

# LAWSAMPO: A Semantic Portal on a Linked Open Data Service for Finnish Legislation and Case Law

Eero HYVÖNEN <sup>a,b</sup>, Minna TAMPER <sup>a</sup>, Esko IKKALA <sup>a</sup>, Sami SARSA <sup>a</sup>,  
Arttu OKSANEN <sup>a,c</sup>, Jouni TUOMINEN <sup>a,b</sup>, and Aki HIETANEN <sup>d</sup>

<sup>a</sup> *Semantic Computing Research Group (SeCo), Aalto University, Finland*

<http://seco.cs.aalto.fi>, [firstname.lastname@aalto.fi](mailto:firstname.lastname@aalto.fi)

<sup>b</sup> *HELDIG – Helsinki Centre for Digital Humanities, University of Helsinki*

<http://heldig.fi>

<sup>c</sup> *Edita Publishing Ltd.*

<http://www.editapublishing.fi>

<sup>d</sup> *Ministry of Justice, Finland*

<http://oikeusministerio.fi>, [firstname.lastname@om.fi](mailto:firstname.lastname@om.fi)

**Abstract.** This paper presents the semantic portal prototype LAWSAMPO for serving end users with legal data from the Linked Open Data service Semantic Finlex of the Finnish Ministry of Justice. The general goal of the system is to aggregate heterogeneous distributed data and represent it as a harmonized knowledge graph on top of which intelligent use-case application perspectives can be created. First applications include faceted semantic search and browsing engines for 1) statutes and 2) court decisions, as well as 3) a service for finding court decisions based on other court decisions. LAWSAMPO will provide the end user with ready to use tooling for analysing legal documents. Linked Data is utilized for creating a “homepage” for each document that is enriched with related data for a more comprehensive reading experience.

**Keywords.** linked data, case law, legislation, semantic portal

## 1. Vision: Publishing Law and Justice as Linked Open Data

Citizens are expected to obey the law, but this is not possible if they do not know the legislation and how law is applied in courts. Governments therefore publish lots of legal information in print and more and more on the Web. In Finland, for example, legislation and case law has been published since 1997 as HTML documents in the Finlex Data Bank<sup>1</sup>, a publicly available online service maintained by the Ministry of Justice [6].

Legal documents on the Web are usually available as HTML or PDF documents and targeted to human readers. This severely hinders their use when re-

---

<sup>1</sup><http://www.finlex.fi>

lated heterogeneous information and documents need to be aggregated from distributed sources (e.g., EU level and national legislation), and when developing intelligent legal applications and web services. Advanced applications usually require that the computer “understands” in one way or another the semantics of the textual documents. This paper argues that Semantic Web standards<sup>2</sup> and the Linked Data publishing paradigm [5] provide a promising approach to address these problems. To demonstrate and evaluate this argument, we present two related Living Laboratory systems in Finland: the Semantic Finlex<sup>3</sup> Linked Data Service and the semantic portal LAWSAMPO<sup>4</sup> on top of it.

This paper first overviews the Semantic Finlex (SF) service and the underlying “7-star” publishing model and Linked Data Finland platform. After this, the underlying new ideas and first experiments of creating application perspectives in the semantic portal LAWSAMPO are presented. In conclusion, related works and contributions of LAWSAMPO are discussed, and next steps ahead outlined.

## 2. Semantic Finlex Linked Open Data Service

The starting point for our work in 2012 was the Finlex Data Bank. Although it was widely used by the public, it did not provide machine-readable legal information as open data, on top of which services and analyses could be built by the ministry or third party vendors. The first version of SF based on Linked Data was published in 2014 [4]. The data included 2413 consolidated laws, 11904 judgments of the Supreme Court, and 1490 judgments of the Supreme Administrative Court. In addition, some 30000 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: ELI European Legislation Identifier [3] and ECLI European Case Law Identifier [2]. The demo dataset also consisted of only one temporal version (2012) of the statutory law and was not updated automatically, as new legislation and case law was published in Finlex. These issues were resolved in the next version of the Semantic Finlex. [11] that was published on-line in 2017.

The Semantic Finlex service adopts the 5-star model suggested by Tim Berners-Lee [1], extended with two more stars, as suggested in the Linked Data Finland model and platform [9]. 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<sup>5</sup> publishing platform that along with a variety of different datasets pro-

---

<sup>2</sup><https://www.w3.org/standards/semanticweb/>

<sup>3</sup>Semantic Finlex is available with documentation at <http://data.finlex.fi>.

<sup>4</sup>A demonstrator of LAWSAMPO is available at <http://test.lakisampo.fi>.

<sup>5</sup><http://ldf.fi>

vides tools and services to facilitate publishing and re-using Linked Data. 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<sup>6</sup> 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. This API 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 intuitive as possible and usable also as pure JSON.

### 3. LawSampo Semantic Portal

To demonstrate and test using the SF data service, the semantic portal LAWSAMPO is under development. The first test demonstrations will be soon on-line<sup>7</sup>. LAWSAMPO is a new member of the Sampo series of semantic portals, based on the “Sampo model” [8], where the data is enriched through a shared ontology and Linked Data infrastructure by related external heterogeneous datasets.<sup>8</sup> The model has been developed and tested in a series of several practical case studies, including CultureSampo<sup>9</sup> (2008), TravelSampo<sup>10</sup> (2011), BookSampo<sup>11</sup> (2011), WarSampo<sup>12</sup> (2015), BiographySampo<sup>13</sup>, and NameSampo<sup>14</sup> (2019). Our experiences suggest that the Sampo model is a promising way to create useful systems that end users like. For example, in 2018, BookSampo had ca. 2 million users and WarSampo 230 000.

The idea of LAWSAMPO is to provide the end user with different application perspectives corresponding to different use cases. The applications demonstrate two ways of re-using the SF data service: First, the data can be used by implementing applications on the client side by using only the SF SPARQL endpoint without server side concerns. Second, the data is available for download and can be used for establishing other kind of backend services and web applications based on them. Similar demonstrators could be implemented by any external party.

Figure 1 depicts the landing page of LAWSAMPO with three perspectives:

**1. Statutes.** By clicking on Statutes, a faceted search interface [15] for searching and browsing statutes is opened (cf. Figure 2). The facets on the left include document type (with seven subtypes), statute type, year, and related EU directive. By making a selection on any facet category in any order, the corresponding fil-

<sup>6</sup><http://www.w3.org/standards/techs/rdfa>

<sup>7</sup><http://test.lakisampo.fi>

<sup>8</sup>In Finnish mythology and the epic Kalevala, “Sampo” is a mythical artefact of indeterminate type that gives its owner richness and good fortune, an ancient metaphor of technology.

<sup>9</sup><https://seco.cs.aalto.fi/applications/kulttuurisampo/>

<sup>10</sup><https://seco.cs.aalto.fi/applications/travelsampo/>

<sup>11</sup><https://seco.cs.aalto.fi/applications/kirjasampo/>

<sup>12</sup><https://seco.cs.aalto.fi/projects/sotasampo/en/>

<sup>13</sup><https://seco.cs.aalto.fi/projects/biografiasampo/en/>

<sup>14</sup><https://seco.cs.aalto.fi/projects/nimisampo/>

September 2019

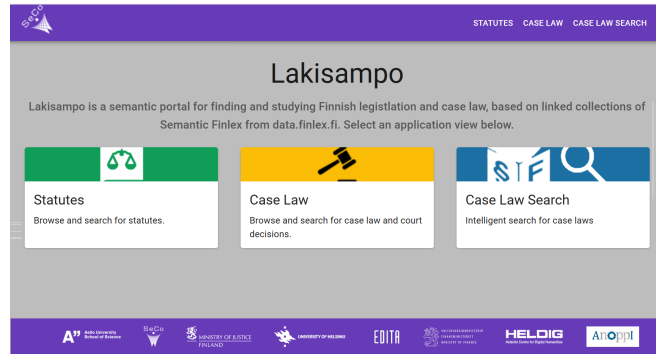


Figure 1. LAWSAMPO landing page with three application perspectives

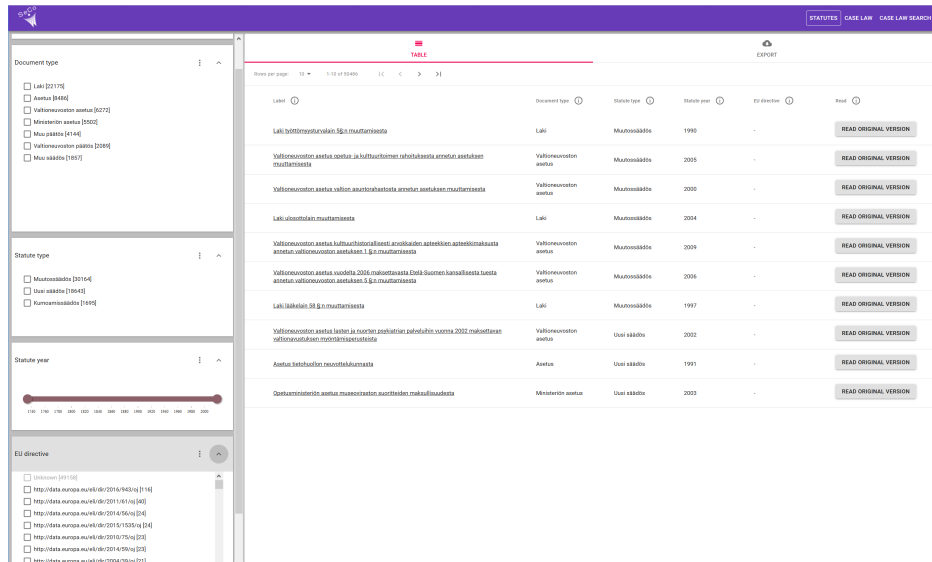


Figure 2. Statutes application perspective with faceted search

tering is performed and the results are shown to the end user. At the same time, hit counts of the facet categories are recomputed. In this way, the user always sees the numerical distribution of results along the facets, is guided towards good filtering selections, and will never make a selection leading to an empty result.

After filtering out a set of documents (or a particular document) of interest, the user is provided with two options. First, the user can select a document from the result list and a “homepage” of the document opens, showing not only the document but also linked contextual information related to it. For example, the court decision in which the statute has been applied can be shown, or the underlying memorandum of the government, or parliamentary discussions related to the law. At the moment, the document itself is shown and work on providing

the additional linkage is going on. Second, it is possible to do data analysis based on the filtered documents. This idea is an adaptation from BiographySampo, where tooling for prosopographical research based on biographical documents is provided. As a simple example, a histogram can be created showing the dates of the filtered documents. Implementing these features is a work in progress at the moment.

**2. Case Law.** In the Case Law perspective, a similar faceted search interface opens for searching and browsing court decisions. In this case, the facets include court, judge, and keywords characterizing the subject matter of the judgement.

**3. Case Law Search.** The third perspective is an application, where a law case judgement, or more generally any document or text, can be used for finding similar other case judgements. For example, if one gets a judgement from a court, this application can be used to find out what kind of similar judgements have been made before. In this way it is possible to determine, whether one has been treated properly and whether it makes sense to appeal to a higher court. Several methods for finding similar cases were tested when implementing this application including TF-IDF, Latent Dirichlet Allocation (LDA), Word2Vec, and Doc2vec [12,13].

#### 4. Related Work and Contributions

Our work on legal Linked Data services was influenced by the MetaLex Document Server<sup>15</sup> [7] that publishes Dutch legislation using the CEN Metalex XML and ontology standards. Other Metalex ontology based implementations include legislation.gov.uk<sup>16</sup> for UK legislation and Nomothesia<sup>17</sup> that also implements ELI-compliant identifiers. Various ELI implementations and prototypes have also been implemented and used in related existing legal information portals nationally, e.g., in Luxembourg<sup>18</sup>, France<sup>19</sup>, and Norway<sup>20</sup>. Many countries already produce ECLI-compliant case law documents to be indexed by the ECLI search engine<sup>21</sup>. A prominent example of publishing EU Law and publications as linked data is the CELLAR system<sup>22</sup>.

LAWSAMPO aims to widen the focus of these related works by providing both legislation and case law to end users through intelligent user interfaces, such as semantic faceted search and document similarity-based search. The documents are automatically enriched with contextual linked data on their “homepages”, and the end user is also provided with ready-to-use data-analytic tooling for analyzing the documents and their relations. In the future, we plan to expand the related enriching datasets to include, e.g., related parliamentary documents and discussions, in the spirit of [16]. In order to be able to publish more legal documents in

---

<sup>15</sup><http://doc.metalex.eu>

<sup>16</sup><http://legislation.gov.uk>

<sup>17</sup><http://legislation.di.uoa.gr>

<sup>18</sup><http://legilux.public.lu/editorial/eli>

<sup>19</sup><http://www.eli.fr/en/constructionURI.html>

<sup>20</sup><http://lovdata.no/eli>

<sup>21</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)

<sup>22</sup><https://data.europa.eu/euodp/data/dataset/sparql-cellar-of-the-publications-office>

a more cost-efficient way, we also work on semi-automatic pseudonymization of court judgements [10] and automatic annotation of legal documents [14].

**Acknowledgments** We thank Saara Packalén, Tiina Husso, and Oili Salminen at Ministry of Justice, and Risto Talo, Jari Linhala, and Sari Korhonen of Edita Publishing Ltd. for collaboration. Our work was funded by the Ministry of Justice and Ministry of Finance of Finland. CSC – IT Center for Science, Finland, provided computational resources for the work.

## References

- [1] T. Berners-Lee. Design Issues: Linked Data, 2006. <https://www.w3.org/DesignIssues/LinkedData.html>.
- [2] 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*, pages 1–7. Publications Office of the European Union, 2011.
- [3] 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*, pages 3–11. Publications Office of the European Union, 2012.
- [4] M. Frosterus, J. Tuominen, and E. Hyvönen. 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*, pages 115–124. IOS Press, 2014.
- [5] T. Heath and C. Bizer. *Linked Data: Evolving the Web into a Global Data Space (1st edition)*. Morgan & Claypool, Palo Alto, California, 2011.
- [6] A. Hietanen. Free access to legislation in Finland: Principles, practices and prospects. In M. Ragona, editor, *Law via the Internet. Free Access – Quality of Information – Effectiveness of Rights*. European Press Academic Publishing, Florence, 2009.
- [7] R. Hoekstra. The MetaLex Document Server legal documents as versioned linked data. In *Proceedings of the ISWC 2011, Bonn, Germany*, pages 128–143. Springer-Verlag, 2011.
- [8] E. Hyvönen. Cultural heritage linked data on the semantic web: Three case studies using the Sampo model, 2016. Invited talk: VIII Encounter of Documentation Centres of Contemporary Art: Open Linked Data and Integral Management of Information in Cultural Centres Artium, Vitoria-Gasteiz, Spain, October 19-20, 2016.
- [9] E. Hyvönen, J. Tuominen, M. Alonen, and E. Mäkelä. Linked Data Finland: A 7-star model and platform for publishing and re-using linked datasets. In V. Presutti, E. Blomqvist, R. Troncy, H. Sack, I. Papadakis, and A. Tordai, editors, *The Semantic Web: ESWC 2014 Satellite Events. ESWC 2014*, pages 226–230. Springer-Verlag, May 2014.
- [10] A. Oksanen, M. Tamper, J. Tuominen, A. Hietanen, and E. Hyvönen. Anoppi: A pseudonymization service for Finnish court documents, 2019. Submitted article under evaluation.
- [11] A. Oksanen, J. Tuominen, E. Mäkelä, M. Tamper, A. Hietanen, and E. Hyvönen. Semantic Finlex: Transforming, publishing, and using Finnish legislation and case law as linked open data on the web. In G. Peruginelli and S. Faro, editors, *Knowledge of the Law in the Big Data Age*, volume 317 of *Frontiers in Artificial Intelligence and Applications*, pages 212–228. IOS Press, 2019.
- [12] S. Sarsa. Information retrieval with finnish case law embeddings. Master’s thesis, University of Helsinki, Dept. of Computer Science, 2019.
- [13] S. Sarsa and E. Hyvönen. Searching case law judgements by using other judgements as a query, 2019. Submitted article under evaluation.
- [14] M. Tamper, A. Oksanen, J. Tuominen, A. Hietanen, and E. Hyvönen. Automatic annotation service: Utilizing a named entity linking tool in legal domain, 2019. Submitted article under evaluation.
- [15] D. Tunkelang. *Faceted search*. Morgan & Claypool Publishers, CA, USA, 2009.

*September 2019*

- [16] A. Van Aggelen, L. Hollink, M. Kemman, M. Kleppe, and H. Beunders. The debates of the European parliament as linked open data. *Semantic Web*, 8(2):271–281, 2017.