LawSampo Portal and Data Service for Publishing and Using Legislation and Case Law as Linked Open Data on the Semantic Web

Eero Hyvönen\textsuperscript{1,2}, Minna Tamper\textsuperscript{1,2}, Esko Ikkala\textsuperscript{1}, Mikko Koho\textsuperscript{2}, Rafael Leal\textsuperscript{2}, Joonas Kesäniemi\textsuperscript{1}, Arttu Oksanen\textsuperscript{1,3}, Jouni Tuominen\textsuperscript{1,2}, and Aki Hietanen\textsuperscript{4}

\textsuperscript{1} Semantic Computing Research Group (SeCo), Aalto University, Finland
\textsuperscript{2} HELDIG – Helsinki Centre for Digital Humanities, University of Helsinki, Finland
\textsuperscript{3} Edita Publishing Ltd, Finland
\textsuperscript{4} Ministry of Justice, Finland

Abstract. This paper argues for the idea of publishing legislation and case law as Linked Open Data (LOD) on the Semantic Web, to cater several user groups, including the general public, legislators, lawyers, researchers of legal informatics, and application developers. To support the argument, the proof-of-concept system LawSampo – Finnish Legislation and Case Law on the Semantic Web is introduced, including a semantic portal and a LOD service. The system is a new application of the more general Sampo Model for LOD publishing. The main novelty of LawSampo is the provision of heterogenous distributed legal data through multiple intelligent interfaces for searching and exploring the data and for data-analysis in legal informatics.

Keywords: Linked data · Case law · Legislation · Semantic portal

1 Introduction

Legislation and case law are widely published online by governments to make jurisdiction transparent and freely accessible to the public, organizations, and lawyers \cite{27}. The Web provides a promising medium for publishing such big data for the masses. There are, e.g., lots of private and public national level portals, such as legislation.gov.uk for the legislation for the UK, Scotland, Wales, and Northern Ireland\textsuperscript{5} and EU level systems, such as the EU Cellar\textsuperscript{6} and the ECLI Search Engine\textsuperscript{7} for the case law. Millions of judicial decisions are available in digital form in databases, and more and more are being published on the Web, too. Several efforts to standardize formats for representing legal texts are underway, such LegalXML formats\textsuperscript{8} and Akoma Ntoso\textsuperscript{9} for advancing legal data.

\textsuperscript{5} https://www.legislation.gov.uk
\textsuperscript{7} https://e-justice.europa.eu/content_ecli_search_engine-430-en.do
\textsuperscript{8} http://www.legalxml.org
\textsuperscript{9} http://www.akomantoso.org
exchange. ELI (European Legislation Identifier) and ECLI (European Case Law Identifier) standards define common identifier and metadata models for legislative and case law documents are used more and more widely.

However, most of the legal documents are available only as texts for the humans to read with little metadata, which makes them hard to use in applications of legal informatics, e.g., in computational law. To address the problem, this paper argues that legislation and case law should be published and used as Linked Open Data (LOD) on the Semantic Web, following the widely accepted FAIR principles for scientific data management and stewardship: the data should be Findable, Accessible, Interoperable, and Re-usable.

To support the argument, a case study based on Finnish legislation and case law is presented as the system LAWSAMPO – Finnish Legislation and Case Law on the Semantic Web that consists of a LOD service and a semantic portal based on it. The system is targeted for several user groups and purposes: 1) Citizens, for making the law more transparent and easy to access. 2) Legislators, to support legislation work by providing better and more data. 3) Researchers, for analyzing court decision data and legislation. 4) Lawyers, to support decision making by better and more data. 5) Developers, for developing applications and business cases using LOD.

LAWSAMPO is based on the more general Sampo Model for collaborative LOD publication that has been applied to a series of portals in Digital Humanities. We show that the model can be applied also to legal document publication.

In the following, we first describe the data underlying LAWSAMPO. After this the Sampo Model and its application to the LAWSAMPO portal and data service are described. In conclusion, related works are discussed and contributions of the paper are summarized with next steps ahead.

2 LawSampo Linked Data

Primary Data Finnish legislation and case law decisions have been published as web documents since 1997 in the Finlex Data Bank. Although this service is widely used, it does not provide machine-readable legal information as open data, on top of which web services, data analyses, and applications could be built. To facilitate this, we published a selection of Finlex data as the Semantic Finlex LOD service that currently contains ca. 28 million triples. In LAWSAMPO this data has been transformed into a simplified form suitable for the semantic portal and enriched by data linking and knowledge extraction techniques.

Data Model LAWSAMPO represents legislation and case law using a simple data model. Its main classes are shown in Table 1 with the number of instances

10 https://en.wikipedia.org/wiki/Legal_informatics
11 https://law.stanford.edu/2021/03/10/what-is-computational-law/
12 https://www.go-fair.org/fair-principles/
13 https://seco.cs.aalto.fi/applications/sampo/
14 http://www.finlex.fi
and descriptions for each class. The legislation data consists of statutes and their sections, whereas the case law data includes court decisions with language versions. Metadata about the instances is given with various properties, mostly aligned with DCMI Metadata Terms. The data model schema is available as an OWL ontology at the namespace URI \[\text{http://ldf.fi/schema/lawsampo/}\].

### Table 1. The main classes of LawSampo

<table>
<thead>
<tr>
<th>Class</th>
<th>Instances</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>:Statute</td>
<td>3725</td>
<td>A statute in consolidated legislation</td>
</tr>
<tr>
<td>:Section</td>
<td>82 145</td>
<td>A section of a statute in consolidated legislation</td>
</tr>
<tr>
<td>:Judgment</td>
<td>15 596</td>
<td>A court decision</td>
</tr>
<tr>
<td>:Expression</td>
<td>21 929</td>
<td>A representation of a court decision</td>
</tr>
<tr>
<td>:TermReference</td>
<td>1 040 537</td>
<td>A term referenced in the text</td>
</tr>
<tr>
<td>:SituationCategory</td>
<td>12</td>
<td>A life situation category of a document</td>
</tr>
<tr>
<td>:SubjectCategory</td>
<td>13 315</td>
<td>A subject keyword of a document</td>
</tr>
</tbody>
</table>

**Data Transformation** The LawSampo data transformation process is presented in Fig. 1. Semantic Finlex data is first transformed and filtered with SPARQL Construct queries. The Legislation RDF data contains only the latest versions of the consolidated legislation. Next, keyword extraction and document classification is employed to the textual contents to link them to corresponding subject keywords and life situations, respectively. After this, the entities are further linked to facet ontologies of time and EU legislation—The LawSampo portal is based on faceted search. The third step involves applying Named Entity Linking to the textual contents of Legislation and Case Law RDF. The facet ontologies are transformed from CSV format into RDF.

**Fig. 1.** LawSampo data transformation process from Semantic Finlex

\[\text{https://www.dublincore.org/specifications/dublin-core/dcmi-terms/}\]
Internal linking The data was inked internally to improve the references to other documents. The links to legal documents needed more processing as the statutes for instance may refer to more concrete part of the statute in a specific version. Unfortunately, the Semantic Finlex data is not complete and requires some human interpretation. A challenge in the data is that the court decisions only have the judgment date but not the dates for the events that are under investigation in the document. However, the judgment is based on legislation that was valid during the time of the events that are being evaluated. Therefore, the linking from decisions to statutes cannot be done to a specific statute version but only to the current consolidated version of it that doesn't take into account what version of the statute was in force at the time of the judged event.

External linking The LawSampo dataset has also been linked to external data sources including EU Cellar and the Finlex service. The links to EU Cellar have been extracted from the original Finlex statute documents [26] and as part of the transformation process added into the LawSampo data with their descriptive texts. The court decisions were linked to the Finlex service to show the original source data.

Terminological linking The legal terms occurring in the texts were linked to their explanations in order to make the texts more readable to the layman by a contextual reader [23]. The Nelli [36] and ARPA tools [22] are used to identify term instances in legal documents and link them to vocabularies with terminological definitions: the Combined Legal Concept Ontology [8], Finnish DBpedia, and the Helsinki Term Bank for the Arts and Sciences[16]. A term instance of a document contains information about its literal representation, links to its definitions in the vocabularies and the document, and a count telling how many times the term in mentioned in the document. The count information is needed for generating tag clouds summarizing the contents of the documents. The instances are represented in RDF and added to the LawSampo dataset.

Keyword extraction and document classification The subject indexing tool Annif [17][35], developed by the National Library of Finland, is used to perform keyword extraction for all the legislative documents. Annif is capable of using different algorithms in order to return suggested keywords and their respective weights for a given input text. The developers also provide a REST API containing various pre-trained projects that combine different algorithms. LawSampo uses the yso-fi pre-trained project, which integrates TF-IDF, Maui and Parabel. The first two are lexical algorithms, so they directly match terms to a vocabulary, whereas Parabel uses an associative approach that is able to also find indirect correlations between words [35]. This mix provides results that are not only grounded on the text of the documents but also able to extrapolate their specific wording. Yso-fi is trained on bibliographical metadata from Finnish museums, archives and libraries. Since the training data is labeled with
terms from the General Finnish Ontology (YSO)\(^{18}\), the API returns keywords identified by unique YSO URIs.

A zero-shot classification system based on the extracted keywords\(^ {21}\) is also used. It works by first transforming the documents into vectorial representations via the word-embedding algorithm fastText\(^ {2}\), using a pre-trained Finnish language model offered by the fastText developers\(^ {9}\). The document representations are then calculated as the average embedding of their respective keywords. A similar treatment involving Annif and fastText is given to a list of category labels representing different life situations, such as *asuminen*, *käsintö* ‘housing, real estate’, *ihmisoikeudet*, *perusoikeudet* ‘human rights, basic rights’ or *omaisuus*, *kaupankäynti*, *kuluttajansuoja* ‘property, commerce, consumer protection’. This results in vectorial representations for each category as well. The classification is then carried out by comparing the document vectors with the category vectors via cosine similarity. Each document is assigned the 5 best-fitting categories whose weight is within 95% of their top category.

Keywords and categories are used in two different ways in LAWSAMPO: they are used to label the respective legislative documents as metadata (respectively as *subject keywords* and *life situation/topic*), and they also form the basis for a semantic search system which will be explained later in this paper.

**Linked Data Service** The LAWSAMPO data service adopts the 5-star Linked Data model\(^ {19}\) extended with two more stars, as suggested in the Linked Data Finland model and platform\(^ {17}\). The 6th star is obtained by providing the dataset schemas and documenting them. The LAWSAMPO schema can be downloaded from the service\(^ {20}\) and the data model is documented using the LODE service\(^ {21}\). The 7th star is achieved by validating the data against the documented schemas to prevent errors in the published data. LAWSAMPO attempts to obtain the 7th star by applying different means of combing out errors in the data within the data conversion process. The LAWSAMPO data model and its integrity constraints are presented in a machine-processable format using the ShEx Shape Expressions language\(^ {22}\)\(^ {38}\). We have made initial validation experiments with the PyShEx\(^ {23}\) validator. Based on the experiments, we have identified errors both in the schema and the data. We plan a full-scale ShEx validation phase integrated in the data conversion and publication process to spot and report errors in the dataset.

The Linked Data service is powered by the Linked Data Finland\(^ {24}\) publishing platform that along with a variety of different datasets provides tools and services to facilitate publishing and re-using Linked Data. All URIs are dereferenceable and support content negotiation by using HTTP 303 redirects. The

\(^{18}\) [https://finto.fi/yso/en/](https://finto.fi/yso/en/)

\(^{19}\) [https://www.w3.org/DesignIssues/LinkedData.html](https://www.w3.org/DesignIssues/LinkedData.html)

\(^{20}\) [https://www.ldf.fi/dataset/lawsampo](https://www.ldf.fi/dataset/lawsampo)

\(^{21}\) [https://essepuntato.it/ldode/](https://essepuntato.it/ldode/)

\(^{22}\) [https://shex.io](https://shex.io)

\(^{23}\) [https://github.com/hsolbrig/PyShEx](https://github.com/hsolbrig/PyShEx)

\(^{24}\) [http://ldf.fi](http://ldf.fi)
data is available as an open SPARQL endpoint. As the underlying triplestore, Apache Jena Fuseki is used as a Docker container, which allows efficient provisioning of resources (CPU, memory), portability, and scaling. Varnish Cache web application accelerator is used for routing URIs, content negotiation, and caching.

3 Applying Sampo Model to LawSampo

Sampo Model is an informal collection of six principles for LOD publishing and designing semantic portals, supported by software tools and an ontology and data infrastructure. In the following, the principles are briefly introduced and their application to LawSampo is described. The principles concern only publishing data, not maintaining linked data directly. It is assumed that there is a separate repeatable and automated pipeline, like in Fig. 1, that creates the linked data in the SPARQL endpoint. If the source data or facet ontologies are updated, the linked data and linkage between resources are recreated to handle change propagation needs.

P1. Support collaborative data creation and publishing. The model is based on the idea of collaborative content creation. The data is aggregated from local data publishers' silos into a global service, based on a shared ontology and publishing infrastructure. The local data are harmonized and enriched with each other by linking and reasoning, based on Semantic Web standards.

In our case, the data and infrastructure components are aggregated from, e.g., the Ministry of Justice, different Finnish courts, National Library of Finland, Universities, EU Cellar, DBpedia, and Term Bank of Finland.

P2. Use a shared open ontology infrastructure. A shared ontology infrastructure is used for aligning the data of the publishers, including mutually aligned shared domain ontologies (thesauri) and metadata schemas. When the publishers provide metadata about their contents, the data is linked and enriched with each other through the infrastructure and forms a knowledge graph.

The national FinnONTO ontology infrastructure is used in LawSampo as the basis with additional components from the data providers above.

P3. Support data analysis and knowledge discovery in addition to data exploration. The model aims, as discussed in [15], not only at data publishing with search and data exploration, but also to data analysis and knowledge discovery with seamlessly integrated tooling for finding, analysing, and even solving research problems in interactive ways, based on AI techniques.

Application of P3 to LawSampo is explained in the two next sections.

P4. Provide multiple perspectives to the same data Sampo model fosters the idea that on top of the data service different thematic application perspectives to the data can be created by re-using the data service, without

25 https://ldf.fi/lawsampo/sparql
26 https://jena.apache.org/documentation/fuseki2/
27 https://varnish-cache.org
28 https://www.w3.org/standards/semanticweb/
modifying the data, which is typically costly. The perspectives are provided on
the landing page of the Sampo system, and enrich each other by data linking.
By selecting one of them the corresponding application is opened.

Fig. 2. LAWAMPO landing page with application perspectives

The landing page of the LAWAMPO portal depicted in Fig. 2 offers four
different application perspectives: Statutes, Sections, Case Law, and Contextual
search. In addition, an external application Case Law Finder can be used. Using
the perspectives is demonstrated in the next two section in more detail.

P5. Standardize portal usage by a simple filter-analyze two-step cy-
cle. The application perspectives can be used by a two-step cycle for research—
the idea is to standardize and make easier the way a semantic portal is used:
Firstly, the focus of interest, the target group, is filtered out using faceted se-
monic search [11,39,40]. Secondly, the target group is visualized or analyzed by
using ready-to-use data analysis tools of the application perspectives.

Principle P5 is used in all four application perspectives of LAWAMPO. For
example, in LAWAMPO a group of judicial decisions can be filtered out first by
selections made in the facets and then visualized on a timeline.

P6. Make clear distinction between the LOD service and the user
interface (UI). The Sampo Model argues for separating the underlying Linked
Data service completely from the user interface via a SPARQL API. The rationale
for this is to simplify the portal architecture and, especially, to be able to open
the data for data analysis research using, e.g., Yasgu29 and Python scripting
in Google Colab30 and Jupyter31 notebooks. For developing Sampo portal user
interfaces (UI) cost-efficiently, the Sampo-UI framework 18 is in use.

29 https://yasgui.triply.cc
30 https://colab.research.google.com/notebooks/intro.ipynb
31 https://jupyter.org
Using the LawSampo portal and the underlying data service independently is explained in the two following sections, respectively.

The Sampo model has evolved gradually over time in 2002–2021 through lessons learned, leading to the “Sampo series” of a dozen of semantic portals. They have had up to millions of users on the Semantic Web, suggesting usability of the model and its tools in practical applications.

4 Using LawSampo Portal

This section illustrates how the application perspectives of LawSampo (cf. Fig. 2), based on the Sampo model and Sampo-UI, is used in practise.

**Statutes Perspective** By clicking on the Statutes perspective box, a faceted search interface for searching and browsing statutes is opened. The facets on the left include document type (with seven subtypes), statute type, year, and related EU regulation. 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 such as the referred EU regulations linked to EU Cellar or other documents from Semantic Finlex referring to it. The LawSampo application utilizes enriched data and shows annotated statute documents. The annotations are highlighted in the text and by hovering over the annotation, the user can see the explanation to the term and links to external portals such as Wikipedia or the Helsinki Term Bank for the Arts and Sciences to learn more about the term. The terms are also used to create tag cloud visualizations to give the user an idea what the text is about.

**Sections Perspective** The Sections perspective operates in the same way, but here it is possible to search and explore consolidated legislation on a more focused section level.

**Case Law Perspective** 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 judgment. Similarly to statutes, the case law view shows the results based on the facet selections as a list for the user. From this point on the user can view the court decision details at its “homepage”. Similarly to statutes, the court decision’s page contains the annotated text document, a tag cloud, and more related information about it. The court decisions also have been enriched in the portal with related case law documents based on Semantic Finlex Case Law Finder. It retrieves documents that are textually similar to the selected document and the results are listed in the table tab.

In addition to the court decision listing and homepages, the user’s choices also influence the other tabs in the case law perspective such as the statistics such as the facet’s pie charts or the by year bar charts for the court decisions.

---

32 See [https://seco.cs.aalto.fi/applications/sampo/](https://seco.cs.aalto.fi/applications/sampo/) for a list of “Sampo portals” and further information.
By selecting a value from a facet, all other facets and results update and the distribution of court decisions by year or by facet (e.g., by court, in the court facet) show the updated results. With these statistical tools the user can view and study the case law data. Fig. 3 is a screenshot from the Case Law perspective’s plot depicting the number of court decisions by year. The plot shows the number of court decisions with the judgment date information on a timeline.

![Fig. 3. Number of court decisions in 1990–2019 in Case Law Perspective](image)

The Case Law perspective also enables the user to export the faceted search query into the Yasgui tool on a separate tab EXPORT. The data can be explored further by editing the SPARQL query there, and the results can be downloaded for further study, for example into a spreadsheet program in CSV form.

**Contextual Search Perspective** The fourth perspective, named Contextual Search, allows for searching legal documents based on the end user’s life situation at hand (e.g., divorce). This search system employs the Relevance Feedback Search (RFBS) paradigm [1, Ch. 5], which works by refining the search parameters iteratively, with input from the user, in order to improve its results. In LAWASAMPO this is done by offering the user keyword and category suggestions based on the results of the previous search round: by activating or deactivating these suggestions, the user gradually redirects the query towards a more satisfying result. This kind of search is argued to be useful in situations where it is difficult for the user to formulate a traditional search query.

Our RFBS framework, described in [21] in more detail, depends on four elements: text query, category, positive keywords, and negative keywords. It builds upon the document and category representations outlined in Section 2. The search is performed by building a vectorial representation for the query and comparing it with the document representations via cosine similarity. This query vector is built by using the keyword extraction→word embeddings pipeline, and refined via keywords –both positive and negative– and category choices made by the user. This Contextual Search perspective splits the search processes into
two phases. In the initial phase, the user can start the search by either typing in a free-form text query or by selecting one of the pre-defined categories. After the initial search, the system switches into iterative mode, where the user can only update the filters through user-provided feedback. If the user starts the search with a textual query, (s)he is presented with both suggested categories and keywords as feedback options. Since the system only allows for one category to be active at a time, suggested categories are disabled in case they are used as the initial search method.

The search continues in an iterative fashion and new results are fetched whenever the user changes the state of the active filters, i.e., selected category or keywords. LAWSAMPO shows the top three categories and top 20 keywords for each iteration. The user can select multiple keywords per iteration as either positive or negative. Negative keywords are used to exclude documents associated with the keywords in question from the search result. The iteration continues until the user removes the initial search criterion, which clears all the filters and returns the search interface to its initial state.

**Similarity-based Case Law Search** In addition to the application perspectives above, the landing page also provides a link to yet another application of searching court decisions [34]. The idea of this application is that the user is able to input a text document as a query either by uploading a file or by writing text directly into the form. Supported file formats for uploading documents are plain text, XML, PDF, and with Tesseract OCR, image formats, such as JPG and PNG. The application also allows the user to choose the algorithm that ranks the documents by using the drop-down menu on the left-bottom in.

This application can be helpful when, for example, one has got a court decision and one is willing to see whether the verdict is fair in comparison with other similar cases. Hopefully, such possibility could lessen unnecessary appeals to higher courts in the future.

This prototype application is based on the assumption that textually similar cases are relevant for the information need. Several methods for finding similar cases were tested when implementing this application including TF-IDF, Latent Dirichlet Allocation (LDA), Word2Vec, and Doc2vec [33,34].

5 Using LawSampo Data Service for Data Analysis

In addition to the ready-to-use statistical applications integrated into the LawSampo portal, the underlying SPARQL endpoint can be used for querying, analyzing, and visualizing the data in flexible ways using HTTP protocol and external tools. For example, the visualization in Fig. 4 shows how the number of court decisions of different kinds in the data change in time during 1980–2020. This graph has been created using Google Colab Jupyter documents and Python scripting. The figure shows how the number of civilian and criminal cases decreases in time, and that the cases of the Administrative court dominate the dataset. There is also a number of court decisions without type of the matter described. Also their number is decreasing but slowly.
6 Discussion

Related Work Our work on legal Linked Data services was influenced by the MetaLex Document Server \[33\] \[10\] that publishes Dutch legislation using the CEN Metalex XML and ontology standards. Other Metalex ontology based implementations include legislation.gov.uk and Nomothesia \[34\] that also implements ELI-compliant identifiers. Various ELI implementations and prototypes have also been implemented in existing legal information portals nationally, e.g., in Luxembourg \[35\], France \[36\], and Norway \[37\]. Many countries already produce ECLI-compliant case law documents to be indexed by the ECLI search engine. A prominent example of publishing EU Law and publications as linked data is the EU Cellar system. Previous related works in the U.S. include, e.g., the Legal Linked Data project aiming at enhanced access to product regulatory information \[3\].

Lots of companies provide legal services for searching and exploring legislation and case law, and even Google Scholar has a specific search application for cases in the various courts of the states \[38\].

Most zero-shot classification methods work by training a classifier and then adapting it to a new set of categories \[6,30,42\], while \[44\] also integrates a knowledge graph into their algorithm. Among unsupervised classification models, \[32\] use skip-gram word embeddings to calculate the semantic similarity between a label and the given documents, while \[43\] treats zero-shot as an entailment problem. Various methods exist for relevance feedback search \[1,31\], which has also been previously employed in various web search settings \[37,28,19\].

\[33\] http://doc.metalex.eu
\[34\] http://legislation.di.uoa.gr
\[35\] http://legilux.public.lu/editorial/eli
\[36\] http://www.eli.fr/en/constructionURI.html
\[37\] http://lovdata.no/eli
\[38\] https://scholar.google.com/scholar_courts
Contributions and Challenges This paper applied the Sampo Model, developed originally for Digital Humanities research, to a novel use case in legal informatics. LAWSAMPO aims to advance the state-of-art in legal document publishing by language technology and a Linked Open Data-based approach. Legislation and case law data are provided to multiple end user groups and purposes through intelligent user interfaces, such as semantic faceted search, relevance feedback search, data analytic views, and document similarity-based search. The documents are automatically enriched with contextual linked data, and the end user is provided with ready-to-use data-analytic tooling for analyzing the documents and their relations.

However, extracting and linking references of legal documents requires still more work. The references to legal documents can be made in various ways and the labels we currently have in our databases are not enough to identify all the ways in which the references are made in texts. There are references made using the official names or nick names that exist in the Finlex database, but some references are made with unidentified acronyms or by twisting the order of words in the names, which may produce unidentifiable wordings for different statute names. It would be much easier to add metadata about related documents manually when indexing the documents than trying to extract the links from unstructured texts afterwards. The biggest semantic challenge we encountered in our work was that the statutes are not stable but their sections are dynamically added, cancelled, and modified in time by other statutes. In the Finnish legislation system, systematic time series of consolidated versions of legislation are not available, but only the initial versions of the statuses and series of changes made to them afterwards. The court decisions are, however, always made based on the legislation in force at the time of the judicial offense, which makes the linking between legislation and case law difficult. LAWSAMPO has access only to the latest versions of manually consolidated statues available in Finlex, and the problem of finding out how the statutes may have changed in time is left to the end user. In many cases, the court decision does not even tell when the judged offence was made, but may only refer to a lower court decision where the date information may be available. From data publishing point of view this information should be added to the decision metadata already at the courts.

In spite of the challenges and complexities of the underlying data, we are confident that that proposed LOD approach is feasible and usable in practice, and plan to make the LAWSamPO prototype publicly available. However, before that a piloting phase with a smaller user group and an evaluation of the usability of the system will be made.

Future Directions In the future, we plan to expand the related enriching datasets to include, e.g., related parliamentary documents and discussions [39] in the spirit of LinkedEP [41]. In order to be able to publish more legal documents in a cost-efficient way, we also work on semi-automatic pseudonymization of court judgments [24] and automatic annotation of legal documents [39].

Acknowledgments We thank Tiina Husso at Ministry of Justice and Risto Talo and Jari Linhala of Edita Publishing Ltd. for collaboration. Our work was funded by the Ministry of Finance; CSC – IT Center for Science provided computational resources for the work.

References

21. Leal, R., Kesäniemi, J., Koho, M., Hyvönen, E.: Relevance feedback search based on automatic annotation and classification of texts (March 2021), submitted