respective tools that rely entirely on Open Web Standards could replace and improve the legal document creation process Beno et al. (2019).

Information provision The lack of coordination in terms of ELI and ECLI implementation concerns the European Union as well as EU member states. Currently, it is a very time-consuming task to find any information about ELI and ECLI implementations in different countries. At the moment the information is cluttered with some countries using the EU e-Justice portal or others providing respective information only on national websites. Furthermore, implementation details can often only be inferred from studying the source code of example documents, rather than by available documentation. Positive examples of countries providing extensive information are, for instance, Denmark,[90] Finland,[91] and Luxembourg[92] who run national websites with implementation information about the ELI. The same applies to the usage of NAL which is encouraged by the ELI and ECLI ontologies. Without additional information about the used NAL it is a tedious task for outsiders to find information which NAL are used. In addition to missing information websites about the NAL some countries use NAL but these NAL cannot be retrieved from the internet or dereferenced. As argued herein, aligning the ELI and ECLI pages at EU level, hence integrating ELI into the EU e-Justice portal, and providing templates for member states about their ELI and ECLI implementation status as well as the usage of national NAL could be highly beneficial. More consistent best practices would also help other, not yet participating

[89] http://www.akomantoso.org/.

[90] https://www.retsinformation.dk/eli/about.

[91] https://data.finlex.fi/en/datamodeling.

[92] http://www.legilux.lu/editorial/casemates.

countries to investigate what and how to implement ELI and ECLI in an overall more aligned manner, which in turn might lower the barrier to participate.

Search interfaces Access to legal information should be as easy as possible for end users as well as data processing professionals. Centralized web search interfaces serving as a *one-stop-shop* with a graphical user interface enabling the access to legal documents from various authorities eases the search process for the end user, citizens and legal professionals. Linked legal data initiatives enable such centralized aggregation of legal information, and can also support common application programming interfaces (API)—such as, e.g. access through the SPARQL protocol—as well as indexes to access and retrieve legal data for subsequent processing.

Multilinguality Legal data is typically presented in the official language(s) of the respective country, some of the legal information systems provide some laws (e.g. civil code and the constitution) in English. As demonstrated herein, one approach to enable better multi-lingual search is to link national indexes with the multi-lingual EuroVoc thesaurus which then acts as a connecting point between legal information provided in different countries and languages. Yet, we also emphasize the importance of national extensions (such as AustroVoc which we proposed in this paper) to cover countrywise specifics, or for keeping ambiguous language use in different legislations/jurisdictions (e.g. Germany and Austria) separate. We envision the creation of similar national extensions, for instance SpainVoc or IrishVoc, by other member states. Another emerging approach to the multilinguality challenge is to create graph-based Linked Data native dictionaries that include lexical knowledge and overcome the disadvantages of tree-based dictionaries (Gracia et al. 2017). Others enrich the underlying ontology with linguistic information, for instance as proposed by the Ontolex-lemon model (McCrae et al. 2017; W3C Ontology-Lexica Community Group 2016). Finally, multilinguality could be further supported by adding linguistic and lexical information to enable NLP applications working with this information contained in an ontology.

Modeling standards In order to achieve the overarching ELI and ECLI goals EU member states should follow the modeling standards outlined in these proposals. Both ELI and ECLI describe a minimum set of non-country specific metadata and are therefore very well suited for national extensions where needed. Our comparison of the linked legal data features in the EU member states (cf. Table 4) shows that most of the participating countries follow the proposed modeling standards. Some countries, for instance Luxembourg provide their JOLUX ontology in their own as well as the ELI format. Individual deviations from these standards undermine the fundamental ideas of easier access to legal information across borders. One of the drawbacks of the current modeling standard, is the need to write queries in order to retrieve certain data as shown by Francesconi et al. (2015). The proposed solution, which involves decoupling the ELI and FRBR ontologies, needs to be approached and initiated in a centralized manner, for instance via a stakeholder engagement process whereby national experts who know their legal system and experts from the responsible EU institutions work together in order to shape future ELI and ECLI enhancements.

7 Related work

The exchange of legal information was already a concern before the advent of (legal) knowledge graphs and started with the standardization of (XML-based) formats that would allow the exchange of legal information across different jurisdictions. Furthermore, also ontologies to model legal information have been proposed. The goal of this section is to present other semantic technology based initiatives in the legal domain beyond work on legal knowledge graphs.

Several formats have been proposed enabling or simplifying the exchange of legal information in a structured and standardized manner. Boer et al. (2002) described the XML standard MetaLex which can be used to encode the structure and the content of legal documents. Another open and extensible XML standard for the exchange of legislative and judiciary documents is Akoma Ntoso[93] providing schemes for the structure and metadata of legal documents. Other standards for the XML-based exchange of legal information are for instance LegalDocML TC[94] based on Akoma-Ntoso aiming at the creation of a standard for a worldwide exchange of legal information using a standardized set of metadata. LegalRuleML (Palmirani et al. 2011; Athan et al. 2013) focuses on the expression of rules and constraints in the legal domain in XML format. The Legal Knowledge Interchange Format (LKIF) proposed by Hoekstra et al. (2007) is an ontology aiming at interchanging legal information between different legal systems modeling the semantics contained in the text of legal documents (Boer et al. 2008).

With respect to legal ontologies there has been research work in the past years mainly dealing with legal domain specific ontologies. A summary of existing legal ontologies has been published by Breuker et al. (2009) listing 23 ontologies and categorizing them by application (information retrieval, general language for expressing legal knowledge,...), type (knowledge representation) or character (general vs domain-specific). A recent extensive study conducted by de Oliveira Rodrigues et al. (2019) analyses legal ontologies found in various digital libraries based on multiple dimensions such as formalization, legal theories, semantic problems and ontology engineering problems in a systematic manner. The study shows that a large number of legal ontologies have been proposed over time and are available for reuse. Leone et al. (2019) classifies legal ontologies according general, modeling and semantic information. Ajani et al. (2016) proposed the European Legal Taxonomy Syllabus (ELTS) as a lightweight ontology that should help to relate national and European legal terminology to represent the differences in the national legal systems of the EU member states. A legal knowledge management system based on ELTS to semi-automatically classify and interlink documents has been proposed by Boella et al. (2019). Besides the generic legal ontologies used in this paper, many domain-specifc legal ontologies have been proposed tailored for the usage in a narrow legal domain. For instance the Open Digital Rights Language (ODRL)[95] (Steyskal and Polleres 2014; Vos et al. 2019), Linked Data

[93] http://www.akomantoso.org/.

[94] https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=legaldocml.

[95] https://www.w3.org/community/odrl/.

Rights (LDR)[96] and the Media Contract Ontology (MCO) (Rodríguez-Doncel et al. 2016) to model policies, LOTED2 (Distinto et al. 2016) and PPROC (Muñoz-Soro et al. 2016) for the procurement domain. Ontologies related to data protection are for instance GDPRtEXT (Pandit et al. 2018) which is an extension of the ELI ontology to model the GDPR, PrivOnto (Oltramari et al. 2018) and PrOnto (Palmirani et al. 2018) to model privacy policies and a similarly named ontology to represent product information called PRONTO (Vegetti et al. 2011). Ontologies are subject to improvement over time. In the legal domain, Francesconi et al. (2015) highlight drawbacks in the modeling of the CDM ontology used by the EU leading to unnecessarily complex queries and show how they could be resolved.

Lastly, ontology design patterns have been proposed to help with the creation of ontologies in a more systematic manner, for instnace based on patterns found in domain-specific documents. An overview of legal ontology design patterns is provided by Gangemi (2007). Examples for specific ontology design patterns in the legal domain are the Complaint Ontology Pattern (COP) by Santos et al. (2016) and the License Linked Data Resources Pattern proposed by Rodríguez-Doncel et al. (2013). Our middle-out ontology engineering method used to extend the existing ontologies described herein can likewise be used and applied alongside ontology design patterns.

8 Conclusion

In this paper, we describe the creation of a legal knowledge graph for Austria and propose the LKG ontology based on a real-world project funded by the Austrian Ministry for Digital and Economic Affairs. We provide detailed information about the modeling of the Austrian legal system using ELI and ECLI and propose different ontology population methods including rule-based and machine learning based approaches. Our comparative evaluation shows that rule-based as well as machine learning based approaches work similarly well for the extraction of legal entities. Furthermore, we enhance our Austrian LKG by linking to external spatial knowledge bases such as Geonames and Open Street Map, thus enabling more fine grained spatial search. We also performed an depth analysis into the existing linked legal data initiatives by the various EU member states, and extended the analysis by presenting the predominant non-governmental linked legal data initiatives that are based on ELI and ECLI. Finally we demonstrated how said initiatives can enhance search possibilities and eases access to legal information by providing example SPARQL queries over several linked legal knowledge sources. The findings show that although the existing initiatives have already started to bear fruit when it comes to making all legal information machine-accessible we have barely scratched the surface.

Future work includes the extension of the corpus used for the evaluation of the legal entities extraction approaches with a study whether these results could be further boosted, for instance by training a state of the art language model based on Austrian legal documents or hyperparameter optimization. Furthermore, analyzing the content of legal documents and including the outputs in our legal knowledge graph, e.g. the

[96] http://vocab.linkeddata.es/ontologies/purl.oclc.orgNETldrns.html.

automatic extraction of rules and constraints of legal provisions, or in analyzing the semantic content of court decisions to predict the outcome of future court decisions. Another possible route for further work involves an extensive linkage of our legal knowledge graph to external knowledge bases, for instance general knowledge bases, news sources, etc. Lastly, while we have shown that integrating the EuroVoc thesaurus supports search across multiple languages, it would be worth investigating the semantic meaning, differences, ambiguities, and similarities of legal expressions across different languages and jurisdictions.

Acknowledgements We thank the anonymous reviewers for their valuable suggestions and inputs. This work was partially funded by the Austrian Research Promotion Agency (FFG) under the "ICT of the Future" program (project "AI@Work", contract #874111), the Federal Ministry of Digital and Economic Affairs of the Republic of Austria and the Jubilee Fund of the City of Vienna. Sabrina Kirrane is funded by the FWF Austrian Science Fund and the Internet Foundation Austria under the FWF Elise Richter and netidee SCIENCE programmes as project number V 759-N. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 860801. Open access funding provided by Vienna University of Economics and Business (WU).

A Appendix

The appendix contains overview tables for the properties used in ELI and ECLI in different EU member states as well as links to the legal databases and example documents we used in this work.

A.1 ECLI properties used in different countries

Table 7 contains all properties from the ECLI ontology and shows which countries use which ECLI properties. Countries for which we use the non-governmental initiatives are highlight gray.

Table 7 Overview of used ECLI properties of countries providing metadata using ECLI

ECLI Property *Data based on*	Austria *LKG*	Finland *Finlex SPARQL Endpoint*	Germany *RDFa*	Netherlands *RDFa*
`dcterms:abstract`		✓		✓
`dcterms:accessRights`	✓		✓	
`dcterms:contributor`	✓	✓		
`dcterms:coverage`	✓		✓	
`dcterms:creator`	✓	✓	✓	✓
`dcterms:date`	✓	✓	✓	
`dcterms:description`		✓		
`dcterms:identifier`	✓		✓	✓
`dcterms:isReplacedBy`				
`dcterms:issued`		✓		✓
`dcterms:isVersionOf`	✓	✓	✓	
`dcterms:language`	✓	✓	✓	✓
`dcterms:publisher`	✓	✓	✓	✓
`dcterms:references`	✓			
`dcterms:subject`	✓			✓
`dcterms:title`				✓
`dcterms:type`	✓		✓	✓

A.2 Overview of used named authority lists

Table 8 shows for which properties NAL are used by different countries. We only list countries which provide metadata in RDF or RDFa format and are using NAL for legislative and judiciary documents. Furthermore, not all countries do provide a dedicated NAL information page although they use NAL and they cannot be retrieved from the internet.

Table 8 Overview of the used NAL in different countries for legislative and judiciary documents

NAL for property *Data based on*	Austria *LKG*	Denmark *RDF*	Finland *Finlex SPARQL Endpoint*	France *RDFa*	Italy *RDF*	Luxembourg *RDF*	Netherlands *RDFa*	Portugal *RDFa*	Spain *RDFa*
`dcerms:type`	✔	–	–	–	–	–	✓	–	–
`dcterms:subject`	✔	–	–	–	–	–	✓	–	–
`eli:is_about`	✔	–	–	–	–	✓	–	–	–
`eli:jurisdiction`	✔	–	–	–	–	–	–	–	✓
`eli:language`	✔	–	–	✓	✓	✓	–	–	✓
`eli:passed_by`	–	✓	✔	–	–	–	–	–	–
`eli:publisher_agent`	–	–	–	–	–	✓	–	✓	–
`eli:relevant_for`	✔	✓	–	–	–	–	–	–	–
`eli:responsibility_of_agent`	–	–	–	–	–	✓	–	✓	–
`eli:rightsholder_agent`	–	–	–	–	–	✓	–	✓	–
`eli:type_document`	✔	✓	–	–	✓	✓	–	✓	✓
`eli:version`	–	–	–	–	✓	–	–	–	✓

Countries for which we use the non-governmental initiatives are highlight bold

A.3 ELI properties used in different countries

Table 9 contains all properties from the ELI ontology and shows which countries use which ELI properties. Countries for which we use the non-governmental initiatives are highlight gray.

Table 9 Overview of used ELI properties of countries providing metadata using ELI including non-governmental initiatives

ELI Property *Data based on*	Austria *LKG*	Denmark *RDF*	Finland *Finlex SPARQL Endpoint*	France *RDFa*	Greece *Nomothesia SPARQL Endpoint*	Ireland *RDF*	Italy *RDFa, RDF*	Luxemburg *RDF*	Portugal *RDFa*	Spain *RDFa*
eli:amended_by	✓		✓							
eli:amends	✓		✓		✓			✓		
eli:applied_by										
eli:applies										
eli:based_on		✓				✓		✓		
eli:basis_for		✓								
eli:changed_by	✓	✓	✓							
eli:changes	✓	✓	✓		✓	✓				
eli:cited_by			✓							
eli:cited_by_case_law	✓									
eli:cited_by_case_law_reference										
eli:cites			✓	✓	✓			✓	✓	
eli:commenced_by										
eli:commences										
eli:consolidated_by		✓								✓
eli:consolidates	✓	✓						✓		
eli:corrected_by										✓
eli:corrects								✓		
eli:date_applicability								✓		
eli:date_document	✓	✓		✓		✓	✓	✓		✓
eli:date_no_longer_in_force	✓	✓	✓					✓		
eli:date_publication	✓		✓	✓	✓		✓	✓	✓	✓

Table 9 continued

ELI Property *Data based on*	Austria *LKG*	Denmark *RDF*	Finland *Finlex SPARQL Endpoint*	France *RDFa*	Greece *Nomothesia SPARQL Endpoint*	Ireland *RDF*	Italy *RDFa, RDF*	Luxemburg *RDF*	Portugal *RDFa*	Spain *RDFa*
`eli:description`						✓		✓	✓	
`eli:embodies`	✓	✓	✓	✓		✓		✓	✓	✓
`eli:first_date_entry_in_force`	✓		✓		✓			✓		
`eli:format`	✓	✓	✓	✓		✓	✓	✓		✓
`eli:has_another_publication`										
`eli:has_member`	✓		✓							✓
`eli:has_part`			✓		✓	✓				
`eli:has_translation`										
`eli:id_local`	✓	✓	✓	✓	✓		✓	✓	✓	✓
`eli:implemented_by`										
`eli:implements`										
`eli:in_force`	✓	✓	✓						✓	
`eli:is_about`	✓		✓						✓	
`eli:is_another_publication_of`										
`eli:is_embodied_by`	✓	✓	✓	✓		✓	✓	✓		✓
`eli:is_exemplified_by`										
`eli:is_member_of`	✓		✓					✓		✓
`eli:is_part_of`			✓	✓				✓		

Table 9 continued

ELI Property *Data based on*	Austria *LKG*	Denmark *RDF*	Finland *Finlex SPARQL Endpoint*	France *RDFa*	Greece *Nomothesia SPARQL Endpoint*	Ireland *RDF*	Italy *RDFa, RDF*	Luxemburg *RDF*	Portugal *RDFa*	Spain *RDFa*
eli:is_realized_by	✓	✓	✓			✓	✓		✓	✓
eli:is_translation_of										
eli:jurisdiction	✓									✓
eli:language	✓	✓	✓	✓		✓	✓	✓	✓	✓
eli:legal_value		✓		✓		✓			✓	
eli:licence						✓		✓	✓	
eli:media_type										
eli:number	✓					✓			✓	✓
eli:passed_by		✓	✓	✓			✓			
eli:published_in		✓		✓	✓	✓				
eli:published_in_format						✓		✓	✓	
eli:publisher				✓		✓	✓	✓	✓	✓
eli:publisher_agent							✓		✓	
eli:publishes										
eli:realized_by	✓				✓					
eli:realizes	✓	✓	✓	✓		✓	✓	✓	✓	✓
eli:related_to			✓			✓				
eli:relevant_for	✓	✓				✓				
eli:repealed_by	✓		✓							
eli:repeals	✓		✓					✓		

Table 9 continued

ELI Property *Data based on*	Austria *LKG*	Denmark *RDF*	Finland *Finlex SPARQL Endpoint*	France *RDFa*	Greece *Nomothesia SPARQL Endpoint*	Ireland *RDF*	Italy *RDFa, RDF*	Luxemburg *RDF*	Portugal *RDFa*	Spain *RDFa*
`eli:responsibility_of`				✓				✓	✓	
`eli:responsibility_of_agent`								✓	✓	
`eli:rights`								✓		
`eli:rightsholder`						✓				
`eli:rightsholder_agent`								✓	✓	
`eli:title`	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
`eli:title_alternative`	✓	✓	✓					✓		
`eli:title_short`	✓	✓								
`eli:transposed_by`	✓				✓					
`eli:transposes`	✓	✓	✓			✓	✓	✓		
`eli:type_document`	✓	✓	✓	✓		✓	✓	✓	✓	✓
`eli:uri_schema`		✓				✓	✓		✓	
`eli:version`	✓		✓			✓	✓			✓
`eli:version_date`			✓							✓

A.4 Overview of legal databases and example documents

Tables 10 (Legislation) and 11 (Jurisdiction) provide an overview over the legal databases and example documents for all EU member states we used for our analysis in Sect. 5.

Table 10 Overview of legal databases and example documents for legislation (URLs of example documents are shortened)

Country	Legislation	Example Document
Austria	https://ris.bka.gv.at/	https://bit.ly/2UEq4E9
Belgium	http://www.ejustice.just.fgov.be/	https://bit.ly/30zHeGx
Bulgaria	https://dv.parliament.bg/	https://bit.ly/2MQ7q83
Croatia	http://nn.hr/	https://bit.ly/3hnXy34
Cyprus	http://www.cylaw.org/	https://bit.ly/3hew1Bc
Czech Republic	https://aplikace.mvcr.cz/sbirka-zakonu/	https://bit.ly/2XVLajg
Denmark	https://www.retsinformation.dk/	https://bit.ly/2YwzBhs
Estonia	https://www.riigiteataja.ee/	https://bit.ly/2XUxLIf
Finland	https://www.finlex.fi/	https://bit.ly/2UEbRXA
France	https://www.legifrance.gouv.fr/	https://bit.ly/2XUy4Tp
Germany	http://www.bgbl.de/	https://bit.ly/3cV7dLh
Greece	http://www.et.gr/	https://bit.ly/2B4bApT
Hungary	http://njt.hu/	https://bit.ly/3d2iQQN
Ireland	http://www.irishstatutebook.ie/	https://bit.ly/2XUhVxd
Italy	https://www.normattiva.it/	https://bit.ly/30zls5Z
Latvia	http://www.likumi.lv/	https://bit.ly/2UDUJBl
Lithuania	https://www.e-tar.lt/	https://bit.ly/2XVquIj
Luxembourg	http://legilux.public.lu/	https://bit.ly/30ycd5Q
Malta	https://legislation.mt/	https://bit.ly/2XSrgpq
Netherlands	https://www.officielebekendmakingen.nl/	https://bit.ly/2Ooq5IV
Poland	http://isip.sejm.gov.pl/	https://bit.ly/3hkOc8i
Portugal	https://dre.pt/	https://bit.ly/3gOtNrn
Romania	http://legislatie.just.ro/	https://bit.ly/37lNJhA
Slovakia	https://www.slov-lex.sk/	https://bit.ly/2XUz4a7
Slovenia	http://www.pisrs.si/Pis.web/	https://bit.ly/3cVWu2Y
Spain	https://boe.es/	https://bit.ly/2AjLPCk
Sweden	http://rkrattsbaser.gov.se/	https://bit.ly/3d2jQV3

Table 11 Overview of legal databases and example documents for jurisdiction (URLs of example documents are shortened)

Country	Judiciary	Example Document
Austria	https://ris.bka.gv.at/	https://bit.ly/37maTo6
Belgium	http://jure.juridat.just.fgov.be/	https://bit.ly/2AjPLTB
Bulgaria	https://legalacts.justice.bg/	https://bit.ly/3hpfdI2
Croatia	https://sudskapraksa.vsrh.hr/home	https://bit.ly/3fiabv0
Cyprus	http://www.cylaw.org/	https://bit.ly/30BXxD0
Czech Republic	http://www.nsoud.cz/	https://bit.ly/2Ywztyu
Denmark	https://domstol.dk/	https://bit.ly/2MQrqan
Estonia	https://www.riigiteataja.ee/	https://bit.ly/2UFChIK
Finland	https://www.finlex.fi/	https://bit.ly/3cS3w93
France	https://www.courdecassation.fr/	https://bit.ly/30CQjyq
Germany	http://www.bundesverfassungsgericht.de/	https://bit.ly/2BXbHUN
Greece	http://www.adjustice.gr/	https://bit.ly/3feYwwT
Hungary	https://birosag.hu/birosagi-hatarozatok-gyujtemenye/	Direct download
Ireland	https://beta.courts.ie/	https://bit.ly/2YpEwRm
Italy	http://www.italgiure.giustizia.it/	Registration required
Latvia	https://manas.tiesas.lv/eTiesas/	https://bit.ly/30Bettm
Lithuania	https://www.lat.lt/	https://bit.ly/30BHfdc
Luxembourg	https://justice.public.lu/	https://bit.ly/3fnuJ5n
Malta	https://justice.gov.mt/	https://bit.ly/2XVMmmK
Netherlands	https://data.rechtspraak.nl/	https://bit.ly/3hlxaqN
Poland	http://orzeczenia.nsa.gov.pl/	https://bit.ly/2zuAq22
Romania	http://www.rolii.ro/	https://bit.ly/2YtigWL
Slovakia	https://obcan.justice.sk/	https://bit.ly/2MO0BDX
Slovenia	http://www.sodnapraksa.si/	https://bit.ly/2XRJEia
Spain	http://www.poderjudicial.es/	https://bit.ly/3fm2ALX
Sweden	https://rattsinfosok.domstol.se/	https://bit.ly/3fckCjq

References

Ajani G, Lesmo L, Boella G, Mazzei A, Rossi P (2007) Terminological and ontological analysis of European directives: multilinguism in law. In: Proceedings of the 11th international conference on artificial intelligence and law, association for computing machinery, New York, NY, USA, ICAIL'07, pp 43–48. https://doi.org/10.1145/1276318.1276327

Ajani G, Boella G, Caro LD, Robaldo L, Humphreys L, Praduroux S, Rossi P, Violato A (2016) The European taxonomy syllabus: a multi-lingual, multi-level ontology framework to untangle the web of European legal terminology. Appl Ontol 11(4):325–375. https://doi.org/10.3233/AO-170174

Akbik A, Blythe D, Vollgraf R (2018) Contextual string embeddings for sequence labeling. In: COLING 2018, 27th international conference on computational linguistics, pp 1638–1649

Athan T, Boley H, Governatori G, Palmirani M, Paschke A, Wyner A (2013) Oasis legalruleml. In: Proceedings of the fourteenth international conference on artificial intelligence and law, association for computing machinery, New York, NY, USA, ICAIL'13, pp 3–12. https://doi.org/10.1145/2514601.2514603

Bacci L, Agnoloni T, Marchetti C, Battistoni R (2019) Improving public access to legislation through legal citations detection: the linkoln project at the Italian senate. Knowl Law Big Data Age 317:149

Beno M, Filtz E, Kirrane S, Polleres A (2019) Doc2RDFa: semantic annotation for web documents. In: Alam M, Usbeck R, Pellegrini T, Sack H, Sure-Vetter Y (eds.), Proceedings of the posters and demo track of the 15th international conference on semantic systems (SEMANTiCS 2019), CEUR-WS.org, Karlsruhe, Germany, CEUR Workshop Proceedings, vol 2451. http://ceur-ws.org/Vol-2451/paper-06.pdf

Berners-Lee T (2006) Linked data design issues. https://www.w3.org/DesignIssues/LinkedData.html. Accessed 15 Mar 2020

Boella G, Caro LD, Graziadei M, Cupi L, Salaroglio CE, Humphreys L, Konstantinov H, Marko K, Robaldo L, Ruffini C, Simov KI, Violato A, Stroetmann VN (2015) Linking legal open data: breaking the accessibility and language barrier in European legislation and case law. In: Sichelman T, Atkinson K (eds.), Proceedings of the 15th international conference on artificial intelligence and law, ICAIL 2015, San Diego, CA, USA, ACM, pp 171–175. https://doi.org/10.1145/2746090.2746106

Boella G, Caro LD, Leone V (2019) Semi-automatic knowledge population in a legal document management system. Artif Intell Law 27(2):227–251. https://doi.org/10.1007/s10506-018-9239-8

Boer A, Hoekstra R, Winkels R, Van Engers T, Willaert F (2002) Metalex: legislation in xml. Legal Knowledge and Information Systems (Jurix 2002) pp 1–10

Boer A, Winkels R, Vitali F (2008) Metalex XML and the legal knowledge interchange format. In: Casanovas P, Sartor G, Casellas N, Rubino R (eds) Computable models of the law, languages, dialogues, games, ontologies, vol 4884. Lecture Notes in Computer Science. Springer, Berlin, pp 21–41. https://doi.org/10.1007/978-3-540-85569-9_2

Breuker J, Casanovas P, Klein MCA, Francesconi E (2009) The flood, the channels and the dykes: Managing legal information in a globalized and digital world. In: Breuker J, Casanovas P, Klein MCA, Francesconi E (eds.), Law, ontologies and the semantic web—channelling the legal information flood, IOS Press, Frontiers in Artificial Intelligence and Applications, vol 188, pp 3–18. https://doi.org/10.3233/978-1-58603-942-4-3

Cardellino C, Teruel M, Alemany LA, Villata S (2017) A low-cost, high-coverage legal named entity recognizer, classifier and linker. In: Keppens J, Governatori G (eds.), Proceedings of the 16th edition of the international conference on artificial intelligence and law, ICAIL 2017, London, United Kingdom, ACM, pp 9–18, https://doi.org/10.1145/3086512.3086514

Casanovas P, Palmirani M, Peroni S, van Engers TM, Vitali F (2016) Semantic web for the legal domain: the next step. Semant Web 7(3):213–227. https://doi.org/10.3233/SW-160224

Chalkidis I, Nikolaou C, Soursos P, Koubarakis M (2017) Modeling and querying greek legislation using semantic web technologies. In: Blomqvist E, Maynard D, Gangemi A, Hoekstra R, Hitzler P, Hartig O (eds.), The Semantic Web—14th international conference, ESWC 2017, Portorož, Slovenia, Proceedings, Part I, Lecture Notes in Computer Science, vol 10249, pp 591–606. https://doi.org/10.1007/978-3-319-58068-5_36

Council of the European Union (2011) Council conclusions inviting the introduction of the European Case Law Identifier (ECLI) and a minimum set of uniform metadata for case law. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52011XG0429(01). Accessed 17 May 2020

Council of the European Union (2012) Council conclusions inviting the introduction of the European Legislation Identifier (ELI). https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52012XG1026(01). Accessed 17 May 2020

Council of the European Union (2017) Council conclusions of 6 November 2017 on the European Legislation Identifier. https://eur-lex.europa.eu/legal-content/GA/TXT/?uri=CELEX:52017XG1222(02). Accessed 17 May 2020

Cunningham H, Cunningham H, Maynard D, Maynard D, Tablan V, Tablan V (1999) JAPE: a java annotation patterns engine

Devlin J, Chang M, Lee K, Toutanova K (2019) BERT: pre-training of deep bidirectional transformers for language understanding. In: Burstein J, Doran C, Solorio T (eds.), Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, NAACL-HLT 2019, Minneapolis, MN, USA, vol 1 (Long and Short Papers), Association for Computational Linguistics, pp 4171–4186. https://doi.org/10.18653/v1/n19-1423

de Oliveira Rodrigues CM, de Freitas FLG, Barreiros EFS, de Azevedo RR, de Almeida Filho A (2019) Legal ontologies over time: a systematic mapping study. Exp Syst Appl 130:12–30

Distinto I, d'Aquin M, Motta E (2016) LOTED2: an ontology of European public procurement notices. Semant Web 7(3):267–293. https://doi.org/10.3233/SW-140151

Dozier C, Kondadadi R, Light M, Vachher A, Veeramachaneni S, Wudali R (2010) Named entity recognition and resolution in legal text. In: Francesconi E, Montemagni S, Peters W, Tiscornia D (eds) Semantic processing of legal texts: where the language of law meets the law of language, vol 6036. Lecture Notes in Computer Science. Springer, Berlin, pp 27–43. https://doi.org/10.1007/978-3-642-12837-0_2

Filtz E, Kirrane S, Polleres A (2018) Interlinking legal data. In: Khalili A, Koutraki M (eds.), Proceedings of the posters and demos track of the 14th international conference on semantic systems co-located with the 14th international conference on semantic systems (SEMANTiCS 2018), Vienna, Austria, CEUR-WS.org, CEUR Workshop Proceedings, vol 2198. http://ceur-ws.org/Vol-2198/paper_118.pdf

Francart T, Dann J, Pappalardo R, Malagon C, Pellegrino M (2018) The European legislation identifier. In: Peruginelli G, Faro S (eds.), Knowledge of the law in the big data age, conference 'Law via the Internet 2018, Florence, Italy, IOS Press, Frontiers in Artificial Intelligence and Applications, vol 317, pp 137–148. https://doi.org/10.3233/FAIA190016

Francesconi E, Küster MW, Gratz P, Thelen S (2015) The ontology-based approach of the publications office of the EU for document accessibility and open data services. In: Ko A, Francesconi E (eds.), Electronic government and the information systems perspective—4th international conference, EGOVIS 2015, Valencia, Spain, Proceedings, Springer, Lecture Notes in Computer Science, vol 9265, pp 29–39. https://doi.org/10.1007/978-3-319-22389-6_3

Gangemi A (2007) Design patterns for legal ontology constructions. In: Casanovas P, Biasiotti MA, Francesconi E, Sagri M (eds.), Proceedings of the 2nd workshop on legal ontologies and artificial intelligence techniques, Stanford University, Stanford, CA, USA, CEUR-WS.org, CEUR Workshop Proceedings, vol 321, pp 65–85. http://ceur-ws.org/Vol-321/paper4.pdf

Ghosh ME, Naja H, Abdulrab H, Khalil M (2016) Towards a middle-out approach for building legal domain reference ontology. Int J Knowl Eng 2(3):109–114

Gracia J, Kernerman I, Bosque-Gil J (2017) Toward linked data-native dictionaries. In: Proceedings of the eLex 2017 conference on electronic lexicography in the 21st Century: lexicography from scratch, pp 19–21

Grishman R, Sundheim B (1996) Message understanding conference- 6: a brief history. In: Proceedings of the Conference on 16th international conference on computational linguistics, COLING 1996, Center for Sprogteknologi, Copenhagen, Denmark, pp 466–471. https://www.aclweb.org/anthology/C96-1079/

Hoekstra R, Breuker J, Bello MD, Boer A (2007) The LKIF core ontology of basic legal concepts. In: Casanovas P, Biasiotti MA, Francesconi E, Sagri M (eds.), Proceedings of the 2nd Workshop on legal ontologies and artificial intelligence techniques, Stanford University, Stanford, CA, USA, CEUR-WS.org, CEUR Workshop Proceedings, vol 321, pp 43–63. http://ceur-ws.org/Vol-321/paper3.pdf

Hogan A, Blomqvist E, Cochez M, d'Amato C, de Melo G, Gutierrez C, Gayo JEL, Kirrane S, Neumaier S, Polleres A, Navigli R, Ngomo ACN, Rashid SM, Rula A, Schmelzeisen L, Sequeda J, Staab S, Zimmermann A (2020) Knowledge graphs

Kucera J, Chlapek D, Klímek J, Necaský M (2015) Methodologies and best practices for open data publication. In: Necaský M, Pokorný J, Moravec P (eds.), Proceedings of the Dateso 2015 annual international workshop on DAtabases, TExts, Specifications and Objects, Neprivec u Sobotky, Jicin, Czech Republic, April 14, 2015, CEUR-WS.org, CEUR Workshop Proceedings, vol 1343, pp 52–64. http://ceur-ws.org/Vol-1343/paper5.pdf

Lafferty JD, McCallum A, Pereira FCN (2001) Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: Brodley CE, Danyluk AP (eds.), Proceedings of the eighteenth international conference on machine learning (ICML 2001), Williams College, Williamstown, MA, USA, Morgan Kaufmann, pp 282–289

Leitner E, Rehm G, Schneider JM (2019) Fine-grained named entity recognition in legal documents. In: Acosta M, Cudré-Mauroux P, Maleshkova M, Pellegrini T, Sack H, Sure-Vetter Y (eds.), Semantic Systems. The Power of AI and Knowledge Graphs—15th International Conference, SEMANTiCS 2019, Karlsruhe, Germany, Proceedings, Springer, Lecture Notes in Computer Science, vol 11702, pp 272–287. https://doi.org/10.1007/978-3-030-33220-4_20

Leone V, Di Caro L, Villata S (2019) Taking stock of legal ontologies: a feature-based comparative analysis. Artif Intell Law 28:207–235

Manning CD, Raghavan P, Schütze H (2008) Introduction to information retrieval. Cambridge University Press, Cambridge

McCrae JP, Bosque-Gil J, Gracia J, Buitelaar P, Cimiano P (2017) The ontolex-lemon model: development and applications. In: Proceedings of eLex 2017 conference, 2017, pp 19–21

Montiel-Ponsoda E, Rodríguez-Doncel V, Gracia J (2017) Building the legal knowledge graph for smart compliance services in multilingual europe. In: Rodríguez-Doncel V, Casanovas P, González-Conejero J (eds.), Proceedings of the 1st workshop on technologies for regulatory compliance co-located with the 30th international conference on legal knowledge and information systems (JURIX 2017), Luxembourg, CEUR-WS.org, CEUR workshop proceedings, vol 2049, pp 15–17. http://ceur-ws.org/Vol-2049/02paper.pdf

Muñoz S, Pérez J, Gutiérrez C (2009) Simple and efficient minimal RDFS. J Web Semant 7(3):220–234. https://doi.org/10.1016/j.websem.2009.07.003

Muñoz-Soro JF, Esteban G, Corcho Ó, Serón FJ (2016) Pproc, an ontology for transparency in public procurement. Semant Web 7(3):295–309. https://doi.org/10.3233/SW-150195

Neumaier S, Polleres A (2019) Enabling spatio-temporal search in open data. J Web Semant 55:21–36. https://doi.org/10.1016/j.websem.2018.12.007

Oksanen A, Tamper M, Tuominen J, Mäkelä E, Hietanen A, Hyvönen E (2019) Semantic finlex: transforming, publishing, and using finnish legislation and case law as linked open data on the web. Knowl Law Big Data Age 317:212–228

Oltramari A, Piraviperumal D, Schaub F, Wilson S, Cherivirala S, Norton TB, Russell NC, Story P, Reidenberg JR, Sadeh NM (2018) Privonto: a semantic framework for the analysis of privacy policies. Semant Web 9(2):185–203. https://doi.org/10.3233/SW-170283

van Opijnen M, Palmirani M, Vitali F, van den Oever J, Agnoloni T (2017a) Towards ECLI 2.0. In: Parycek P, Edelmann N (eds.), 2017 conference for e-democracy and open government, CeDEM 2017, Krems, Austria, IEEE Computer Society, pp 135–143. https://doi.org/10.1109/CeDEM.2017.17

van Opijnen M, Peruginelli G, Kefali E, Palmirani M (2017b) On-line Publication of Court Decisions in the EU. Report of the policy group of the project 'building on the European case law identifier, BO-ECLI. http://bo-ecli.eu/uploads/deliverables/Deliverable

Palmirani M, Governatori G, Rotolo A, Tabet S, Boley H, Paschke A (2011) Legalruleml: Xml-based rules and norms. In: Olken F, Palmirani M, Sottara D (eds.), Rule-based modeling and computing on the semantic web, 5th international symposium, RuleML 2011-America, Ft. Lauderdale, FL, Florida, USA. Proceedings, Springer, Lecture Notes in Computer Science, vol 7018, pp 298–312. https://doi.org/10.1007/978-3-642-24908-2_30

Palmirani M, Martoni M, Rossi A, Bartolini C, Robaldo L (2018) Pronto: Privacy ontology for legal reasoning. In: Ko A, Francesconi E (eds.), Electronic government and the information systems perspective—7th international conference, EGOVIS 2018, Regensburg, Germany, Proceedings, Springer, Lecture Notes in Computer Science, vol 11032, pp 139–152. https://doi.org/10.1007/978-3-319-98349-3_11

Pandit HJ, Fatema K, O'Sullivan D, Lewis D (2018) Gdprtext - GDPR as a linked data resource. In: Gangemi A, Navigli R, Vidal M, Hitzler P, Troncy R, Hollink L, Tordai A, Alam M (eds.), The semantic web—15th international conference, ESWC 2018, Heraklion, Crete, Greece, Proceedings, Springer, Lecture Notes in Computer Science, vol 10843, pp 481–495. https://doi.org/10.1007/978-3-319-93417-4_31

Presutti V, Gangemi A (2008) Content ontology design patterns as practical building blocks for web ontologies. In: Li Q, Spaccapietra S, Yu ESK, Olivé A (eds.), Conceptual Modeling—ER 2008, 27th international conference on conceptual modeling, Barcelona, Spain. Proceedings, Springer, Lecture Notes in Computer Science, vol 5231, pp 128–141. https://doi.org/10.1007/978-3-540-87877-3_11

Publications Office of the European Union (2020a) Common Data Model (CDM). https://op.europa.eu/en/web/eu-vocabularies/cdm/. Accessed 27 Nov 2020

Publications Office of the European Union (2020b) European Legislation Identifier (ELI). https://op.europa.eu/en/web/eu-vocabularies/eli/. Accessed 27 Nov 2020

Rodríguez-Doncel V, Suárez-Figueroa MC, Gómez-Pérez A, Poveda-Villalón M (2013) License linked data resources pattern. In: Gangemi A, Gruninger M, Hammar K, Lefort L, Presutti V, Scherp A (eds.), Proceedings of the 4th workshop on ontology and semantic web patterns co-located with 12th international semantic web conference (ISWC 2013), Sydney, Australia, CEUR-WS.org, CEUR Workshop Proceedings, vol 1188. http://ceur-ws.org/Vol-1188/paper_7.pdf

Rodríguez-Doncel V, Delgado J, Llorente S, Rodríguez E, Boch L (2016) Overview of the MPEG-21 media contract ontology. Semant Web 7(3):311–332. https://doi.org/10.3233/SW-160215

Sanh V, Debut L, Chaumond J, Wolf T (2019) Distilbert, a distilled version of BERT: smaller, faster, cheaper and lighter. http://arxiv.org/abs/1910.01108

Santos C, Pruski C, Silveira MD, Rodríguez-Doncel V, Gangemi A, van der Torre L, Casanovas P (2016) Complaint ontology pattern—COP. In: Hammar K, Hitzler P, Krisnadhi A, Lawrynowicz A, Nuzzolese AG, Solanki M (eds.), Advances in ontology design and patterns [revised and extended versions of the papers presented at the 7th edition of the Workshop on Ontology and Semantic Web Patterns, WOP@ISWC 2016, Kobe, Japan, 18th October 2016], IOS Press, Studies on the Semantic Web, vol 32, pp 69–83. https://doi.org/10.3233/978-1-61499-826-6-69

Steyskal S, Polleres A (2014) Defining expressive access policies for linked data using the ODRL ontology 2.0. In: Sack H, Filipowska A, Lehmann J, Hellmann S (eds.), Proceedings of the 10th international conference on semantic systems, SEMANTICS 2014, Leipzig, Germany, ACM, pp 20–23. https://doi.org/10.1145/2660517.2660530

Uschold M, Gruninger M (1996) Ontologies: principles, methods and applications. Knowl Eng Rev 11(2):93–136. https://doi.org/10.1017/S0269888900007797

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. In: Guyon I, von Luxburg U, Bengio S, Wallach HM, Fergus R, Vishwanathan SVN, Garnett R (eds.), Advances in Neural Information Processing Systems 30: Annual conference on neural information processing systems 2017, 4-9 December 2017, Long Beach, CA, USA, pp 5998–6008. http://papers.nips.cc/paper/7181-attention-is-all-you-need

Vegetti M, Leone HP, Henning GP (2011) PRONTO: an ontology for comprehensive and consistent representation of product information. Eng Appl Artif Intell 24(8):1305–1327. https://doi.org/10.1016/j.engappai.2011.02.014

Vos MD, Kirrane S, Padget JA, Satoh K (2019) ODRL policy modelling and compliance checking. In: Fodor P, Montali M, Calvanese D, Roman D (eds.), Rules and Reasoning—Third international joint conference, RuleML+RR 2019, Bolzano, Italy, September 16-19, 2019, Proceedings, Springer, Lecture Notes in Computer Science, vol 11784, pp 36–51. https://doi.org/10.1007/978-3-030-31095-0_3

W3C JSON for Linking Data Community Group (2012) JavaScript Object Notation for Linking Data (JSON-LD). https://www.w3.org/community/json-ld/. Accessed 17 May 2020

W3C Ontology-Lexica Community Group (2016) Lexicon Model for Ontologies: Final Community Group Report, 10 May 2016. https://www.w3.org/2016/05/ontolex/. Accessed 27 Nov 2020

W3C Recommendation (2012) A direct mapping of relational data to RDF. https://www.w3.org/TR/2012/REC-rdb-direct-mapping-20120927/. Accessed 17 May 2020

Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, Cistac P, Rault T, Louf R, Funtowicz M, Brew J (2019) Huggingface's transformers: state-of-the-art natural language processing. arXiv:1910.03771