

# Semantic Finlex: Finnish Legislation and Case Law as a Linked Open Data Service

Arttu Oksanen<sup>1,3</sup>, Jouni Tuominen<sup>1</sup>, Eetu Mäkelä<sup>1</sup>,  
Minna Tamper<sup>1</sup>, Aki Hietanen<sup>2</sup>, and Eero Hyvönen<sup>1</sup>

<sup>1</sup> 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](mailto:firstname.lastname@aalto.fi)

<sup>2</sup> Ministry of Justice, Finland

[firstname.lastname@om.fi](mailto:firstname.lastname@om.fi)

<sup>3</sup> Edita Publishing Ltd

## 1 The Problem: Legislation and Case Law as Data

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.

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<sup>4</sup>, a publicly available online service since 1997, maintained by the Ministry of Justice [4]. 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.

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) [2] and ECLI (European Case Law Identifier) [1] standards that define

---

<sup>4</sup> <http://www.finlex.fi>

common identifier and metadata models for legislative and case law documents by applying Linked Data principles.

Semantic Finlex<sup>5</sup>, a national Linked Open Data Service for Finnish legislation and case law addresses these problems and trends based on new European and Semantic Web standards. The service hosts and publishes a central part of the Finnish legislation along with judgments of the Supreme Court and the Supreme Administrative Court. All of the datasets are automatically updated regularly.

## 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)<sup>6</sup> by Talentum Oyj and Edilex<sup>7</sup> 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 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

---

<sup>5</sup> <http://data.finlex.fi>

<sup>6</sup> <http://www.suomenlaki.com>

<sup>7</sup> <http://www.edilex.fi>

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 [6]. 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.

### 3 Related Work

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<sup>8</sup> [5], that provides regularly updated Dutch legislation as Linked Open Data utilizing the CEN Metalex XML and ontology standards. Another known example of a Metalex based legal Linked Data service is legislation.gov.uk<sup>9</sup> 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<sup>10</sup>, 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<sup>11</sup>, France<sup>12</sup>, and Norway<sup>13</sup>. Many countries already produce ECLI-compliant case law documents to be indexed by the ECLI search engine<sup>14</sup>. 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.

---

<sup>8</sup> <http://doc.metalex.eu>

<sup>9</sup> <http://legislation.gov.uk>

<sup>10</sup> <http://legislation.di.uoa.gr>

<sup>11</sup> <http://legilux.public.lu/editorial/eli>

<sup>12</sup> <http://www.eli.fr/en/constructionURI.html>

<sup>13</sup> <http://lovdata.no/eli>

<sup>14</sup> [https://e-justice.europa.eu/content\\_ecli\\_search\\_engine-430-en.do](https://e-justice.europa.eu/content_ecli_search_engine-430-en.do)

## 4 Paper Outline

Our work on Semantic Finlex started in 2012, and the first version of the service was published in 2014 [3]. 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 paper, 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.

**Acknowledgements** The work is a joint effort between Aalto University, HELDIG – Helsinki Centre for Digital Humanities, 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. 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)
2. 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)
3. 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)
4. 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)
5. 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)
6. Lindholm, J., Derlén, M.: Festina lente – Europarättens genomslag i svensk rättspraxis 1995–2015. Europarättslig tidskrift 1, 151–177 (2015)