Law and Justice as a Linked Open Data Service

Arttu Oksanen[1], Jouni Tuominen[1], Eetu Mäkelä[1],
Minna Tamper[1], Aki Hietanen[2], and Eero Hyvönen[1]

[1] Semantic Computing Research Group (SeCo), Aalto University, Finland
HELDIG – Helsinki Centre for Digital Humanities, University of Helsinki
http://seco.cs.aalto.fi/, firstname.lastname@aalto.fi
[2] Ministry of Justice, Finland
firstname.lastname@om.fi

Abstract. Everybody is expected to know and obey the law in today's society. Governments therefore publish legislation and case law widely in print and on the web. Such legal information is provided for human consumption, but the information is usually not available *as data* for algorithmic analysis and applications to use. However, this would be beneficial in many use cases, such as building more intelligent juridical online services and conducting research into legislation and legal practice. To address these needs, this paper presents Semantic Finlex, a national in-use data resource and system for publishing Finnish legislation and related case law as a Linked Open Data service with applications. The system transforms and interlinks on a regular basis data from the legacy legal database Finlex of the Ministry of Justice into Linked Open Data, based on the new European standards ECLI and ELI. The data is hosted on a "7-star" SPARQL endpoint with a variety of related services available that ease data re-use. Rich Internet Applications using only SPARQL for data access are presented as first application demonstrators of the data service.

Keywords: legislation, case law, Linked Data publishing, data modeling, data linking

1 Introduction

Governments provide publicly available legal information on the web usually in the form of HTML or PDF documents targeted to human readers. In Finland, for example, legislation and case law are published as HTML documents in the Finlex Data Bank[3], a publicly available online service since 1997, maintained by the Ministry of Justice [8]. However, Finlex does not provide publicly available machine-readable legal information as open data, on top of which services and analyses can be built by the ministry or third party vendors.

This paper presents Semantic Finlex[4]—a national Linked Open Data Service for Finnish legislation and case law. The service hosts and publishes a central part

[3] http://www.finlex.fi
[4] http://data.finlex.fi

of the Finnish legislation along with judgments of the Supreme Court and the Supreme Administrative Court. All of the datasets are automatically updated regularly.

Our work on Semantic Finlex started in 2012, and the first version of the service was published in 2014 [7]. The data included 2413 consolidated laws, 11 904 judgments of the Supreme Court, and 1490 judgments of the Supreme Administrative Court. In addition, some 30 000 terms used in 26 different thesauri were harvested for a first draft of a consolidated vocabulary. During this work, some shortcomings of the initial RDF data model became evident as well as the need for using the then emerging new standards for EU level interoperability. The demo dataset also consisted of only one temporal version (2012) of the statutory law and was not updated. These issues have now been resolved in the work reported in this paper.

In the following, we first explicate the motivation and use cases for publishing law and justice as linked open data. Then the underlying data models and the data conversion process applied in the service are presented, followed by a discussion on enriching the data with semantic and structural annotations. We also introduce the Semantic Finlex publishing platform and semantic portal. Finally, the service is evaluated in practice by presenting data analysis and application demonstrators built on top of the system.

2 Motivating Use Cases

Many actors would benefit from access to legislative and judicial content as data:

Information portals. Within the online services provided by different sectors, it is often necessary to refer to various sections of acts and decrees and display these to users. This requires that such sections be referable and readable as online data. For example, various regulations referring to law are published in the fields of construction, defense, and chemical safety.

The media. Since news on fields such as politics and the business world often refer to various sections of statutes, it is sometimes useful to guide readers to the original legal texts. However, this is not possible if the sections in question are not referable or available in data format.

Juridical online services. In Finland, these include services such as Suomen Laki (Finnish Law)[5] by Talentum Oyj and Edilex[6] by Edita Publishing Ltd, which primarily provide juridical information for professionals in law, such as judges and legal counsels, as well as private persons. Maintaining data in current systems is tedious and largely based on manual work, because the data is not available in a form ”understood” by computers, but only as documents in PDF, Word and other formats.

Legislative drafting. When new statutes are drafted in order to complement and supersede previous ones, the drafters have to examine previous

[5] `http://www.suomenlaki.com`
[6] `http://www.edilex.fi`

statutes in order to evaluate the effects of the changes and avoid discrepancies. However, semantic information on the various versions of and interdependencies between statutes has been available only in text format.

Editing and publishing of legislative datasets. Today, legislation-related information is produced in an inconsistent manner, by using various text formats and index term vocabularies to describe information content. If documents were drafted at the production stage in the form of structured data and in accordance with mutually agreed standards, this would facilitate their further processing and linking to other documents, such as materials in Parliament and in publishing systems such as Finlex.

Intelligent services. Legislative information related to problematic juridical situations, such as divorce or estate distribution, is often scattered between various acts, decrees, and legal practice cases. The availability of statutes and legal cases as such is of little help if the reader, such as an ordinary citizen, finds it impossible to piece the issue together. Presenting legislative documents in a form that can be interpreted by a computer, i.e., as semantic data, would enable the development of more intelligent applications, which would in turn enable making law and justice more comprehensible to citizens. For example, legal texts can be automatically linked to other related texts, legal cases, and vocabularies explaining legal terminology.

Research into legislation and legal practice. The enactment of legislation and legal practice are fields of research in which data analysis methods can be used. The topic of such a research might, for instance, be the impact of EU law to national legal practice [13]. However, data analysis methods require that statutes, the connections between them, and case-law-based information on their implementation are available in the form of systematically presented data.

Moreover, authorities in Europe strive to improve the semantic interoperability between EU and Member State legal systems, as the methods in use now for storing and displaying legal documents differ among countries. Therefore, the Council of the European Union has invited the introduction of ELI (European Legislation Identifier) [4] and ECLI (European Case Law Identifier) [3] standards that define common identifier and metadata models for legislative and case law documents by applying Linked Data principles.

3 Modeling Juridical Documents with RDF

Several features specific to legal documents need to be taken into account when modeling legislation with RDF. First of all, law is not constant but changes over time. The data model needs to be able to identify the different temporal versions of the law. Secondly, statutory citations refer to both entire statutes and their individual parts. Therefore, we need to identify the different document parts as well as their temporal versions. Moreover, different language versions of the same document may exist and content can be represented in multiple formats that all need to be identified.

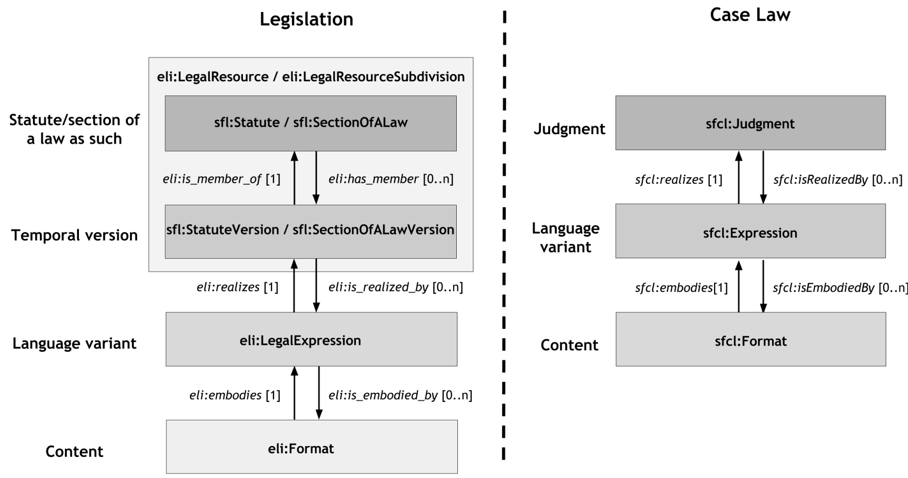

Fig. 1. The FRBR-inspired data models for both legislative and case law documents

The ELI standard applies the well-established conceptual model FRBR (Functional Requirements for Bibliographic Records) [12] to its ontology definition to distinguish between 1) statutes as such (*work*), 2) their language versions (*expression*), and 3) different content formats (*manifestation*). However, ELI does not provide an unanimous way to group together different temporal versions of a document. Such grouping is important, however, when maintaining a version history of statutes. According to the previously developed CEN Metalex ontology [2], both temporal versions and language versions of legal documents should be considered FRBR expression level entities. However, the ELI implementation guide [6] states that in multilingual environments, such as in Finland where both Finnish and Swedish are official languages, expressions should be used only to model the different language versions, and thus the work level should be used to denote the temporal versions.

We therefore decided to extend the ELI ontology with our own ontology SFL (Semantic Finlex Legislation) that defines separate classes for the different work level entities, i.e., the legislative documents and the document parts as such (*sfl:Statute* and *sfl:SectionOfALaw*), and their temporal versions (*sfl:StatuteVersion* and *sfl:SectionOfALawVersion*). The FRBR-inspired, ELI-compliant, extended data model is presented on the left side of Fig. 1. Different namespaces and prefixes used in the schemas are listed in Table 1. Statutes and their parts are linked to their temporal versions via *eli:has_member* property. Temporal versions in turn have two language variants in Finnish and Swedish. These language variants are modeled as instances of the *eli:LegalExpression* class and linked to the temporal versions with the property *eli:realizes*. Finally, the different content formats (text and HTML) of the language variants are modeled as instances of the *eli:Format* class and linked to the language variant with *eli:embodies* property.

Table 1. Prefixes and namespaces used in the RDF data models

Prefix	Namespace
common	http://data.finlex.fi/common/
dcterms	http://purl.org/dc/terms/
eli	http://data.europa.eu/eli/ontology#
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
sfcl	http://data.finlex.fi/schema/sfcl/
sfl	http://data.finlex.fi/schema/sfl/
skos	http://www.w3.org/2004/02/skos/core#

Table 2. ELI-compliant URI patterns for legislative documents

URI pattern	Description
/eli/sd/{year}/{id}	Statute as such
/eli/sd/{year}/{id}/luku/{#}/pykala/{#}/...	Section of a law as such
/eli/sd/{year}/{id}/.../alkup	Original version
/eli/sd/{year}/{id}/.../ajantasa/{date}	Consolidated version
/eli/sd/{year}/{id}/.../{version}/{language}	Language variant
/eli/sd/{year}/{id}/.../{version}/{language}/{format}	Content

Besides an ontology, ELI defines a URI pattern schema to unambiguously identify legislation. The URI patterns developed for Semantic Finlex are presented in Table 2. The original versions are denoted with parameter *alkup* in the URI and consolidated versions with *ajantasa/{date}*, where the date corresponds to the date of entry into force. Notice that the documents as well as their parts, their temporal versions, language versions, and content formats can all be identified uniquely using these patterns.

In addition to an extension to the FRBR structural classes of ELI, we have extended the ELI model with an additional property *sfl:statuteType* to describe the functionality of a statute, i.e., whether it is a new statute, an amendment or a repeal. ELI itself defines descriptive properties *eli:type_document*, that we use to describe the level of a statute in the hierarchy of norms (i.e., whether the statute is an act, decree, or decision), and *eli:version* to denote whether the statute is an original version or a consolidated version.

The internal, physical structure of a statute is modeled at the temporal version level, as it can vary between versions. To describe the document structure, the documents and their individual parts are annotated with *eli:has_part* and

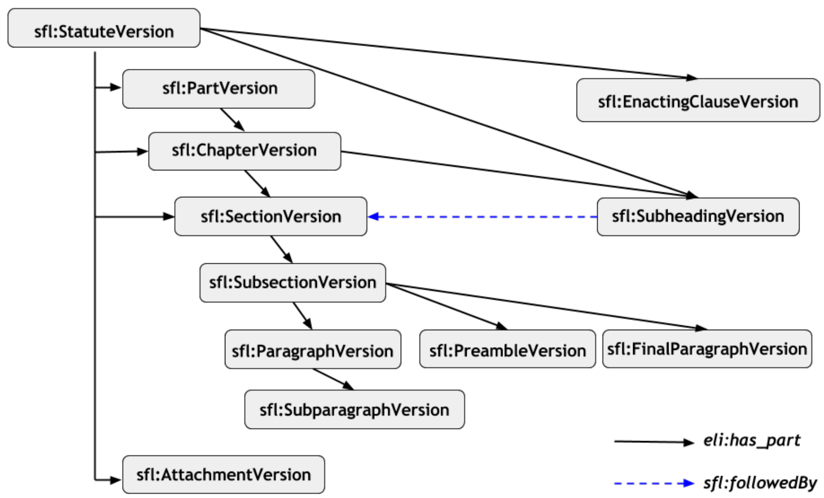

Fig. 2. Model of the physical structure of a statute

eli:is_part_of properties. Each type of section of a law is modeled as a separate class. For example, a statute can consist of multiple sections, while sections consist of subsections and subsections consist of paragraphs. The model of the physical structure is presented in Fig. 2.

Table 3. ELI-mimicing URI patterns for case law documents

URI pattern	Description
/ecli/{court}/{year}/{id}	Judgment
/ecli/{court}/{year}/{id}/{language}	Language variant
/ecli/{court}/{year}/{id}/{language}/{format}	Content

For case law, we adapted the ECLI standard. In contrast to the sophisticated functional model of ELI, ECLI only defines a set of Dublin Core properties to be used to annotate case law documents. Thus we have developed our own ontology called SFCL (Semantic Finlex Case Law) that wraps the ECLI metadata model to an FRBR-inspired model reminiscent of ELI. However, we can omit temporal versions from the model since there is only one temporal version of each judgment in both Finnish and Swedish. The FRBR model for case law is presented on the right side of Fig. 1. In addition, for the case law documents we generate URIs that mimic the ELI pattern, because the standard format for document identifiers defined by ECLI is not an HTTP URI. The URI pattern is presented in Table 3. The physical structure of the case law documents is not modeled in RDF, and therefore no identifiers for the document parts need to be generated.

ELI and ECLI seem to be very different standards, developed separately by separate teams. It is also unclear why the ELI model was not made more congruent with the Metalex ontology[7] than it is. In the ELI ontology[8] it is stated that the decision was made not to align with the Metalex class names that express the FRBR hierarchy (namely *BibliographicWork*, *BibliographicExpression*, *BibliographicManifestation*) because "ELI is specifically about publishing legal metadata on the web, and not structuring legal content in any document". We argue that publishing metadata about legal documents on the web is largely about structuring content in the documents because of the aforementioned requirements to be able to identify and annotate physical parts, temporal versions and language versions.

4 Data Conversion

The Semantic Finlex service currently consists of four different datasets:

Original legislation. This dataset consists of approximately 47 000 acts and decrees as they originally appeared in the Statutes of Finland, the official publication of Finnish Law. Besides new acts and decrees, the dataset includes amendments and repeals targeted on previously enacted statutes.

Consolidated legislation. Consolidated texts of acts and decrees incorporate their successive amendments. Editorial work has been carried out by a publishing company. Currently the dataset includes approximately 2 800 statutes.

Judgments of the Supreme Court. This dataset comprises around 5 500 precedents published in the Yearbook of the Supreme Court since 1980.

Judgments of the Supreme Administrative Court. This dataset includes roughly 9 300 judgments of the Supreme Administrative Court since 1987.

All of the above datasets are transformed from different legacy XML formats to RDF adapting the data models described in the previous section. New and updated documents are fetched weekly from the Finlex service and converted into RDF.

When converting a consolidated text of a statute, the parts of the statute that have changed between updates need to be identified in order to produce ELI-compliant identifiers and metadata. Identifier for a section of a law that has been changed needs to be resolved by locating the date when the amendment in question comes into force. Fig. 3 presents how the identifiers of the temporal versions are managed in the data. At time t_1 there is a section that consist of two subsections. At time t_2 a new subsection is added between the other two subsections, and new temporal versions with a different version dates (t_2) are generated for section 1 and subsections 2 and 3. As the content of subsection 3, however, remains the same and it only moves in hierarchy, we link it to

[7] `http://justinian.leibnizcenter.org/MetaLex/metalex-cen.owl`
[8] `http://publications.europa.eu/mdr/resource/eli/eli.owl`

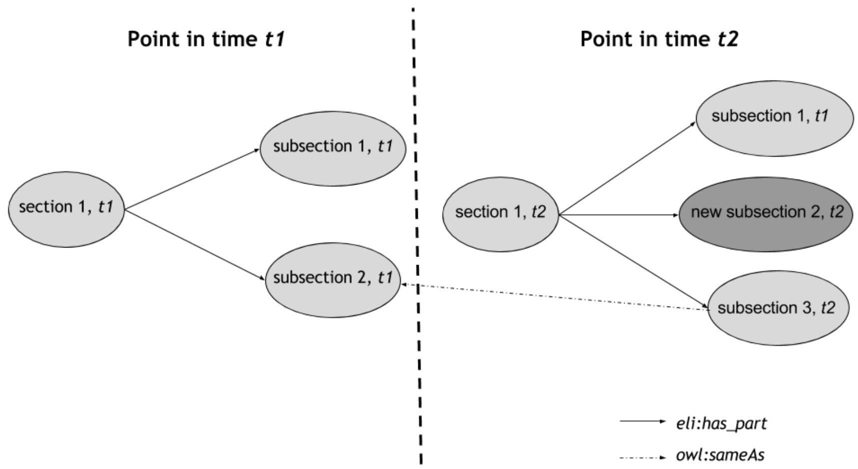

Fig. 3. An example of a subsection added between two other subsections

the previous version with the *owl:sameAs* property. The version date $t2$ is also inherited by all of the elements hierarchically above section 1, but the version date of subsection 1 remains the same (t_1), because neither its content or its position has changed.

Resolving the date of entry into force is not always a trivial task, since the legacy XML files do not provide these dates as explicit metadata. Therefore the information must be extracted from plain text. It is often straightforward to recognize some of the dates from the list of information about the amendments contained at the end of the consolidated document, but there are anomalies and exceptions, too. For example, different parts of a single amendment might enter into force and stop being in force at different dates.

We could use the original versions of the acts and their amendments to automatically compile the consolidated versions but, unfortunately, we do not have a complete history of all the original documents. Moreover, the XML structure of an amendment does not contain identifiers for the subsections that have changed. Those identifiers would have to be extracted from natural language in the enacting clause of the amendment.

Characteristic to the legislative XML formats is the lack of ELI-compliant metadata. In order to produce the RDF metadata it must be extracted from the document text during the conversion process using regular expressions. If the required text is not in the assumed place in the assumed form, then the conversion process may result in missing values and possibly errors in the converted data. Therefore, rules need to be defined to validate the data before allowing it to be published in the service. These rules can be expressed in the form of SPARQL queries performed against the converted RDF data. An example of such a query is one that verifies that the date of entry into force does not precede the date of publication of a statute. In addition, the ELI validator[9] developed by Sparna Labs can be used to check the conformance of the RDF data against the ELI ontology. The validator applies predefined SHACL shapes[10] (RDF Shapes Constraints Language) to the input data and files a report on the identified constraint violations.

Converting the judgments into RDF is quite straightforward as most of the metadata fields mentioned in the definition of ECLI are included in the original XML files. The ECLI identifiers appearing in the XML documents are converted to HTTP URIs that mimic the ELI pattern. For example the ECLI identifier

ECLI:FI:KKO:2016:1

is transformed to

http://data.finlex.fi/ecli/kko/2016/1

Document contents are stored at the manifestation level in text and HTML formats as values to RDF properties *sfl:text*, *sfl:html*, *sfcl:text* and *sfcl:html*. This enables the HTML or text content of a specific section of a law to be retrieved with a single SPARQL query from the triple store. We therefore do not have a separate store for the documents except for the original XML files provided as archive packages.

5 Enriching the Data with Relationships and Key Concepts

To describe the interconnectedness of the law with RDF, references to different sources of law are extracted from the documents. These relationship links are always assigned to the most detailed part of a legislative document.

First of all, the relationship between the consolidated versions and the original amendments is modeled by linking the amendments to the corresponding consolidated versions with the *eli:amends* and *eli:amended_by* properties. This model is described in Fig. 4. If the section of a law in question has been repealed, a link to the repealing statute is created instead, as depicted in Fig. 5.

The linkage to the EU legislation provides one way to link legislative documents between individual member states. ELI defines the *eli:transposes* property to describe which legal acts of the EU a statute transposes into national law. The original versions of the statutes contain references to EU directives and regulations for which ELI-compliant URIs can be generated automatically by following the ELI URI pattern.

To further enrich the metadata of the case law documents, the names of the justices of the Supreme Court and other staff linked to a specific case are extracted from the text of each document. This is done by using regular

[9] `http://labs.sparna.fr/eli-validator/`

[10] `http://www.w3.org/TR/shacl/`

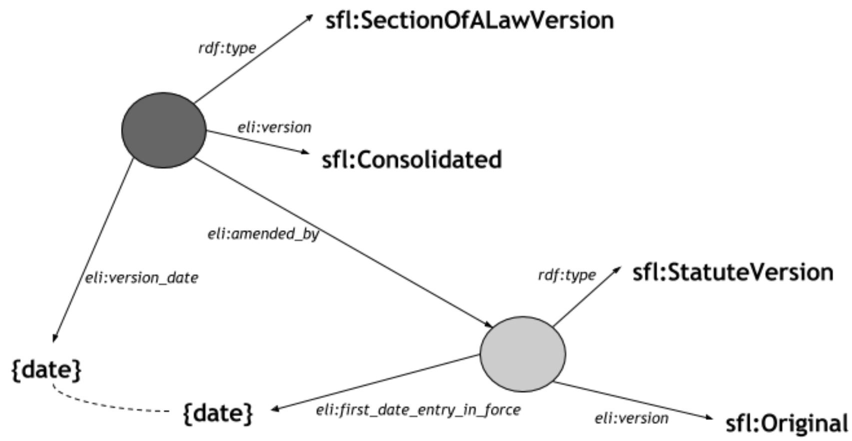

Fig. 4. Consolidated versions are linked to the corresponding original amendments.

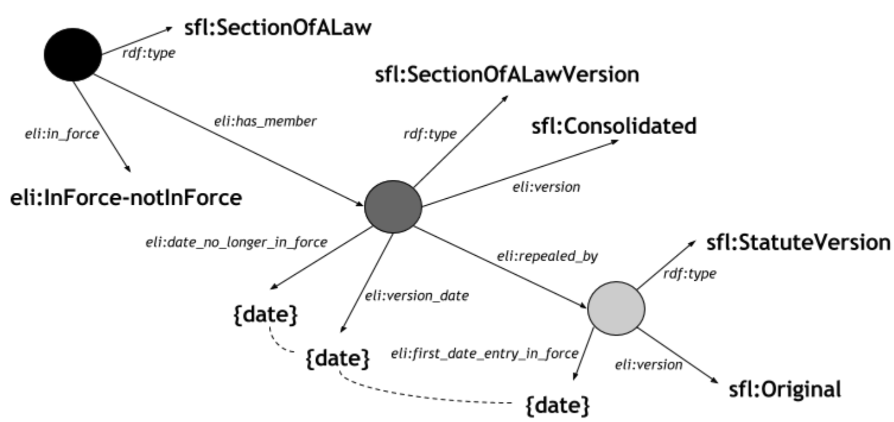

Fig. 5. Repealed consolidated versions are linked to the corresponding original repeals.

expressions that match known types of names. The personnel are modeled as *dcterms:Agent* type of resources and linked to the judgment with the property *dcterms:contributor* in accordance with the ECLI specification.

Besides the names of the contributors, references to legislative texts are extracted from the case law documents. These are annotated with the property *sfcl:refToLegislation*. However, since we do not know which version of a legislative text the citation refers to, we always resolve the link to the abstract work level of a statute or a section of a law, and not any specific temporal version.

To support search and discovery of legal texts, key concepts relating to pieces of legislation were automatically mined from the texts of the documents. These semantic annotations were selected from the following vocabularies: The Bank of Finnish Terminology in Jurisprudence[11], Eurovoc[12], the legal terminology sections of the KOKO ontology[13], and DBpedia[14].

Before querying the vocabularies, the texts were filtered using stop word lists and linguistic tools. First, the entire text was lemmatized using SeCo Lexical Analysis Services [14]. After this, stopword lists were applied to filter out words too general in this context (such as the term *legislation* itself). After this, n-grams from the preprocessed texts were compared against terms in the vocabularies to discover candidate key concepts. Once the results were at hand, the final step was to use a weighting scheme (TF-IDF) to pick only the relevant candidates.

Once the best annotations had been selected the results were written in RDF format and uploaded to the Semantic Finlex service. The annotations are placed in their own graph[15] using the property *common:autRecSubject*.

6 Publishing Platform and Application Programming Interfaces

The Semantic Finlex service adopts the 5 star deployment scheme suggested by Tim Berners-Lee [1]. The Semantic Computing Research Group has previously proposed an extension to the 5 star scheme by adding two more stars to it [11]. The 6th star is obtained by providing the dataset schemas and documenting them. Semantic Finlex schemas can be downloaded from the service and the data models are documented under the `data.finlex.fi` domain. The 7th star is achieved by validating the data against the documented schemas to prevent errors in the published data. Semantic Finlex attempts to obtain the 7th star by applying different means of combing out errors in the data within the data conversion process. The service is powered by the Linked Data Finland[16] publishing platform that along with a variety of different datasets provides tools and services to facilitate publishing and re-using Linked Data.

[11] `http://tieteentermipankki.fi/wiki/Oikeustiede`
[12] `http://eurovoc.europa.eu`
[13] `http://finto.fi/koko/fi/`
[14] `http://dbpedia.org`
[15] `http://data.finlex.fi/annotation/sd`
[16] `http://ldf.fi`

Following the Linked Data principles all URIs are dereferenceable and support content negotiation by using HTTP 303 redirects. In accordance with the ELI specification, RDF is embedded in the HTML presentations of the legislative documents as RDFa[17] markup. In addition to the converted RDF data, the original XML files are also provided.

To support easier use by programmers without knowledge of SPARQL or RDF, a simplified REST API is provided. This API can be accessed by using the URI patterns and specifying JSON as the preferred content type in the header of the HTTP request. In truth, this API also returns its data in the JSON-LD RDF format, but much thought has been given to organize the returned data in such a way that it is as intuitively as possible and usable also as pure JSON. For example, the affordances provided by JSON-LD @*context* definitions are used to encode language versions of texts in *content_fi* and *content_sv* properties, instead of the user needing to filter the rich literals for their desired language. In addition, URL parameters are provided for retrieving the information pertinent to most common use cases in a stable structure, such as being able to specify which temporal, language, and format versions (of txt and html) of the legislation are required. Finally, a *tree* parameter is provided to build and return the entire subtree of a legislative document without the need to resort to complicated SPARQL queries.

On the other hand, for queries that go beyond fetching information on individual pieces of legislation (such as relational or data analysis queries), a SPARQL endpoint is also available, and a number of sample SPARQL queries are provided to draw inspiration from.

7 Evaluation and Applications

7.1 Data analysis

Regarding data analysis, sample SPARQL queries were drafted to extract interesting information from the data. These were then fed through to Google Charts for visualization. Information queried was, for example:

- Laws most often referred to from other legislation as well as court decisions
- Laws that have been changed or amended the most
- The years in which the above laws were laid
- The number of EU transpositions by year
- The members of the supreme court with the most decisions, as well as their tenures
- The most common topics for supreme court cases by their key concepts

7.2 Applications.

In addition to drafting examples of data analysis, the following application demonstrators were built on top of the Semantic Finlex Linked Data service.

[17] http://www.w3.org/standards/techs/rdfa

Legal recommender The HTML representations of the documents are enriched with recommendations to related sources of law similar to [19]. For example, links to relevant EU legislation are queried from the CELLAR SPARQL service[18] by matching their Eurovoc based keywords with the semantic annotations in the Semantic Finlex datasets.

Semantic search. A text-based search tool with semantic autocomplete [10] for the documents has been implemented in connection with the service. The search tool works for both legislative and case law documents. The search is targeted on different entities in the following order of importance: keywords, document titles, section headings, and texts.

Tag clouds. Tag clouds were used to visualize the contents of the documents. For each statute, a tag cloud was generated using the same process as with the semantic annotations of key concepts described earlier.

Contextual reader. To support users in making sense of the legal terminology in the law and court decision texts themselves, the CORE contextual reader [15] was configured with the legal terminology stored in Semantic Finlex. This enables highlighting each instance of specialist terminology in the documents with a popover providing the definition of that term on a mouse-over. Further, clicking on any such mention brings in other laws, decisions and legal news articles pertaining to that topic, thus facilitating further semantic browsing and delving more deeply into matters either interesting or unclear.

Annotation widget. Publishing legislation as Linked Data enables the use of statutes as a reference ontology for linking and integrating heterogeneous datasets that refer to law. The SPARQL endpoint of data.finlex.fi can be utilized as an ontology service [5] for finding relevant statutes or their parts to be used in metadata descriptions. We have prototyped an autocompletion annotation widget that allows the user to search for statutes and fetch their URIs into a cataloging system by typing in a part of the statute name, in the same vain as in ONKI [18].

Content widget. As the legislation dataset includes the textual contents of statutes, it can be used to enrich external websites with up-to-date law texts as a web widget [16]. For example, a news article or government announcement informing of a critical change in a statute can be accompanied with the new versions of the relevant parts of the statute. Once the widget is configured with the URI of the desired statute (or its specific part), it will perform a SPARQL query to fetch the text content of the statute, and display it on the web page.

8 Related Work and Discussion

Similar efforts to publish legislation and case law as Linked Open Data have been conducted in various countries. The main inspiration for our work was the MetaLex Document Server[19] [9], that provides regularly updated Dutch legislation as Linked Open Data utilizing the CEN Metalex XML and ontology

[18] `http://publications.europa.eu/webapi/rdf/sparql`

[19] `http://doc.metalex.eu`

standards. Another known example of a Metalex based legal Linked Data service is legislation.gov.uk[20] that hosts UK legislation in local XML formats together with RDF metadata based on the Metalex ontology. There is also a Metalex ontology based implementation of a legal Linked Data service in Greece, named Nomothesia[21], that also implements ELI-compliant identifiers.

Various ELI implementations and prototypes have also been implemented, usually by resolving ELI-compliant URIs and rendering ELI metadata to existing legal information portals such as in Luxembourg [22], France [23], and Norway [24]. Many countries already produce ECLI-compliant case law documents to be indexed by the ECLI search engine[25]. Semantic Finlex aims to widen focus by providing both legislation and case law as Linked Open Data through simple Linked Data APIs and linking the datasets with each other. The attempt to unify the ELI and ECLI standards is also something we haven't seen in other implementations.

A valuable lesson learned during the project was that the complexity of the legislative system leads to the fact that the automation of publishing up-to-date legislation without the need for manual editorial work won't be obtained overnight. In the future we may apply parts of the Metalex as well as some other standards to more accurately model the different types of legislative modifications as events. However, to produce documents with the metadata required for such a rich model to be applied, and without the need for costly editorial work or error-prone automated NER techniques, the legislative XML and RDF standards, such as the Metalex-compliant Akoma Ntoso [17], should be applied as early as in the legislative and judicial processes where the documents are drafted.

Another issue is that due to copyrights caused by editorial work carried out by the publishing company, we've had to publish the consolidated legislation under a license that restricts its commercial use. Moreover, we cannot use the reference database of legislation maintained by the publishing company to look up the dates of entry into force associated with the original statutes and the amendments. If we were allowed to use the reference database we would not have to resort to NER techniques as often as we do now. The simplest method to circumvent the copyright issues altogether would be to eliminate the need for consolidation by changing the legislative process so that new versions of complete statutes would be published instead of amendments consisting of individual sections.

The new version of the Semantic Finlex service was released on March 10, 2016, and has been in use since. The development of the service is carried on by further enriching and validating the legal datasets and implementing new

[20] `http://legislation.gov.uk`
[21] `http://legislation.di.uoa.gr`
[22] `http://legilux.public.lu/editorial/eli`
[23] `http://www.eli.fr/en/constructionURI.html`
[24] `http://lovdata.no/eli`
[25] `https://e-justice.europa.eu/content_ecli_search_engine-430-en.do`

applications on top of them. To further evaluate the usefulness of the service in practice we are planning to arrange a public hackathon later in 2017.

Acknowledgements The work is a joint effort between Aalto University, the Ministry of Justice, the Ministry of Finance, Edita Publishing Ltd., Tekes and Viestintäalan tutkimussäätiö. We thank Risto Talo, Jari Linhala and Sari Korhonen of Edita Publishing Ltd. for collaboration.

References

1. Berners-Lee, T.: Design Issues: Linked Data, `https://www.w3.org/DesignIssues/LinkedData.html`
2. Boer, A., van Engers, T.: A MetaLex and metadata primer: Concepts, use, and implementation. In: Sartor, G., Palmirani, M., Francesconi, E., Biasiotti, M.A. (eds.) Legislative XML for the Semantic Web: Principles, Models, Standards for Document Management, pp. 131–149. Springer Netherlands, Dordrecht (2011)
3. Council of the European Union: Council conclusions inviting the introduction of the European Case Law Identifier (ECLI) and a minimum set of uniform metadata for case law. In: Official Journal of the European Union, C 127, 29.4.2011. pp. 1–7. Publications Office of the European Union (2011)
4. Council of the European Union: Council conclusions inviting the introduction of the European Legislation Identifier (ELI). In: Official Journal of the European Union, C 325, 26 October 2012. pp. 3–11. Publications Office of the European Union (2012)
5. d'Aquin, M., Noy, N.F.: Where to publish and find ontologies? A survey of ontology libraries. Web Semantics: Science, Services and Agents on the World Wide Web 11, 96–111 (2012)
6. ELI Task Force: ELI: A Technical Implementation Guide. Publications Office of the European Union (2015)
7. Frosterus, M., Tuominen, J., Hyvönen, E.: Facilitating re-use of legal data in applications–Finnish law as a linked open data service. In: Proceedings of the 27th International Conference on Legal Knowledge and Information Systems JURIX 2014, Krakow, Poland. pp. 115–124. IOS Press (2014)
8. Hietanen, A.: Free access to legislation in finland: Principles, practices and prospects. In: Ragona, M. (ed.) Law via the Internet. Free Access – Quality of Information – Effectiveness of Rights. European Press Academic Publishing, Florence (2009)
9. Hoekstra, R.: The MetaLex Document Server legal documents as versioned linked data. In: Proceedings of the ISWC 2011, Bonn, Germany. pp. 128–143. Springer-Verlag (2011)
10. Hyvönen, E., Mäkelä, E.: Semantic autocompletion. In: Proceedings of the First Asia Semantic Web Conference (ASWC 2006), Beijing. Springer–Verlag (2006)
11. Hyvönen, E., Tuominen, J., Alonen, M., Mäkelä, E.: Linked Data Finland: A 7-star model and platform for publishing and re-using linked datasets. In: Proceedings of ESWC 2014 Demo and Poster Papers, Springer-Verlag (May 2014)
12. IFLA: Functional Requirements for Bibliographic Records: Final Report. K. G. Saur (1998)
13. Lindholm, J., Derlén, M.: Festina lente – Europarättens genomslag i svensk rättspraxis 1995–2015. Europarättslig tidskrift 1, 151–177 (2015)

14. Mäkelä, E.: LAS: an integrated language analysis tool for multiple languages. The Journal of Open Source Software 1(6) (oct 2016), http://dx.doi.org/10.21105/joss.00035
15. Mäkelä, E., Lindquist, T., Hyvönen, E.: CORE - a contextual reader based on linked data. In: Proceedings of Digital Humanities 2016, long papers. pp. 267–269 (July 2016), http://dh2016.adho.org/abstracts/2580
16. Mäkelä, E., Viljanen, K., Alm, O., Tuominen, J., Valkeapää, O., Kauppinen, T., Kurki, J., Sinkkilä, R., Känsälä, T., Lindroos, R., Suominen, O., Ruotsalo, T., Hyvönen, E.: Enabling the semantic web with ready-to-use web widgets. In: Proceedings of the First Industrial Results of Semantic Technologies Workshop, ISWC2007 (November 11 2007)
17. Palmirani, M.: Legislative Change Management with Akoma-Ntoso. In: Sartor, G., Palmirani, M., Francesconi, E., Biasiotti, A.M. (eds.) Legislative XML for the Semantic Web: Principles, Models, Standards for Document Management, pp. 101–130. Springer Netherlands, Dordrecht (2011)
18. Tuominen, J., Frosterus, M., Viljanen, K., Hyvönen, E.: ONKI SKOS server for publishing and utilizing SKOS vocabularies and ontologies as services. In: Proceedings of the 6th European Semantic Web Conference (ESWC 2009). Springer–Verlag (2009)
19. Winkels, R., Boer, A., Vredebregt, B., von Someren, A.: Towards a legal recommender system. In: Proceedings of the 27th International Conference on Legal Knowledge and Information Systems JURIX 2014, Krakow, Poland. pp. 169–178. IOS Press (2014)