

Lainsäädännön ja oikeuskäytännön mallintaminen ja julkaiseminen linkitettynä avoimena datana

Arttu Oksanen

Sähkötekniikan korkeakoulu

Diplomityö, joka on jätetty opinnäytteenä tarkastettavaksi
diplomi-insinöörin tutkintoa varten Espoossa 10.10.2016.

Työn valvoja:

Prof. Eero Hyvönen

Työn ohjaajat:

FM Jouni Tuominen

TkT Eetu Mäkelä

Tekijä: Arttu Oksanen

Työn nimi: Lainsäädännön ja oikeuskäytännön mallintaminen ja julkaiseminen linkitettyinä avoimena datana

Päivämäärä: 10.10.2016

Kieli: Suomi

Sivumäärä: 7+62

Tietotekniikan laitos

Professuuri: Mediatekniikka

Työn valvoja: Prof. Eero Hyvönen

Työn ohjaajat: FM Jouni Tuominen, TkT Eetu Mäkelä

Viranomaiset julkaisevat oikeudellista tietoa verkossa avoimesti usein ihmisluettavissa PDF- ja HTML-muodoissa. Kuitenkin tiedonhaun tehostamiseksi ja tiedon ymmärtämisen helpottamiseksi tarvitaan älykkäitä palveluita sekä älykästä tietomallinnusta. Lisäksi lainsäädännön kansainvälistyessä eri organisaatioilla on tarve edistää oikeudellista tiedonvaihtoa yli kansallisten rajojen, mikä edellyttää aineistojen esitystavan yhtenäistämistä. Tässä diplomityössä tutkitaan, miten linkitetyn datan teknologioilla voidaan mallintaa ja julkaista lainsäädäntö sekä oikeuskäytäntö siten, että julkaisu palvelee laajasti eri oikeudellisen tiedon käyttötapauksia. Työ sisälsi RDF-tietomallien ja datamuunnoksen kehittämisen, datan rikastamisen sekä ohjelmointirajapintojen ja sovellusprototyyppien toteuttamisen. Lopputuloksena syntyi Semanttinen Finlex -palvelu, jossa Suomen lainsäädäntö sekä korkeimman oikeuden ja korkeimman hallinto-oikeuden ratkaisut on julkaistu keskeisiltä osin linkitettyinä avoimena datana noudattaen eurooppalaisia tunniste- ja metatietostandardeja.

Avainsanat: lainsäädäntö, oikeuskäytäntö, linkitetyn datan julkaiseminen, tietomallinnus, yhteentoimivuus, datan linkitys

Author: Arttu Oksanen

Title: Modeling and Publishing Legislation and Case Law as Linked Open Data

Date: 10.10.2016

Language: Finnish

Number of pages: 7+62

Department of Computer Science

Professorship: Media Technology

Supervisor: Prof. Eero Hyvönen

Advisors: M.Sc. Jouni Tuominen, D.Sc. (Tech.) Eetu Mäkelä

Governments publish legal information openly online usually in the form of human readable PDF and HTML documents. However, to facilitate search and understanding of legal information intelligent services and intelligent data modeling are required. Moreover, as law becomes international the authorities have a growing need for enhancing information exchange which calls for standardized ways of presenting information. This thesis examines how Linked Data technologies can be applied to modeling and publishing legislation and case law to support different use cases extensively. The work involved development of RDF data models and data conversion, data enriching and implementation of application programming interfaces and application prototypes. The end result is Semantic Finlex, a legal Linked Open Data service, that hosts central part of the Finnish legislation, judgments of the Supreme Court and judgments of the Supreme Administrative Court published using European standards for identifiers and metadata schemas.

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

Esipuhe

Haluan kiittää työni valvojaa professori Eero Hyvöstä sekä ohjaajiani Jouni Tuomista ja Eetu Mäkelää saamistani hyvistä neuvoista ja rakentavasta palautteesta. Lisäksi kiitän kaikkia muita Semanttisen laskennan tutkimusryhmän (SeCo) jäseniä hyvästä työilmapiiristä sekä mielenkiintoisista keskusteluista.

Kiitän myös Edita Publishing Oy:n Risto Taloa sekä Jari Linhalaa hyvästä yhteistyöstä sekä asiantuntevista vastauksista esittämiini kysymyksiin oikeudellisista aineistoista. Kiitos kuuluu myös European Legislation Identifier -työryhmän (ELI) Thomas Francartille neuvoista ELI:n käyttöön ottamisessa.

Ennen kaikkea tunnustus kuuluu oikeusministeriön Aki Hietaselle, joka mahdollisti tämän diplomityön tekemisen.

Espoossa, 10.10.2016

Arttu Oksanen

Sisällysluettelo

Tiivistelmä	ii
Tiivistelmä (englanniksi)	iii
Esipuhe	iv
Sisällysluettelo	v
Lyhenteet	vii
1 Johdanto	1
2 Oikeudelliset aineistot	3
2.1 Lainsäädäntöprosessi Suomessa	3
2.2 Säädösten rakenne	3
2.3 Lainsäädännölliset aineistot	5
2.4 Oikeuden ratkaisut	6
3 Oikeudellisen tiedon esittäminen verkossa	7
3.1 Käyttötapaukset	7
3.2 Linkitetty lakitieto	8
3.3 Oikeudelliset ontologiat	9
3.4 ELI-standardi lainsäädännön esittämiseksi	10
3.5 ECLI-standardi oikeuskäytännön esittämiseksi	12
3.6 Oikeudelliset datapalvelut	13
4 Lainsäädännön tietomalli	15
4.1 Aineistojen XML-rakenteet	15
4.2 Säädösten URI-tunnisteet	17
4.3 FRBR-kuvailumallin soveltaminen	18
4.4 Asiakirjojen rakenteellinen RDF-mallinnus	25
4.5 Säädösten kontekstuaalinen metatieto	26
4.6 Säädösten semanttinen metatieto	27
5 Oikeuden ratkaisujen tietomalli	29
5.1 Ratkaisujen URI-tunnisteet	29
5.2 Tietomallin luokat ja ominaisuudet	29
5.3 Nimien tunnistaminen ratkaisun tekstistä	33
5.4 Asiasanastot	34
6 Datamuunnos ja tiedonhallinta	35
6.1 XML-RDF-muunnin	35
6.2 Aineiston päivitys	36
6.3 Validointi	37
6.4 Datamuunnoksen haasteet ja puutteet	38

7	Datajulkaisu	42
7.1	Kuuden tähden datajulkaisu	42
7.2	Käyttöehdot	43
7.3	Ohjelmointirajapinnat	44
8	Sovellukset	46
8.1	Linkitetyn datan selaus ja semanttinen suosittelu	46
8.2	Haun tehostus	49
9	Johtopäätökset	51
9.1	Vastaukset tutkimuskysymyksiin	51
9.2	Vertailu olemassa oleviin datapalveluihin	52
9.3	Jatkokehitys	52
10	Yhteenveto	54

Lyhenteet

CDM	Common Data Model
ECLI	European Case Law Identifier
EEJP	European e-Justice Portal
ELI	European Legislation Identifier
FRBR	Functional Requirements for Bibliographic Records
HTTP	Hypertext Transfer Protocol
IFLA	International Federation of Library Associations and Institutions
JSON-LD	JavaScript Object Notation for Linked Data
KHO	Korkein hallinto-oikeus
KKO	Korkein oikeus
NEL	Named Entity Linking
NER	Named Entity Recognition
NLP	Natural Language Processing
OWL	Web Ontology Language
PTJ	Valtioneuvoston päätöksentekojärjestelmä
REST	Representational State Transfer
RDF	Resource Description Framework
RDFS	RDF Schema
SeCo	Semanttisen laskennan tutkimusryhmä
SFCL	Semantic Finlex Case Law
SFL	Semantic Finlex Legislation
SHACL	Shapes Constraint Language
SKOS	Simple Knowledge Organization System
SPARQL	SPARQL Protocol and RDF Query Language
SPIN	SPARQL Inferencing Notation
URI	Uniform Resource Identifier
W3C	World Wide Web Consortium
WWW	World Wide Web

1 Johdanto

Lainsäädännön ja muun oikeudellisen tiedon määrä on jatkuvassa kasvussa. Sekä yksityishenkilöt että yritykset ovat koko ajan monimutkaistuvan ja uusia elämän osa-alueita koskehtavan sääntelyn vaikutuspiirissä. Lainsäädäntö ja oikeuskäytäntö myös kansainvälistyvät, mikä luo tarpeen tarjota oikeudellista tietoa paitsi kansalaisille, myös yli kansallisten rajojen. Pelkkä tiedon tarjoaminen avoimesti ei kuitenkaan riitä vastaamaan käytännön haasteisiin, joita aineistojen käyttäjät kohtaavat sekä etsiessään oikeudellista tietoa valtavista aineistoista että yrittäessään ymmärtää sitä.

Viime vuosina hallinnon läpinäkyvyyttä on eri valtioissa haluttu edistää julkaisemalla hallinnollisia aineistoja, kuten keskeisimmät oikeusnormit, avoimella lisenssillä koneluettavana datana. Avoimena julkaistu data mahdollistaa lisäarvoa tuottavien sovellusten ja analyysien rakentamisen dataa hyödyntäen. Erityisesti linkitetyn datan teknologiat [6] nähdään mahdollisuutena sekä tehostaa oikeudellista tiedonhakua että helpottaa oikeudellisen tiedon ymmärtämistä [11].

Suomessa lainsäädäntö ja oikeuskäytäntö ovat olleet keskeisiltä osin saatavissa ilmaiseksi ihmisluettavissa PDF- ja HTML-tiedostomuodoissa oikeusministeriön ylläpitämässä Finlex-palvelussa¹. Kustannusyhtiö Edita Publishing Oy toimittaa Finlexin sisältämät oikeudelliset aineistot. Finlex-tietopankin lisäksi tarjolla on erityisesti oikeusalan ammattilaisille suunnattuja kaupallisia verkkopalveluita, jotka tarjoavat asiakkailleen lainopillisia lisäarvopalveluja maksusta. Näitä ovat Talentum Oyj:n toimittama Suomen Laki² sekä Edita Publishing Oy:n Edilex³.

Oikeudelliset aineistot eivät kuitenkaan ole Suomessa olleet aiemmin saatavissa avoimena, rakenteellisena datana siten, että tietotekniset sovellukset voisivat helposti hyödyntää niitä avoimien rajapintojen kautta. Tässä diplomityössä esitellyssä Semanttinen Finlex -palvelussa oikeudelliset aineistot on nyt julkaistu linkitettyinä avoimena datana. Projektissa rakennettiin pysyviä URI-tunnisteita ja oikeudellisia ontologioita käyttämällä tietoinfrastrukturi kansalliselle lainsäädännölle ja oikeuskäytännölle. Linkitetyn datan teknologiat ja käyttöön otetut kansainväliset standardit mahdollistavat semanttisten yhteyksien luomisen sekä kansallisten että kansainvälisten oikeudellisten aineistojen välille.

Linked Data Finland -projektissa⁴ julkaistiin ensimmäinen versio Semanttinen Finlex -palvelusta [28] jo vuonna 2012. Julkaisu sisälsi Finlex-palvelun ajantasaisen lainsäädännön kokoelman sekä korkeimman oikeuden (KKO) ennakkopäätökset ja korkeimman hallinto-oikeuden (KHO) vuosikirjaratkaisut. Myöhemmin havaittiin, että datajulkaisun toteutukseen liittyi ongelmia. Julkaisun tietomalli ei vastannut parhaita käytäntöjä, vaan oli suureksi osaksi käännetty suoraan alkuperäisestä XML-merkintäkielestä toistaen sen rakenteita, mikä vaikeutti datan käyttöä sovelluksissa. Lisäksi aineistolle ei ollut olemassa päivityskäytäntöä, ja julkaistu lainsäädännöllinen aineisto sisälsi vain yhden version ajantasaistetusta laista.

Keväällä 2016 julkaistiin uusi, jatkokehitetty versio Semanttinen Finlex

¹<http://www.finlex.fi>

²<http://www.suomenlaki.com>

³<http://www.edilex.fi>

⁴<http://www.ldf.fi>

-datapalvelusta. Palvelun uudessa versiossa on otettu käyttöön eurooppalaiset standardit dokumenttien metatietojen mallintamiseksi. Uudessa versiossa ovat mukana sekä alkuperäiset säädökset että päivittyvä ajantasainen lainsäädäntö.

Tässä diplomityössä keskitytään tarkastelemaan Semanttinen Finlex -palvelun uuden version kehitystyötä. Työssä pyritään vastaamaan seuraaviin tutkimuskysymyksiin:

- Miten data kannattaa julkaista, jotta julkaisu palvelee mahdollisimman laajasti eri käyttötapauksia?
- Miten lainsäädännölle ja oikeustapauksille luodaan metatietomäärittelyiltään riittävän kattava, laajennettavissa oleva ja muiden aineistojen kanssa yhteentoimiva tietomalli?
- Minkäláisten rajapintojen kautta data tulee tarjota, jotta se on mahdollisimman kustannustehokkaasti hyödynnettävissä sovelluksissa?
- Miten aineistojen metadatta voidaan rikastaa automaattisesti?
- Miten aineistosta voidaan luoda yhteyksiä muihin datoihin?

Uusi versio Semanttinen Finlex -julkaisusta on kehitetty Aalto-yliopiston Semanttisen laskennan tutkimusryhmässä (SeCo) yhteistyössä oikeusministeriön ja Edita Publishing Oy:n kanssa. Kehitystyö on tehty valtiovarainministeriön erillisrahoituksella osana julkisten tietovarantojen avaamiseen sekä innovatiiviseen hyödyntämiseen tähtäävää avoimen tiedon ohjelmaa⁵.

Diplomityön sisältö jakautuu seuraavasti: luvussa 2 käydään läpi suomalainen lainsäädäntöprosessi sekä mitä julkaistavat oikeudelliset asiakirjat sisältävät. Luvussa 3 esitellään avoimen oikeudellisen datan käyttötapaukset sekä oikeudellisen tiedon esittämiseen kehitetyt laskennalliset mallit ja standardit sekä olemassa olevia oikeudellisia datapalveluita. Semanttinen Finlex -datajulkaisun lainsäädännöllisten aineistojen esittämiseen kehitetyt tietomallit käsitellään luvussa 4 ja oikeuden ratkaisujen tietomallit luvussa 5. Luvussa 6 esitellään palvelun datamuunnostyökalu sekä tiedonhallinnan haasteita. Datajulkaisu ja siihen liittyvät ohjelmointirajapinnat käsitellään luvussa 7. Dataa hyödyntävät prototyyppisovellukset esitellään luvussa 8. Luvussa 9 muotoillaan vastaukset tutkimuskysymyksiin, vertaillaan Semanttinen Finlex -palvelua olemassa oleviin oikeudellisiin datapalveluihin sekä hahmotellaan suuntaviivoja jatkokehitykselle ja tulevaisuuden tutkimukselle. Lopuksi luvussa 10 tehdään yhteenveto tutkimuksesta.

⁵<http://vm.fi/avointieto>

2 Oikeudelliset aineistot

Tässä luvussa kuvataan aluksi lyhyesti lainsäädäntöprosessin kulku Suomessa. Sen jälkeen tarkastellaan prosessissa syntyvien säädösasiakirjojen rakennetta. Lopuksi esitellään Semanttinen Finlex -palvelussa julkaistavat lainsäädännölliset aineistot ja oikeuden ratkaisut.

2.1 Lainsäädäntöprosessi Suomessa

Lain säätäminen tulee Eduskunnassa vireille joko hallituksen esityksellä tai kansanedustajan aloitteella (Perustuslaki 70§). Käytännössä valtaosa lakialoitteista syntyy hallituksen esittämänä.

Hallituksen esityksen valmistelee se ministeriö, jonka toimialaan säänneltävä asia kuuluu. Valmistelutyö ministeriöissä on monivaiheinen prosessi, johon sisältyy hallituksen esityksen esi-, perus- ja jatkovalmistelu sekä lausuntokierrokset, joissa eri sidosryhmiltä kerätään palautetta esitysluonnoksesta. Lopulta hallitus päättää hallituksen esityksen antamisesta eduskunnalle valtioneuvoston yleisistunnossa. Tämän jälkeen esitys menee eduskunnan käsittelyyn. [60]

Hallituksen esityksen käsittely käynnistyy eduskunnassa lähetekeskustelulla, jossa päätetään, mikä valiokunta ottaa esityksen käsittelyyn. Valiokunta kuulee ministeriön virkamiehiä ja asiantuntijoita ja laatii mietinnön, jossa se suosittaa täysistunnolle esityksen hyväksymistä, hylkäämistä tai hyväksymistä muutettuna. Tämän jälkeen lakiehdotus joko hyväksytään tai hylätään kaksivaiheisessa käsittelyssä täysistunnossa. [22]

Eduskunnan hyväksynnän jälkeen tasavallan presidentin on vielä vahvistettava laki ennen sen voimaantuloa (Perustuslaki 78§). Vahvistettu laki julkaistaan lopulta Finlex-palvelun sähköisessä säädöskokoelmassa⁶ PDF-dokumenttina [60].

Tasavallan presidentti, hallitus tai ministeriöt voivat lisäksi antaa lakia tarkentavia asetuksia eduskunnan säätämän lain perusteella [60]. Sekä lakeja että asetuksia kutsutaan säädöksiksi. Säädökset muodostavat arvohierarkian, jossa ylimpänä ovat lait ja sen jälkeen tulevat järjestyksessä tasavallan presidentin asetukset, valtioneuvoston asetukset, ministeriöiden asetukset ja ministeriöitä alempien viranomaisten antamat oikeussäännöt [58].

Säädös voi olla joko uusi säädös, muutossäädös tai kumoamissäädös. Muutossäädöksellä olemassa olevan säädöksen sisältöön kohdistetaan jokin muutos. Muutossäädös sisältää ainoastaan muutettavat lainkohdat, jotka voivat olla mistä tahansa kohdasta muutettavaa säädöstä. Kumoamissäädöksellä ainoastaan kumotaan olemassa oleva säädös tai osa siitä.

2.2 Säädösten rakenne

Säädösasiakirjalla on tarkkaan määritelty rakenne, joka alkaa säädösnumerolla, nimikkeellä ja johtolauseella. Kuvassa 1 on esimerkki säädöksen alusta. Säädösnumero

⁶<http://finlex.fi/fi/laki/kokoelma/>

81/2016

Julkisten alojen eläkelaki

Eduskunnan päätöksen mukaisesti säädetään:

1 luku

Lain tarkoitus ja soveltamisala

1 §

Lain tarkoitus

Tässä laissa säädetään työansioiden perusteella määräytyvästä Kevan jäsenyhteisön, valtion, evankelis-luterilaisen kirkon ja Kansaneläkelaitoksen palveluksessa olevan eläketurvasta.

2 §

Eläketurvan toimeenpano

Tämän lain mukaisesta eläketurvasta huolehtii julkisoikeudellinen Keva-niminen eläkelaitos. Kevasta säädetään Kevasta annetussa laissa (66/2016).

Kuva 1: Säädöksen säädösnumero, nimike, johtolause ja kaksi ensimmäistä pykälää (kuva: Sähköinen säädöskokoelma)

on säädöksen antamisvuotensa sisällä yksilöivä numero. Esimerkin tapauksessa säädösnumero on 81/2016, missä 2016 on antovuosi ja 81 on säädöksen yksilöivä numero antovuoden sisällä. Nimike on säädöksen nimi, joka on esimerkin tapauksessa ”Julkisten alojen eläkelaki”. Eduskunnan säätämän uuden lain johtolause on tyypillisesti sisällöltään ”Eduskunnan päätöksen mukaisesti säädetään:”, kuten esimerkissä.

Johtolauseen jälkeen säädös jatkuu pykälillä, joita voi olla yksi tai useampi. Pykälät numeroidaan ja ne voidaan otsikoida. Pykälä koostuu edelleen momenteista, jotka ovat tekstikappaleita pykälän sisällä. Jos pykälää on paljon, ne voidaan ryhmitellä luvuiksi ja luvut edelleen osiksi [58]. Pykälät voidaan myös ryhmitellä väliotsikoiden avulla. Kuvan 1 esimerkissä näkyvät lain kaksi ensimmäistä pykälää, jotka on ryhmitelty luvun 1 alle ja joista molemmat sisältävät yhden momentin.

Momentti voi edelleen koostua numeroiduista kohdista, ja kohdat pienillä aakkosilla merkityistä alakohdista. Jos momentti koostuu kohdista, on sen alussa kaksoispisteeseen päättyvä johdantokappale. Poikkeuksena ovat rangaistussäännökset, joissa johdantokappaleessa ei ole kaksoispistettä ja lisäksi momentin lopussa on vielä loppukappale. Kuvassa 2 on esimerkki rangaistussäännöksistä. Johdantokappaletta ”Joka tahallaan tai huolimattomuudesta” seuraa kaksi numeroitua kohtaa ja loppukappale ”on tuomittava puolustustarvikkeiden maastavientirikkomuksesta sakkoon”. Lisäksi esimerkin pykälä sisältää vielä toisen momentin.

Jokaisesta säädöksestä löytyy myös voimaantulosäännös, joka kertoo milloin säädöksen lainkohdat tulevat voimaan [58]. Uudessa säädöksessä voimaantulosäännös on tavallinen numeroitu pykälä, muutossäädöksissä se on erotettu muutettavista

38 §**Rangaistussäännökset**

Joka tahallaan tai huolimattomuudesta

- 1) jättää toimittamatta 33 §:n 1 momentissa tarkoitetun selvityksen tai
 - 2) rikkoo 33 §:n 2 tai 3 momentissa tarkoitettuja säännöksiä kirjanpidosta,
- on tuomittava *puolustustarvikkeiden maastavientirikkomuksesta* sakkoon.

Rangaistus puolustustarvikkeiden maastavientirikoksesta säädetään rikoslain (39/1889) 46 luvun 11 §:ssä.

Kuva 2: Esimerkki rangaistussäännöksistä (kuva: Finlex)

23 §*Voimaantulo*

Tämä laki tulee voimaan 1 päivänä tammikuuta 2016. Lain 4 ja 5 § tulevat kuitenkin voimaan vasta 1 päivänä tammikuuta 2017.

Kuva 3: Esimerkki voimaantulosäännöksestä (kuva: Sähköinen säädöskokoelma)

lainkohdista palstaviivalla. Esimerkki voimaantulosäännöksestä on esitetty kuvassa 3. Esimerkin lain 4 pykälä ja 5 pykälä tulevat voimaan eri ajankohtana kuin lain muut pykälät.

2.3 Lainsäädännölliset aineistot

PDF-muotoisia säädöksiä sisältävän sähköisen säädöskokoelman lisäksi Finlex-palvelussa julkaistaan lainsäädännöllisiä aineistoja myös HTML-muodossa. Nämä aineistot koostuvat kolmesta keskeisestä kokonaisuudesta:

- **Säädökset alkuperäisinä**⁷. Alkuperäiset säädöstekstit on esitetty HTML-muodossa ilman oikaisuja sellaisina kuin ne on julkaistu alunperin säädöskokoelmassa. Kokoelmasta löytyvät kaikki säädökset vuodesta 1987 eteenpäin ja keskeinen osa sitä vanhemmista säädöksistä – yhteensä noin 45 000 säädöstä sekä suomeksi että ruotsiksi.
- **Ajantasainen lainsäädäntö**⁸. Edita Publishing Oy toimittaa Finlex-palveluun ajantasaistettua lainsäädäntöä. Ajantasaistamisella tarkoitetaan, että säädöskokoelmassa julkaistut muutossäädökset on sisällytetty alkuperäisten säädösten tekstiin. Muutos- tai kumoamissäädöksen julkaisun yhteydessä ei siis lainsäätäjien toimesta julkaista itse muutetusta säädöksestä uutta versiota, vaan

⁷<http://finlex.fi/fi/laki/alkup/>

⁸<http://finlex.fi/fi/laki/ajantasa/>

voimassaoleva säädösteksti kootaan yhteen dokumenttiin Edita Publishing Oy:n toimitustyönä. Ajantasaisten säädösten kokoelmaan sisältyy noin 2 800 säädöstä sekä suomeksi että ruotsiksi.

- **Säädösmuutosten hakemisto**⁹. Hakemisto sisältää kustannustoimittajan käsityönä laatimia säädösten metatietoja kuten voimaantulopäivä, mahdollinen kumoamispäivä sekä oikaisutiedot. Lisäksi hakemistosta löytyvät tiedot säädöksen nojalla annetuista muista säädöksistä, säädöstä muuttaneista muutossäädöksistä sekä muutossäädösten nojalla annetuista muista säädöksistä.

Semanttinen Finlex -datajulkaisu sisältää Finlex-palvelun aineistoista alkuperäisten sekä ajantasaisten säädösten kokoelmat. Säädösmuutosten hakemisto ei kuulu julkaistavien aineistojen joukkoon tekijänoikeudellisista syistä.

2.4 Oikeuden ratkaisut

Tuomioistuinten ratkaisuihin Semanttinen Finlex -palvelussa julkaistaan Finlex-palvelusta löytyvät korkeimman oikeuden ennakkopäätökset sekä korkeimman hallinto-oikeuden vuosikirjassa julkaistut ratkaisut. Korkein oikeus toimii Suomessa ylimpänä muutoshakuasteena riita- ja rikosasioissa. Sen antamat ennakkopäätökset toimivat oikeusohjeina alempien oikeusasteiden tuomioistuimille vastaavanlaisissa oikeusriidoissa. Korkein hallinto-oikeus käsittelee pääsääntöisesti viranomaisten päätöksistä tehtyjä valituksia [61].

Korkeimman oikeuden ennakkopäätökset koostuvat yleensä neljästä osasta, jotka ovat tiivistelmä, alempien oikeuksien ratkaisut, muutoksenhaku korkeimmassa oikeudessa ja korkeimman oikeuden ratkaisu. Korkeimman hallinto-oikeuden ratkaisujen rakenne vaihtelee enemmän, eikä niihin ole koostettu sisällysluetteloita Finlex-palvelussa. Usein korkeimman hallinto-oikeuden ratkaisusta löytyy tiivistelmän lisäksi tieto päätöksestä, josta valitetaan, asian aiempi käsittely sekä korkeimman hallinto-oikeuden ratkaisu. Sekä korkeimman oikeuden ennakkopäätökset että korkeimman hallinto-oikeuden vuosikirjaratkaisut julkaistaan sekä suomeksi että ruotsiksi, tosin ruotsin kielellä ratkaisusta on saatavilla usein vain tiivistelmä.

⁹<http://finlex.fi/fi/laki/smur/>

3 Oikeudellisen tiedon esittäminen verkossa

Tässä luvussa tarkastellaan avoimen oikeudellisen datan käyttötapauksia sekä sen RDF-perustaiseen esittämiseen käytettyjä malleja ja standardeja. Standardeista esitellään tarkemmin Semanttinen Finlex -datajulkaisussa käyttöön otetut eurooppalaiset standardit ELI ja ECLI. Lopuksi tarkastellaan olemassa olevia oikeudellisia datapalveluita.

3.1 Käyttötapaukset

Avointa oikeudellista dataa voivat hyödyntää niin yksityishenkilöt, yritykset, juristit kuin viranomaisetkin [38]. Sekä yksityishenkilöiden että yritysten tulee toimia voimassaolevaa lainsäädäntöä noudattaen, mikä edellyttää voimassaolevan sääntelyn tuntemista. Juristeilla on aikaisemmin nostettuja oikeusjuttuja tarkastelleessaan usein tarve tuntea voimassaolevan lain lisäksi myös aiempi versio laista. Viranomaisten tarve avoimelle lakitiedolle ulottuu aina lainsäätäjien tarpeista organisaatioiden välisen tiedonvaihdon tehostamiseen.

Koneellisesti luettavaa avointa oikeudellista dataa voidaan hyödyntää suoraan esimerkiksi seuraavissa käyttötapauksissa:

- **Avointa dataa hyödyntävät sovellukset.** Avoimen oikeudellisen datan julkaisemisen tavoitteena on mahdollistaa uusien dataa hyödyntävien palvelujen rakentaminen [29]. Tämä edellyttää, että data on kokonaisuudessaan vapaasti ladattavissa tai käytettävissä avointen ohjelmointirajapintojen kautta.
- **Lainvalmistelu.** Laki on monimutkainen, keskinäisiä riippuvuussuhteita sisältävä kokonaisuus. Ristiriitaisuuksia välttääkseen lainlaatijan tulee tuntea laadittavan sääntelyn vaikutus voimassaolevaan lainsäädäntöön [74]. Suomessa lainlaatijan tulee tuntea perustuslain lisäksi säänneltävän oikeudenalan yleislait, kansainväliset sopimukset, yleiset oikeusperiaatteet sekä kaikki muu laadittavaan sääntelyyn vaikuttava lainsäädäntö [58]. Lainsäätäjän työtä voidaan helpottaa tarjoamalla semanttista tietoa eri säädösten välisistä riippuvuuksista.
- **Data-analyysi.** Kun lainsäädäntö ja oikeuskäytäntö ovat käytössä datana, voidaan niitä hyödyntää data-analyysin menetelmillä oikeustieteellisessä tutkimustyössä. Tutkimuksen kohteena voivat olla esimerkiksi oikeuskäytännön eurooppalaistumiskehitys [49] tai lainsäädännön sisältämien ristiviittausten mallintaminen ja analyysi monimutkaisena verkostona [46].
- **Tiedonvaihto eri organisaatioiden välillä.** Datan julkaiseminen avoimena helpottaa julkisyhteisöjen välistä tiedonjakoa [11]. Erillistä asiakirjaliikennettä organisaatioiden välillä ei datan avaamisen jälkeen enää tarvita.
- **Lainsäädännön ja oikeuskäytännön kansainvälinen yhteentoimivuus.** EU-sopimukset ja -asetukset ovat suoraan sovellettavissa jäsenvaltioiden tuomioistuimissa. Direktiivit ovat sovellettavissa kansallisen täytäntöönpanon

kautta. Jäsenvaltioiden tuomioistuimissa on usein tarve tietää, miten EU-lainsäädäntöä on tulkittu muiden jäsenmaiden tuomioistuinten ratkaisuisa [7]. Euroopan unionissa onkin viime vuosina käynnistetty useita eri hankkeita oikeudellisen tiedonhaun helpottamiseksi yli jäsenmaiden kansallisten rajojen.

- **Viittaukset lakiin.** Eri verkkopalveluissa on tarve viitata yksittäiseen lainkohtaan ja tarjota tämän sisältö lukijalle. Tämä käyttötapaus luo tarpeen lainsäädännölliselle ontologiapalvelulle [18], joka mahdollistaa viittaamisen lainkohtaan yksilöivällä tunnisteella. Sitä käyttäen myös lainkohdan sisältö on saatavissa datana.

Edellä mainitut käyttötapaukset edellyttävät, että data on verkosta vapaasti ladattavissa. Linkitetyn datan teknologiat ja standardit mahdollistavat datan avaamisen lisäksi semanttisen tiedon tarjoamisen, yksilöivien tunnisteiden luomisen sekä uuden tiedon päättelyn ontologiaperustaista tietoa hyödyntäen. Kansainvälinen yhteentoimivuus saavutetaan, kun eri aineistot ovat identifioitavissa yhdenmukaista mallia noudattavilla yksilöivillä tunnisteilla, ja aineistojen sisältämä data on esitetty yhteistä standardia noudattavilla tietomalleilla.

3.2 Linkitetty lakitieto

Linkitetyllä datalla tarkoitetaan käytäntöjen ja teknologioiden joukkoa, jota hyödyntäen data voidaan julkaista verkossa siten, että kaikkien aineistojen sisältämät asiat ovat koko verkon laajuisesti yksilöitävissä tunnisteella ja eri aineistojen ja niiden sisältämien asioiden välille pystytään luomaan globaaleja yhteyksiä.

Web-dokumenttien tai multimedian lisäksi linkitettyinä datana esitettävät asiat voivat olla todellisia objekteja tai käsitteitä [36]. Lain kontekstissa asiat voivat olla esimerkiksi säädösasiakirjoja, niiden laatimiseen ja julkaisemiseen liittyviä toimijoita tai asiakirjojen kuvailuun käytettyjä asiasanoja.

WWW:n (World Wide Web) keksijä Tim Berners-Lee [4] linjasi linkitetyn datan neljä pääperiaatetta:

1. Asiat tulee nimetä yksilöivin URI-tunnistein (Uniform Resource Identifier).
2. Asian URI-tunnisteena tulee käyttää HTTP-osoitetta (Hypertext Transfer Protocol).
3. URI-tunnistetta käyttämällä asiasta löytyy lisätietoa, joka tarjotaan käyttäjälle soveltaen linkitetyn datan standardeja, kuten RDF (Resource Description Framework) [16] ja SPARQL (SPARQL Protocol and RDF Query Language) [84]. RDF on linkitetyn datan tietomalli, jossa tieto esitetään subjekti-predikaatti-objekti-kolmikoita käyttäen. SPARQL on kyselykieli, jonka avulla RDF-muotoista tietoa voidaan hakea ja hyödyntää sovelluksissa.
4. Koska asiat sekä tunnistetaan että kuvaillaan HTTP URI -tunnisteiden avulla, voidaan tunnisteita myös seurata HTTP-protokollalla aina uuden lisätiedon löytämiseksi.

Linkitetyn datan visio on koko WWW:n laajuinen datapilvi, jossa aineistot täydentävät toisiaan niiden välille muodostettujen yhteyksien kautta. Tämä linkettyjen aineistojen verkko yhdessä siihen liittyvän teknologiapinon kanssa tunnetaan myös semanttisena webinä (engl. Semantic Web) tai tiedon verkkona (engl. Web of Data) [6] [36].

3.3 Oikeudelliset ontologiat

Ontologian käsite on peräisin filosofiasta, missä sillä tarkoitetaan kaiken olemassa olevan tunnistamista ja kuvailua. Tietotekniikassa käsitteen merkitys on muuttunut alkuperäisestä. Abstraktin ontologian käsitteen sijaan puhutaan laskennallisista ontologioista [32], jotka Studer et al. [72] määrittelevät ”formaaleiksi, eksplisiittisiksi määrittelyiksi yhteisistä käsitteistöistä”.

Laskennallinen ontologia mallintaa jotakin järjestelmää luomalla sen sisältämistä entiteeteistä käsitteitä (luokkia) ja määrittelemällä niiden väliset suhteet [32]. Luokat muodostavat luokkahierarkioita. Esimerkiksi lainsäädännön asiayhteydessä luokkia voivat olla säädös, laki ja asetus. Laki ja asetus voidaan määritellä säädöksen alaluokiksi, koska kaikki lait ja asetukset ovat säädöksiä. Ontologiat voivat luokkasuhteiden lisäksi sisältää myös ominaisuuksia (esim. *Eduskunta säätää lakeja*) tai rajoitteita (esim. *ainoastaan pykälä voi sisältää momentteja*).

Semanttisessa webissä ontologioilla pyritään vakiinnuttamaan yhteinen terminologia eri agenttien välille [3]. Ontologioiden RDF-perustaiseen esittämiseen on kehitetty useita eri sanastokieliä kuten RDFS (RDF Schema) [9] ja OWL (Web Ontology Language) [51].

RDFS sopii yksinkertaisten ontologioiden kuvailuun. RDFS-sanastolla voidaan määritellä luokkia (*rdfs:Class*), luokkahierarkioita (*rdfs:subclassOf*), ominaisuuksia (*rdfs:Property*), joita käytetään RDF-predikaatteina, sekä ominaisuushierarkioita (*rdfs:subPropertyOf*). Ominaisuudelle voidaan määritellä lähtöjoukoksi (*rdfs:domain*) ja arvojoukoksi (*rdfs:range*) yksittäisiä luokkia. RDF-resurssin voi merkitä kuuluvaksi tiettyyn luokkaan ominaisuudella *rdf:type*.

OWL on RDFS-sanastoa ilmaisuvoimaisempi ja soveltuu ontologiaperustaisen tiedon rikkaaseen kuvailuun. OWL mahdollistaa ominaisuuksien kuten kardinaliteetin, transitiivisuuden, symmetrisyyden ja yhtäläisyyden kuvaamisen. Siten OWL-kielillä voidaan kuvailla monimutkaisempia ontologioita kuin RDFS-sanastolla.

Oikeudellisen tiedon esittämiseen semanttisessa webissä on kehitetty keskenään hyvin erilaisia ja eri käyttötapauksia vastaavia ontologioita. Kehitystä ajaa eteenpäin erityisesti tarve kehittää uusia tietoteknisiä sovelluksia eri käyttäjäryhmille [10]. Valenten [81] mukaan oikeudellisia ontologioita voidaan hyödyntää oikeudellisen tiedon järjestämisessä ja rakenteistamisessa, päättelyssä ja ongelman ratkaisussa, semanttisessa indeksoinnissa ja tiedonhaussa, eri aineistojen semanttisen yhtentoimivuuden edistämässä sekä oikeudellisen tiedon ymmärtämisessä.

Oikeudelliset ontologiat voidaan jakaa niiden hienojakoisuuden perusteella ydin- ja toimialaontologioihin [10]. Toimialaontologia esittää yksittäiselle toimialalle ominaista tietämystä, kun sitä vastoin ydinontologia sisältää eri toimialaontologioille yhteisiä yleisempiä käsitteitä [69].

Oikeudellisia ydinontologioita rakennetaan tavallisesti asiantuntijaperusteisesti ylhäältä alaspäin-menetelmällä [5]. Ylhäältä alaspäin-lähestymistavassa ontologian rakentaminen aloitetaan kaikkein yleisimmistä käsitteistä. Esimerkkejä oikeudellisista ydinontologioista ovat LKIF-Core [39] sekä The Core Legal Ontology (CLO) [31].

Oikeudellisia ontologioita voidaan rakentaa myös alhaalta ylöspäin-menetelmällä. Menetelmä perustuu oikeudellisten käsitteiden eristämiseen juridisesta tekstistä luonnollisen kielen käsittelytekniikoita (NLP, Natural Language Processing) hyödyntäen [27]. NLP-tekniikoiden käyttöön liittyy yleisesti monia haasteita ja juridisten tekstien käsittelyä monimutkaistavat vielä oikeudellisen kielen erityispiirteet [83].

Gandonin [30] mukaan alhaalta ylöspäin-menetelmä on taipuvainen tuottamaan hyvin spesifisiä ontologioita ja yksityiskohtaisia käsitteitä. Francesconi et al. [27] mukaan alhaalta ylöspäin- ja ylhäältä alaspäin-lähestymistapojen yhdistäminen vaikuttaa kaikista lupaavimmalta menetelmältä oikeudellisten ontologioiden rakentamiseksi. Lähestymistapojen yhdistäminen mahdollistaa samanaikaisesti sekä oikeusteoriaan että oikeudellisten tekstien sisältöön tukeutuvan ontologian rakentamisen.

3.4 ELI-standardi lainsäädännön esittämiseksi

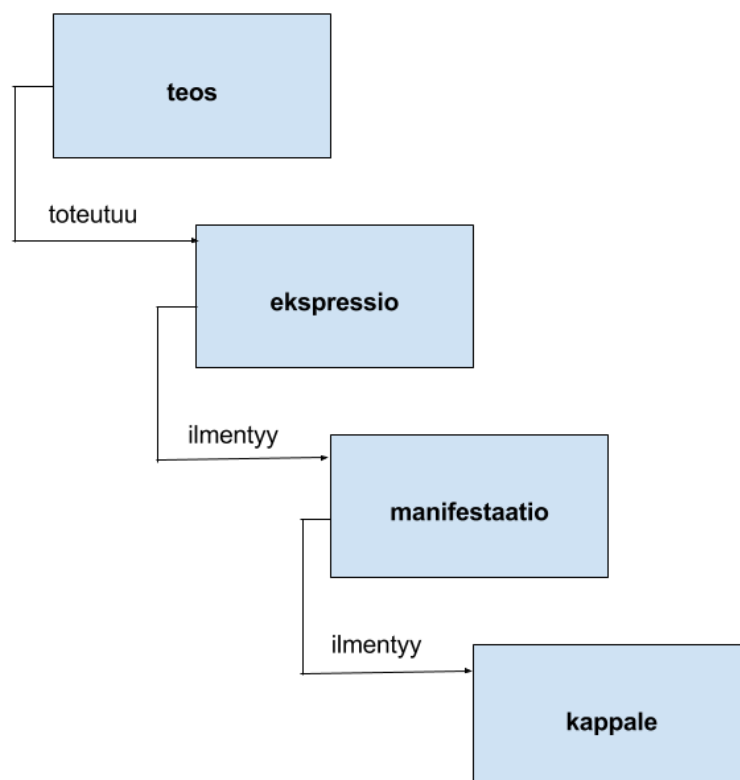
ELI (European Legislation Identifier) [14] on Euroopan unionin neuvoston suositus tunniste- ja metatietomalliksi Euroopan unionin sekä sen jäsenvaltioiden lainsäädännöllisille asiakirjoille. Sen tavoitteena on helpottaa lainsäädännöllisten aineistojen hakua ja yhdistämistä yli kansallisten rajojen standardisoimalla jäsenvaltioiden lainsäädännöllisten aineistojen esitystapaa. Lainsäädännön RDF-perustaiseen esittämiseen ELI tarjoaa URI-tunnisteratkaisuksi mallipohjan ja metatietomalliksi OWL-kieleen perustuvan ontologian.

ELI-standardin mukainen URI-tunniste mahdollistaa viittamisen tiettyyn dokumenttiin tai sen osaan sellaisena kuin se on ollut voimassa eri ajanhetkillä. URI-malli ottaa myös huomioon, että samasta asiakirjasta voi olla useita eri kieliversioita ja nämä kieliversiot voivat olla useassa eri sisältömuodossa.

ELI:n ontologia perustuu IFLA:n (International Federation of Library Associations and Institutions) bibliografisten tietueiden funktionaalisten vaatimusten työryhmän kehittämään FRBR-malliin (Functional Requirements for Bibliographic Records) [43]. FRBR on bibliografisten tietuiden kuvailemiseksi kehitetty käsitelmä, jonka on todettu soveltuvan myös lainsäädännön esittämiseen [48].

FRBR-malli erottaa käsitteet teos (engl. work), ekspressio (engl. expression), manifestaatio (engl. manifestation) ja kappale (engl. item). Teoksella tarkoitetaan jotakin tiettyä kirjallista teosta abstraktilla käsitteellisellä tasolla. Ekspressio tarkoittaa teoksen tiettyä versiota, kuten kirjan tietynkielistä käännöstä. Manifestaatio tarkoittaa ekspression fyysistä ilmentymää, joka voi olla esimerkiksi kirjan tietystä käännöksestä julkaistu PDF-dokumentti. Kappaleella tarkoitetaan yksittäistä manifestaation fyysistä ilmentymää, joka voi olla PDF-dokumentin kopio tietokoneen kovalevyllä. Terminologiaa on havainnollistettu kuvassa 4, joka esittää FRBR-mallin entiteetit ja niiden väliset suhteet.

ELI soveltaa FRBR-perustaista kuvailua luomalla lainsäädännölliselle asiakirjoille FRBR-vastinparit seuraavasti:



Kuva 4: FRBR-käsittemallin entiteetit ja niiden väliset suhteet.

- *eli:LegalResource* vastaa teosta eli säädöstä tai lainkohtaa käsitteellisellä tasolla
- *eli:LegalExpression* vastaa ekspressiota eli säädöksen tai lainkohdan toteutumaa eri ajanhetkinä tai eri kielillä
- *eli:Format* vastaa manifestaatiota eli kieliversion fyysistä ilmentymää tietyssä sisältöformaatissa

Kuvailumallin mukaisten luokkien lisäksi ELI määrittelee jokaiselle luokalle omat RDF-perustaiset metatietokentät, joista osa on vapaaehtoisia, osa suositeltuja ja osa pakollisia. RDF-metadatan suositellaan esitettäväksi upotettuna HTML-muotoisiin säädösasiakirjoihin RDFa-määrittelyn [1] mukaisesti [24].

Koska ELI:n halutaan olevan sovellettavissa eri maiden epäyhtenäisten lainsäädännöllisten aineistojen mallintamiseen, on sen ontologia hyvin yleinen ja joustava. Siksi ELI ei sinänsä sovellu lakiaineistojen yksityiskohtaiseen semanttiseen kuvailuun eikä säädösasiakirjan sisäisen rakenteen esittämiseen [23]. ELI-ontologiaa suositellaankin käytettäväksi yhdessä räätälöityjen metatietoskeemojen kanssa [14].

ELI-standardin käyttöönotto Euroopan unionin jäsenmaissa perustuu vapaaehtoisuuteen, ja jäsenmaat voivat edetä ELI:n käyttöönotossa omaa tahtiaan [24]. ELI-standardi on otettu käyttöön Semanttinen Finlex- palvelussa seuraavista syistä:

- **Metatietomäärittelyjen riittävyys.** ELI-ontologia sisältää riittävän määrän metatietomäärittelyjä suhteessa julkaistavan datan tarjoamiin metatietoihin.
- **Aineistojen yhteentoimivuus.** ELI pyrkii edistämään eurooppalaisten oikeudellisten tietoaaineistojen yhteentoimivuutta luomalla yhtenäisen tunniste- ja metatietomallin, joka on jo käytössä eri datajulkaisuissa.
- **Laajennettavuus.** ELI-standardiin perustuva tietomalli on myöhemmin laajennettavissa muilla oikeudellista tietoa yksityiskohtaisemmin kuvaavilla ontologioilla.

Semanttinen Finlex -palvelun lisäksi ELI on otettu käyttöön esimerkiksi Yhdistyneessä kuningaskunnassa¹⁰, Ranskassa¹¹, Luxemburgissa¹², Irlannissa¹³, Norjassa¹⁴ sekä Euroopan unionin julkaisuutoimistossa¹⁵. Suurin osa toteutuksista on kuitenkin vielä kehitysvaiheessa, eivätkä kaikki niistä tarjoa ELI-perustaista metadataa dokumentteihin upotettuna. Yksikään toteutuksista ei tarjoa ELI-perustaista RDF-metadataa SPARQL-palvelupisteen kautta siten, että siihen voisi suoraan kohdistaa SPARQL-kyselyitä.

3.5 ECLI-standardi oikeuskäytännön esittämiseksi

ECLI (European Case Law Identifier) [13] on eurooppalaisen oikeuskäytännön esittämiseksi kehitetty tunniste- ja metatietostandardi. Oikeuslaitoksilla on perinteisesti ollut hyvin poikkeavia käytäntöjä ratkaisujen tunnisteiden luomisessa [63]. ECLI syntyi tarpeesta helpottaa eri ratkaisuasikirjoihin viittaamista luomalla yhteinen tunnistestandardi eurooppalaisille oikeuslaitoksille.

ECLI-asiakirjatunniste koostuu viidestä seuraavassa järjestyksessä olevasta, pakollisesta elementistä:

1. *ECLI*
2. maatunnus
3. tuomioistuimen tunnus
4. ratkaisun antovuosi
5. ratkaisun yksilöivä tunniste, joka voi olla enintään 25 merkkiä pitkä ja sisältää vain alfanumeerisia merkkejä.

¹⁰<http://legislation.gov.uk>

¹¹<http://www.eli.fr/en/>

¹²<http://eli.legilux.public.lu/>

¹³<http://www.irishstatutebook.ie/>

¹⁴<https://lovdata.no/>

¹⁵<https://publications.europa.eu/en/home>

ECLI-tunnisteen elementit on erotettu toisistaan kaksoispisteellä. Esimerkiksi *ECLI:FI:KKO:2016:55* on korkeimman oikeuden vuoden 2016 55:nnen ennakkopäätöksen ECLI-tunnus. Vastaavasti korkeimman hallinto-oikeuden vuoden 2016 55:nnen vuosikirjaratkaisun ECLI-tunnus on *ECLI:FI:KHO:2016:55*.

Tunnisteen lisäksi standardi määrittelee asiakirjoille joitakin pakollisia ja valinnaisia metatietokenttiä. Metatietomäärittelyt perustuvat Dublic Core-sanastoon [19], jota käytetään yleisesti sähköisten asiakirjojen metatietojen määrittelyyn.

EEJP (European e-Justice Portal)¹⁶ pyrkii toimimaan EU:n oikeudellisen tiedon ja palvelujen kokoajana [64]. Portaalista löytyvä ECLI-sivusto kokoaa yhteen yleistä tietoa ECLI-standardista sekä sen käyttöön otosta eri EU-maissa. Portaalista löytyy ECLI-hakukone, jolla dokumentteja voi hakea niiden ECLI-perustaisten metatietojen perusteella.

Monet EU-maat, Suomi mukaanlukien, tuottavat jo ECLI-standardin mukaisia oikeuden ratkaisiasiakirjoja julkaisuprosessiensa yhteydessä. Edita Publishing Oy tuottaa Finlex-julkaisun yhteydessä oikeuden ratkaisusta ECLI-metatietomäärittelyn mukaisia XML-muotoisia asiakirjoja. Asiakirjoista luodaan indeksointitiedosto, jonka perusteella EEJP:n hakukone hakee ratkaisujen ECLI-perustaiset metatiedot [76].

ECLI-standardi valittiin perustaksi Semanttinen Finlex -julkaisun oikeuden ratkaisujen tietomallille samoista syistä kuin ELI-standardi. ECLI-standardin valintaa puoltaa myös se, että standardi on jo otettu käyttöön Finlex-julkaisun yhteydessä. Siten KKO:n ja KHO:n päätökset saadaan Edita Publishing Oy:ltä valmiiksi ECLI-standardin mukaisessa XML-rakenteessa, mikä tekee niiden muuntamisesta Semanttinen Finlex -julkaisun RDF-tietomallin mukaisiksi melko suoraviivaista.

3.6 Oikeudelliset datapalvelut

Oikeudellisia aineistoja on julkaistu linkitettynä avoimena datana eri avoimen tiedon edistämishankkeissa. Seuraavassa tarkastellaan joitakin eurooppalaisia oikeudellisen datan julkaisuja.

Hollannin kansallinen lainsäädäntö on julkaistu linkitettynä avoimena datana Metalex Document Server -palvelussa¹⁷ [38]. Julkaisussa säädösasiakirjojen rakenne on esitetty CEN MetaLex-XML-rakenteessa [8].

MetaLex on standardoitu esitystapa oikeudellisten asiakirjojen esittämiseksi verkossa. Sen tarkoitus ei ole korvata kansallisia käytäntöjä asiakirjojen valmistelussa, vaan tarjota yleistetty standardi asiakirjojen esittämiseen ja siten lisätä eri aineistojen yhteentoimivuutta. Yleistetyn rakenteensa ansiosta MetaLex on otettavissa käyttöön riippumatta kieli- ja hallintoaluekohtaisista käytännöistä. XML-rakenteen lisäksi MetaLex sisältää RDF-perustaisen ontologian, joka määrittelee tapahtumamallin lainsäädännöllisten asiakirjojen muutoksille. Metalex Document Server -palvelussa säädösten metadata on julkaistu RDF-muodossa käyttäen MetaLex-ontologian lisäksi PROV-ontologiaa (Provenance Ontology) [47], SEM-mallia (Simple Event Model) [34] sekä OWL-Time-ontologiaa [15].

¹⁶<https://e-justice.europa.eu/home.do?action=home>

¹⁷<http://doc.metalex.eu/>

Metalex Document Server -palvelun aineisto päivittyy joka päivä. Jokaisen asiakirjan RDF-metatiedot sekä XML-muotoisen sisällön voi ladata palvelusta kerralla. Dataan voi myös kohdistaa SPARQL-kyselyitä palvelupisteen kautta. Julkaisun URI-tunnisteiden käyttö perustuu sisältöneuvotteluun (engl. content negotiation), jossa asiakasohjelma lähettää HTTP-kutsun ylätunnisteessa (engl. header) pyynnön tietystä sisältömuodosta, ja datapalvelin tarjoaa sisällön kyseisessä muodossa. Metalex Document Server -palvelun tarjoama sisältö on joko säädösasiakirjan RDF-perustainen metadata, XML-dokumentti tai HTML-dokumentti.

Yhdistyneessä kuningaskunnassa lainsäädännöllisiä aineistoja on julkaistu linkitetynä datana legislation.gov.uk-palvelussa. Säädösten rakenne on kuvattu paikallisen XML-rakenteen mukaisena, mutta RDF-metatiedot perustuvat osittain MetaLex-ontologiaan. Palvelussa käyttöön otetut URI-tunnisteet ovat yhdenmukaisia ELI-standardin kanssa ja toimivat sisältöneuvotteluperustaisesti.

EU:n oikeudelliset asiakirjat löytyvät kaikilla EU:n virallisilla kielillä EUR-Lex-palvelusta¹⁸. Palvelussa on otettu käyttöön ELI-muotoiset tunnisteet ja metadata. Julkaisutoimiston ELI-toteutus perustuu CDM-tietomallilla (Common Data Model) esitettyjen metatietojen ELI-muunnokseen. CDM on FRBR-malliin sekä OWL- ja RDFS-tekniikoihin perustuva Euroopan Unionin julkaisutoimiston resurssien kuvaamiseksi kehitetty ontologia [25]. Aineistoihin voi kohdistaa SPARQL-kyselyitä CELLAR-rajapinnan¹⁹ kautta. CELLAR-tietokannasta ei löydy ELI-perustaista metatietoa, mutta ELI-tunnistetta käyttäen pystyy hakemaan EU-säädösten CDM-mallilla kuvattuja metatietoja.

¹⁸<http://eur-lex.europa.eu/>

¹⁹<http://publications.europa.eu/webapi/rdf/sparql>

4 Lainsäädännön tietomalli

Tässä luvussa esitellään Semanttinen Finlex -palvelussa julkaistavien lainsäädännöllisten aineistojen XML-rakenteet sekä kuvataan Semanttinen Finlex -julkaisun ELI-standardiin perustuvat URI-käytännöt ja RDF-perustainen tietomalli.

4.1 Aineistojen XML-rakenteet

Alkuperäiset säädökset

```
<saa:Johtolause>
  <saa:SaadosKappaleKooste>
    Eduskunnan päätöksen mukaisesti
  </saa:SaadosKappaleKooste>
  <saa:SaadosKappaleKooste>
    <saa:SaadosKursiiviKooste>muutetaan</saa:SaadosKursiiviKooste>
    liikennevalvontaa vaikeuttavien laitteiden kieltämisestä annetun
    lain (546/1998) 7 ja 8 § seuraavasti:
  </saa:SaadosKappaleKooste>
</saa:Johtolause>
```

Kuva 5: Esimerkki muutossäädöksen johtolauseen XML-rakenteesta.

Alkuperäiset säädökset saadaan Edita Publishing Oy:ltä Tieto Finland Oy:n valtioneuvoston päätösasiakirjoille laatiman XML-rakenteen [80] mukaisina. Säädöksiä ei kuitenkaan laadita lainvalmistelussa suoraan XML-muotoon, koska rakenteisen sisällöntuotannon käyttöönotossa lainvalmisteluprosessissa on ollut haasteita [37]. Valtioneuvoston kanslia on esittänyt tietopyynnön [82], jolla se hakee riittävän helppokäyttöistä teknistä ratkaisua asiakirjojen laatumiseksi suoraan rakenteiseen XML-muotoon.

Tällä hetkellä säädösasiakirjat laaditaan Word-dokumenteiksi Valtioneuvoston päätöksentekojärjestelmällä (PTJ), jonka avulla hoidetaan myös tiedonvälitys eduskuntaan sekä Finlex-julkaisuun. Muunnoksen onnistuminen Word-dokumentista XML-muotoon riippuu Word-asiakirjapohjan käytön tarkkuudesta. Säädösvalmistelijat käyttävät Word-asiakirjapohjalla laaditun hallituksen esityksen, lain, asetuksen tai valtioneuvoston kirjelmän rakenteen validoimiseksi erillistä VaRa-nimistä sovellusta. Finlex-julkaisua varten Word-asiakirja lähetetään Edita Publishing Oy:lle, missä se muunnetaan valtioneuvoston päätösasiakirjoille laaditun skeeman mukaiseen XML-rakenteeseen [76]. Taittajat muokkaavat dokumenttia vielä FrameMaker-julkaisuohjelmalla ennen sen julkaisemista sähköisessä säädöskokoelmassa.

Valtioneuvoston päätösasiakirjoille laadittu rakennemäärittely ei täysin vastaa prosessissa syntyvien XML-dokumenttien rakennetta, koska joitakin elementtejä ei kyetä tunnistamaan Word-asiakirjapohjasta XML-muunnoksen yhteydessä. Lisäksi rakennemäärittelyjen monimutkaisuus aiheuttaa virheitä myös FrameMaker-työkalua käytettäessä [76]. Esimerkiksi rangaistussäännösten loppukappaleet on toistuvasti merkitty XML-rakenteessa virheellisesti alakohdiksi ja loppukappaletta edeltävä kohta on myös virheellisesti merkitty alakohdaksi.

Alkuperäisten säädösten XML-rakenteesta löytyy niukasti metatietoja. XML-skeema on lähinnä dokumentin rakenteen kuvaus, joka ei sisällä rakenteista tietoa itse dokumentin sisällöstä. Esimerkiksi kuvasta 5 löytyvän muutossäädöksen XML-rakenteessa ei ole tietoa siitä, mitä lakia ja lainkohtia muutossäädös muuttaa. Tieto muutetuista lainkohdista löytyy ainoastaan johtolauseen tekstisisällöstä.

Ajantasaiset säädökset

```
<py t="VM" id="sd19980546:00:L0:P8" sdk="20110866">
  <nu>8 §</nu>
  <vtref vtid="22.7.2011/866">(22.7.2011/866)</vtref>
  <ot>Henkilöntarkastus</ot>
  <mo id="sd19980546:00:L0:P8:M1">
    <te>
      Jos on erittäin todennäköisiä perusteita epäillä, että on tehty 3
      §:n 1 kohdassa tarkoitettu paljastinlaiterikkomus, voidaan
      toimittaa pakkokeinolaissa tarkoitettu henkilöntarkastus
      moottorikäyttöisen ajoneuvon kuljettajalle ja matkustajalle sen
      estämättä, mitä pakkokeinolain 8 luvun 31 §:n 1 momentissa
      säädetään.
    </te>
  </mo>
</py>
```

Kuva 6: Esimerkki ajantasaisen säädöksen XML-rakenteesta.

Ajantasaiset säädökset saadaan Edita Publishing Oy:ltä heidän itse kehittämässään XML-rakenteessa. Kuvassa 6 on esimerkki kyseisessä XML-rakenteessa esitetystä pykälästä. *py* kertoo elementin olevan pykälä ja *id*-attribuutin arvo *sd19980546:00:L0:P8* kertoo pykälän olevan säädöksen 546/1998 kahdeksas pykälä. Kirjaimet *O* ja *L* viittaavat osaan ja lukuun, joita kyseisessä säädöksessä ei ole. Elementin *t*-attribuutin arvo *VM* kertoo pykälän olevan voimassa oleva muutos ja *sdk*-attribuutin arvo *20110866* kertoo pykälän olevan peräisin muutossäädöksestä 866/2011. *t*-attribuutin arvon ensimmäinen kirjain voi olla *V*, *S* tai *M* ja se ilmaisee, onko lainkohta voimassa, voimassa väliaikaisesti vai myöhemmin voimaantuleva. Toinen kirjain voi olla *A*, *M*, *L* tai *K* ja ilmaisee, onko lainkohta alkuperäinen, muutos, lisäys vai kumottu.

```
<ref id="22.7.2011/866" tark="T">
  <refot>22.7.2011/866:</refot>
  <te>Tämä laki tulee voimaan 1 päivänä tammikuuta 2014.</te>
  <ev ty="HE">
    <viite t="DTHE" kohde="he20100222" VALID="YES">HE 222/2010</viite>
  </ev>
  <ev ty="MT">
    <viite t="DTMT" kohde="lavm20100044" VALID="NO">LaVM 44/2010</
viite></ev>
  <ev ty="EV">EV 374/2010</ev>
</ref>
```

Kuva 7: Esimerkki muutossäädöksen viitetiedosta ajantasaisen säädöksen XML-rakenteessa.

Pykälän sisällä oleva *vtref*-elementti on viittaus dokumentin lopussa oleviin muutossäädöksen voimaantulotietoihin. Kuvassa 7 on esimerkki dokumentin lopusta löytyvästä viitetiedosta. Muutossäädöksen antopäivä löytyy otsikosta ja voimaantulopäivä on kirjoitettu *te*-elementin sisään. Lisäksi viitetiedoista löytyvät viitteet muutossäädöksen esitöihin. Kuvan esimerkistä löytyvät viitteet hallituksen esitykseen (*HE*), lakivaliokunnan mietintöön (*La VM*) sekä eduskunnan vastaukseen (*EV*).

Edita Publishing Oy:n on tarkoitus uudistaa ajantasaisten säädösten XML-rakennetta lähitulevaisuudessa [76]. Tavoitteena on tarjota enemmän metatietoja XML-rakenteessa ja ottaa käyttöön jokin päätösasiakirjoille tarkoitettu XML-standardi.

4.2 Säädösten URI-tunnisteet

Semanttinen Finlex -palvelussa säädösasiakirjojen URI-tunnisteet on muodostettu ELI-standardia noudattaen. ELI:n URI-mallin elementeistä on valittu elementit siten, että yksittäiseen lakiin tai lainkohtaan voidaan viitata yksikäsitteisesti sellaisena kuin se on ollut voimassa eri ajanhetkillä, eri kielillä ja eri sisältömuodoissa.

Käyttöön otettu URI-malli on muotoa:

$$\begin{aligned} &/eli/{asiakirjatyyppe}/\{antovuosi}/\{säädösnumero}/\{lainkohta} \\ &\quad /{\versio}/\{ajankohta}/\{kieli}/\{formaatti} \end{aligned}$$

Taulukossa 1 on luetteloitu käytössä olevat URI-elementit muotoineen ja mahdollisine arvojoukkoineen.

Taulukko 1: Lainsäädännöllisten aineistojen URI-elementit

ELI-elementti	Muoto	Arvojoukko
Asiakirjatyyppe	merkkijono	<i>sd</i>
Antovuosi	YYYY	
Säädösnumero	merkkijono	
Lainkohta	merkkijono	
Versio	merkkijono	<i>ajantasa, alkup</i>
Ajankohta	YYYYMMDD	
Kieli	ISO 639-2	<i>fin, swe</i>
Formaatti	merkkijono	<i>xml, html, txt</i>

URI-tunnisteessa ensimmäisenä esiintyvä elementti on asiakirjatyyppe, joka voi saada arvokseen *sd*, joka viittaa säädökseen. Säädösten lisäksi URI-malli on siten laajennettavissa myös muille asiakirjatyypeille.

Asiakirjatyypin jälkeen URI:ssa esiintyvät antovuosi- ja säädösnumero viittaavat alkuperäisen säädöksen antovuoteen ja säädösnumeroon. Lainkohta merkitään URI-tunnisteeseen säilyttäen rakenteen hierarkia. Esimerkiksi 2 luvun 3a pykälä merkitään URI-tunnisteessa *luku/2/pykala/3a*. Poikkeus hierarkian säilyttämiseen

URI-tunnisteessa ovat vuodesta 1993 eteenpäin annettujen säädösten osat. Osa jätetään pois URI-tunnisteesta, koska lukujen numerointi on jatkuva osien välillä.

Versio voi olla joko *ajantasa* tai *alkup*. Jos lainkohtaa on muutettu muutossäädöksellä tai se on kumottu, on versio *ajantasa*. Jos lainkohtaa ei olla muutettu, on versio *alkup*. Ajankohta-elementti viittaa lainkohdan voimaantulopäivään ja se löytyy URI-tunnisteesta vain siinä tapauksessa, että lainkohta on muutettu, jolloin sen versio on *ajantasa*.

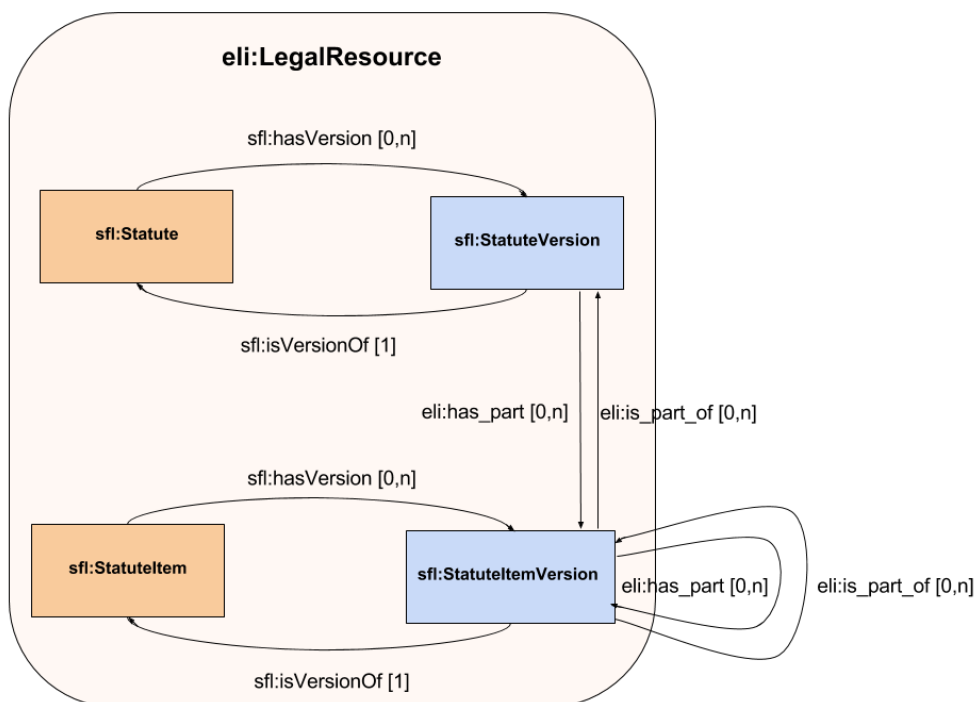
Kieli-elementti ilmaisee lainkohdan kielen. ELI-työryhmä [23] suosittelee käyttämään standardin ISO 639-2 mukaisia kolmimerkkisiä kielikoodeja. URI-tunnisteen formaatti-elementti ilmaisee, onko lainkohdan sisältö esitetty HTML-, XML- vai tekstimuodossa.

4.3 FRBR-kuvailumallin soveltaminen

Semanttinen Finlex -palvelussa on käytössä ELI-ontologian versio 1.0. Luvussa 3 todettiin ELI-ontologian määrittelevän kolme FRBR-kuvailumallin mukaista käsitteellistä tasoa säädöksille. Koska samasta säädöksestä tai lainkohdasta voi eri kieliversioiden ja sisältömuotojen lisäksi olla olemassa myös sisällöltään erilaisia muutossäädöksillä muutettuja versioita, on perusteltua ottaa käyttöön vielä neljäs käsitteellinen taso. Tämän vuoksi ELI:n ontologiaa on Semanttinen Finlex -palvelussa laajennettu palvelua varten kehitetyllä SFL-ontologialla (Semantic Finlex Legislation), joka määrittelee FRBR-mallin teostasoa kuvaavalle luokalle *eli:LegalResource* neljä alaluokkaa seuraavasti:

- *sfl:Statute* (Säädös), jolla kuvataan säädöstä abstraktina tunnisteena (esimerkiksi Rikoslaki (39/1889)), ja joka kokoaa säädöksen eri versiot yhteen
- *sfl:StatuteItem* (Lainkohta), joka vastaavasti kokoaa eri lainkohdan versiot yhteen, ja jolla kuvataan lainkohtaa abstraktina tunnisteena (esimerkiksi Rikoslain 15 luku)
- *sfl:StatuteVersion* (Säädöksen versio), jolla kuvataan säädöstä tietyllä ajanhetkellä (esimerkiksi Rikoslaki sellaisena kuin se oli voimassa 1.1.2016)
- *sfl:StatuteItemVersion* (Lainkohdan versio), jolla kuvataan lainkohtaa tietyllä ajanhetkellä (esimerkiksi Rikoslain 15 luku sellaisena kuin se oli voimassa 1.1.2016)

Sekä säädösten ja säädösten versioiden että lainkohtien ja lainkohtien versioiden välisiä suhteita kuvataan SFL-ontologiaan määritetyillä ominaisuuksilla *sfl:hasVersion*-sekä *sfl:isVersionOf*. Ominaisuudet on määritelty Dublin Core -sanaston samannimisten ominaisuuksien alaominaisuuksiksi. Yhdellä lainkohdalla voi olla yksi tai ei yhtään alkuperäistä versiota (alkuperäistä ei välttämättä ole saatavilla datassa) ja mahdollisesti ajantasaistettuja versioita, jos lainkohtaa on muutettu muutossäädöksillä.



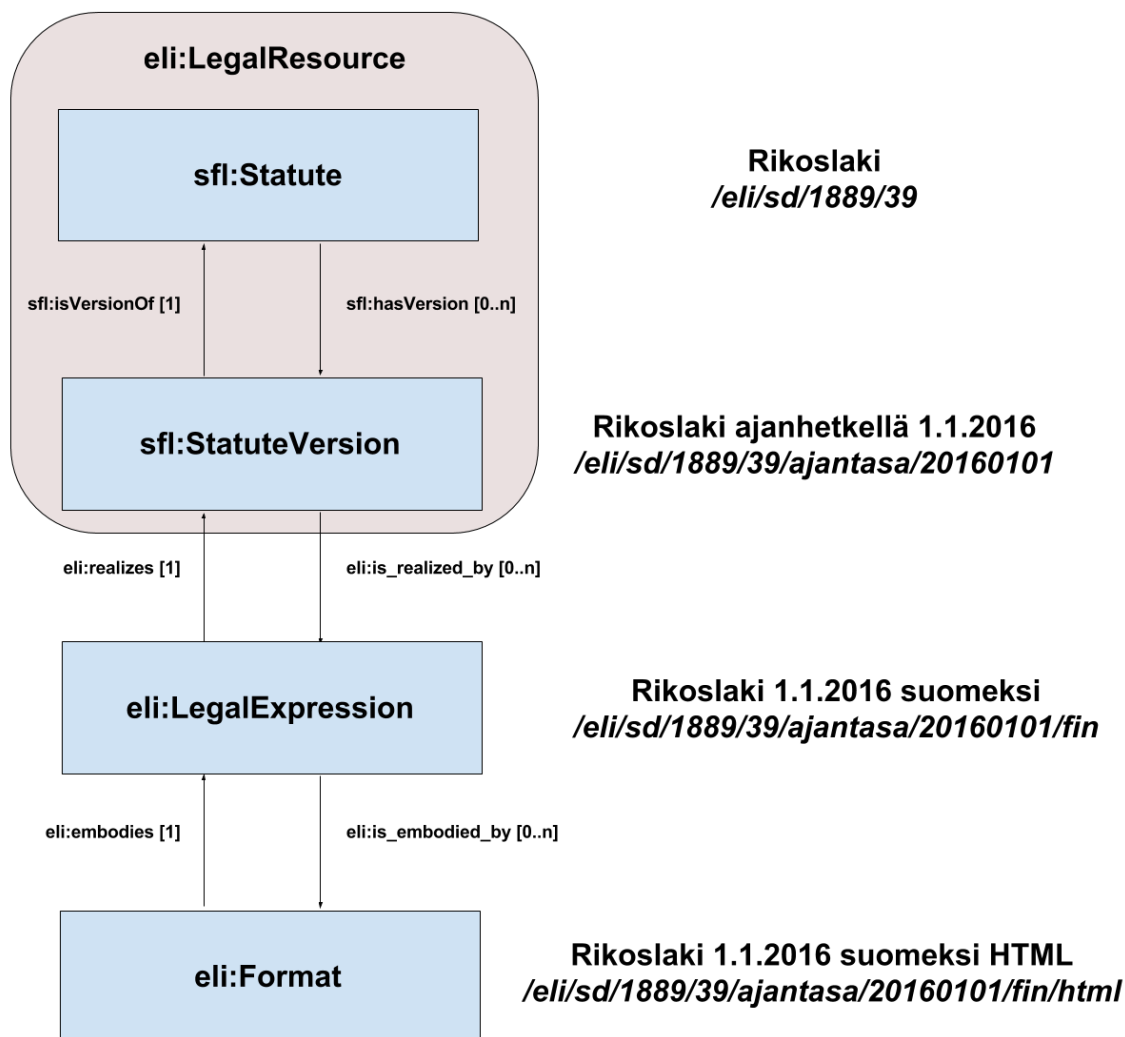
Kuva 8: *eli:LegalResource*-luokan jakautuminen alaluokkiin SFL-ontologiassa ja luokkien ilmentymien väliset suhteet

Kuvassa 8 on havainnollistettu edellämainittujen luokkien suhdetta toisiinsa sekä luokkien ilmentymien välisiä ominaisuuksia. Ominaisuuden perään on merkitty sen kardinaliteetti kulmasulkeiden sisään.

ELI-työryhmä [23] linjaa, että eri aikaversiot voi kuvata joko *eli:LegalResource*-tai *eli:LegalExpression*-luokilla, paitsi jos säädökset on esitetty usealla eri kielellä. *eli:LegalExpression*-luokkaa ei silloin voi käyttää eri aikaversioiden kuvaamisen, vaan ainoastaan eri kieliversioiden erotteluun. Tämä on tilanne Semanttinen Finlex -datajulkaisun tapauksessa, koska asiakirjat julkaistaan sekä suomeksi että ruotsiksi. Toisaalta IFLA:n FRBR-loppuraportti [43] ei tiukasti rajaa sitä, mitä pidetään teoksena ja mitä ekspressiona. Erityisesti lainsäädännön tapauksessa rajanvetoa on vaikea tehdä, koska saman lainkohdan eri aikaversioilla ei välttämättä ole sisällöllisesti tekemistä toistensa kanssa etenkin silloin, jos lainkohdat ovat siirtyneet rakenteessa. Näistä syistä Semanttinen Finlex -datajulkaisun tietomallissa on päädytty hajottamaan *eli:LegalResource*-luokka eli teostasoa kahdeksi eri käsitteelliseksi tasoksi, jotka ovat abstrakti tunniste ja versio, jolla tarkoitetaan tietyn säädöksen tai lainkohdan toteutumaa tietyllä ajanhetkellä.

Säädöksen tai lainkohdan version tiettyä kieliversiota mallinnetaan suoraan ELI-ontologian määrittelemälle *eli:LegalExpression*-luokalla, joka vastaa FRBR-mallissa ekspressiotasoa. Edelleen kieliversion tiettyä sisältömuotoa mallinnetaan *eli:Format*-luokalla, joka vastaa FRBR-mallissa manifestaatiotasoa.

Tietomallin käyttöä ja luokkien ilmentymien välisiä suhteita on havainnollistet-



Kuva 9: SFL-ontologia, ELI-muotoiset tunnisteet ja Rikoslaki esimerkkinä

tu kuvassa 9. Kuvan vasemmalla puolella olevien luokkien väliin piirretyt nuolet kuvaavat luokkien ilmentymien välille määriteltyjä ominaisuuksia. Ominaisuuden perään on merkitty sen kardinaliteetti kulmasulkeiden sisään. Oikealla puolella kuvassa on käytetty Rikoslakia (39/1889) esimerkkinä jokaisen luokan ilmentymästä. Vastaavat ELI-muotoiset URI-tunnisteet löytyvät myös esimerkistä. Eri luokkien kaikki ominaisuudet on listattu taulukoissa 3, 4, 5, 6, 7 ja 8. Taulukoissa listatut ominaisuudet ja niiden kuvaama data esitellään tarkemmin seuraavissa aliluvuissa. Käytetyt nimiavaruudet löytyvät taulukosta 2.

Taulukko 2: Lainsäädännön mallinnuksessa käytetyt nimiavaruudet

Lyhenne	Nimiavaruus
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#
sfl	http://data.finlex.fi/schema/sfl/
skos	http://www.w3.org/2004/02/skos/core#
xsd	http://www.w3.org/2001/XMLSchema#

Taulukko 3: *sfl:Statute*-luokan ominaisuudet

Ominaisuus	RDF-tunniste	Arvojoukko
säädösnumero	<i>eli:id_local</i>	<i>rdfs:Literal</i>
versio	<i>sfl:hasVersion</i>	<i>sfl:StatuteVersion</i>
säädöstaso	<i>eli:type_document</i>	<i>eli:ResourceType</i>
säädöksen tyyppi	<i>sfl:statuteType</i>	<i>sfl:StatuteType</i>

Taulukko 4: *sfl:StatuteItem*-luokan ominaisuudet

Ominaisuus	RDF-tunniste	Arvojoukko
versio	<i>sfl:hasVersion</i>	<i>sfl:StatuteItemVersion</i>

Taulukko 5: *eli:LegalExpression*-luokan ominaisuudet

Ominaisuus	RDF-tunniste	Arvojoukko
versio	<i>eli:realizes</i>	<i>sfl:StatuteVersion</i> <i>sfl:StatuteItemVersion</i>
manifestaatio	<i>eli:is_embodied_by</i>	<i>eli:Format</i>
kieli	<i>eli:language</i>	<i>eli:Language</i>
nimike/otsikko	<i>eli:title</i>	<i>rdfs:Literal</i>
lempinimi	<i>eli:title_alternative</i>	<i>rdfs:Literal</i>

Taulukko 6: *eli:Format*-luokan ominaisuudet

Ominaisuus	RDF-tunniste	Arvojoukko
ekspressio	<i>eli:embodies</i>	<i>eli:LegalExpression</i>
sisältömuoto	<i>eli:format</i>	URI
teksti	<i>sfl:text</i>	<i>rdfs:Literal</i>
HTML	<i>sfl:html</i>	<i>rdfs:Literal</i>
XML	<i>sfl:xml</i>	<i>rdf:XMLLiteral</i>

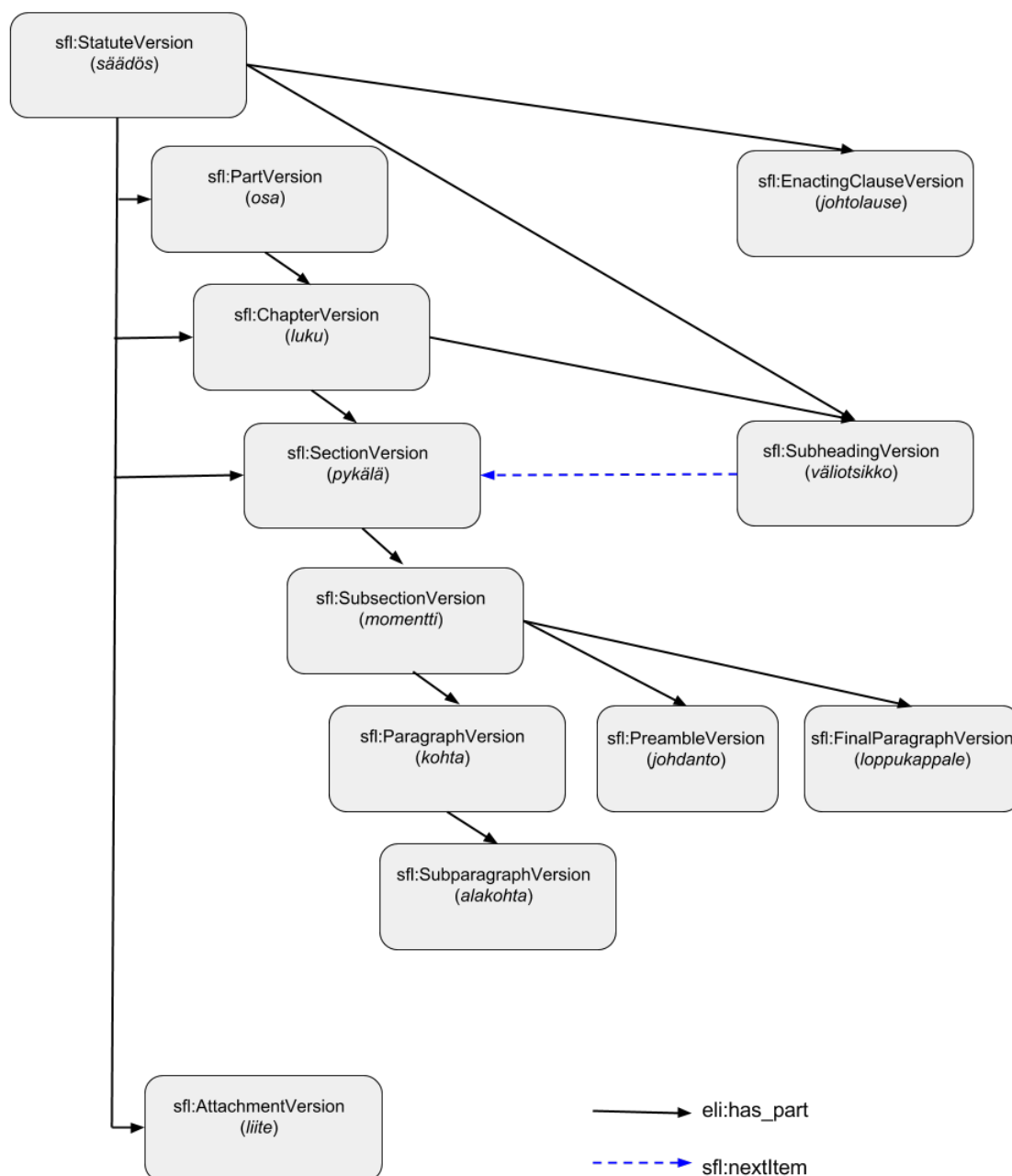
Taulukko 7: *sfl:StatuteVersion*-luokan ominaisuudet
 (*alkuperäiset versiot, **muutossäädöksellä muutetut versiot)

Ominaisuus	RDF-tunniste	Arvojoukko
säädös (tunnistetaso)	<i>sfl:isVersionOf</i>	<i>sfl:Statute</i>
kieliversio	<i>eli:is_realized_by</i>	<i>eli:LegalExpression</i>
sisältää lainkohdan	<i>eli:has_part</i>	<i>sfl:StatuteItemVersion</i>
version tyyppi	<i>eli:version</i>	<i>eli:Version</i>
voimaantulopäivä*	<i>eli:first_date_entry_in_force</i>	<i>xsd:date</i>
allekirjoituspäivä*	<i>eli:date_document</i>	<i>xsd:date</i>
julkaisupäivä*	<i>eli:date_publication</i>	<i>xsd:date</i>
voimaantulopäivä**	<i>eli:version_date</i>	<i>xsd:date</i>
kumoamispäivä**	<i>eli:date_no_longer_in_force</i>	<i>xsd:date</i>
asiasanat	<i>eli:is_about</i>	<i>skos:Concept</i>
säädöksen hyväksynyt elin*	<i>eli:passed_by</i>	<i>eli:Agent</i>
nojautuu*	<i>eli:based_on</i>	<i>sfl:StatuteVersion</i> <i>sfl:StatuteItemVersion</i>
muuttaa*	<i>eli:changes</i>	<i>sfl:StatuteItemVersion**</i>
kumoo*	<i>sfl:repeals</i>	<i>sfl:StatuteVersion**</i> <i>sfl:StatuteItemVersion**</i>
säädöksen nojalla säädetty	<i>eli:basis_for</i>	<i>sfl:StatuteVersion*</i>
säädöksen muuttama**	<i>eli:changed_by</i>	<i>sfl:StatuteVersion*</i>
säädöksen kumoama**	<i>sfl:repealedBy</i>	<i>sfl:StatuteVersion*</i>
esityöviittaus*	<i>sfl:related_to</i>	<i>sfl:GovernmentProposal</i>
esityöviittaus EU-sääd.*	<i>eli:transposes</i>	<i>eli:LegalResource</i>
voimassaolo	<i>eli:in_force</i>	<i>eli:InForce</i>

Taulukko 8: *sfl:StatuteItemVersion*-luokan ominaisuudet
 (*alkuperäiset versiot, **muutossäädöksellä muutetut versiot, +väliotsikon versio)

Ominaisuus	RDF-tunniste	Arvojoukko
lainkohta (tunnistetaso)	<i>sfl:isVersionOf</i>	<i>sfl:StatuteItem</i>
kieliversio	<i>eli:is_realized_by</i>	<i>eli:LegalExpression</i>
sisältää lainkohdan	<i>eli:has_part</i>	<i>sfl:StatuteItemVersion</i>
version tyyppi	<i>eli:version</i>	<i>eli:Version</i>
voimaantulopäivä*	<i>eli:first_date_entry_in_force</i>	<i>xsd:date</i>
voimaantulopäivä**	<i>eli:version_date</i>	<i>xsd:date</i>
kumoamispäivä**	<i>eli:date_no_longer_in_force</i>	<i>xsd:date</i>
asiasanat	<i>eli:is_about</i>	<i>skos:Concept</i>
lainkohdan nojalla säädetty	<i>eli:basis_for</i>	<i>sfl:StatuteVersion*</i>
säädöksen muuttama**	<i>eli:changed_by</i>	<i>sfl:StatuteVersion*</i>
säädöksen kumoama**	<i>sfl:repealedBy</i>	<i>sfl:StatuteVersion*</i>
seuraava pykälä+	<i>sfl:nextItem</i>	<i>sfl:SectionVersion</i>
voimassaolo	<i>eli:in_force</i>	<i>eli:InForce</i>

4.4 Asiakirjojen rakenteellinen RDF-mallinnus



Kuva 10: Luokkakuvaus säädösasiakirjan rakenteesta

Eri lainkohtia ja niiden versioita kuvaavat luokat on määritelty SFL-ontologiassa. *sfl:StatuteItem* jakautuu alaluokkiin, jotka määrittelevät kaikki eri tyyppiset lainkohdat, kuten luku (*sfl:Chapter*), pykälä (*sfl:Section*) ja momentti (*sfl:Subsection*). Vastaavasti *sfl:StatuteItemVersion* jakautuu eri lainkohtien versioita kuvaaviin alaluokkiin, kuten luvun versio (*sfl:ChapterVersion*), pykälän versio (*sfl:SectionVersion*) ja momentin versio (*sfl:SubsectionVersion*). Säädösasiakirjan sisäinen rakenne esi-

tetään versiotasolla, koska dokumentin rakenne voi poiketa eri ajankohtien välillä. Lainkohtien versioita kuvaavat luokat ja asiakirjan hierarkkinen rakenne on esitetty kuvassa 10. Lainkohdan version sisältyminen toiseen lainkohdan versioon on merkitty ominaisuuksilla *eli:has_part* (sisältää lainkohdan) ja *eli:is_part_of* (sisältyy lainkohtaan tai säädökseen). Lisäksi väliotsikon paikka merkitään *sfl:nextItem*-ominaisuudella, joka osoittaa väliotsikkoa seuraavaan pykälään. Itse säädösasiakirjan sisäinen rakenne esiteltiin tarkemmin luvussa 2.2.

Säädöksen tai lainkohdan sisältö ilmaistaan aina manifestaatiotasolla sekä teksti- (*sfl:text*) että HTML-muodossa (*sfl:html*). HTML-muotoisen sisällön ilmaiseminen mahdollistaa sen hakemisen suoraan SPARQL-kyselyllä, jolloin se voidaan näyttää ulkopuolisella verkkosivustolla leijukkeena (engl. widget) [55]. Säädöksen alkuperäisen version sisällöstä tarjotaan myös XML-versio (*sfl:xml*) sellaisena kuin se on saatu Edita Publishing Oy:ltä.

4.5 Säästösten kontekstuaalinen metatieto

Salmisen [66] mukaan kontekstuaalisella metatiedolla tarkoitetaan sisältöyksiköiden ympäristön kuvaamista jossakin tietyssä tilanteessa. Kontekstuaaliseksi metatiedoksi voidaan lukea esimerkiksi sisältöyksikön tapahtumatiedot (säädöksen julkaisupäivä), sisältöyksikön suhteet muihin sisältöyksiköihin (laki muuttaa toista lakia) sekä sisältöyksikön käsittelyprosessiin liittyvät toimijat (lain on säätänyt eduskunta).

Säädöksestä ilmaistaan versioiden välillä muuttumattomat kontekstuaaliset metatiedot abstraktilla tunnistetasolla. Tällaisia metatietoja ovat säädösnumero, säädöksen tyyppi, joka voi olla uusi säädös, muutossäädös tai kumoamissäädös sekä säädöstaso (laki, tasavallan presidentin asetus, valtioneuvoston asetus, ministeriön asetus, valtioneuvoston päätös, muu päätös tai muu säädös). XML-dokumenteissa ei säädöstasoa eikä säädöksen tyyppiä olla eritelty metatiedoiksi, joten ne päätellään säädöksen nimikkeestä. Nimikkeen alusta ilmenee säädöksen tyyppi (esim. ”Valtiovarainministeriön asetus Karl Fazer ja suomalaisuus -juhlarahasta”) ja lopusta, mikäli kyseessä on muutos- tai kumoamissäädös (esim. ”Valtioneuvoston asetus asukkaiden valinnasta arava- ja korkotukivuokra-asuntoihin annetun valtioneuvoston asetuksen muuttamisesta”).

Versiotasolle tallennetaan sellaiset kontekstuaaliset metatiedot, jotka voivat muuttua eri versioiden välillä. Versiotasolla ilmaistaan ominaisuudella *eli:version*, onko kyseessä alkuperäinen (*sfl:Original*) vai muutettu (*sfl:Consolidated*) versio säädöksestä tai lainkohdasta. Alkuperäisissä säädöksissä voimaantulopäivä tunnistetaan voimaantulosäännöksen tekstistä ja ilmaistaan ominaisuudella *eli:first_date_entry_in_force*. Kumotussa säädöksessä kumoamispäivä ilmaistaan ominaisuudella *eli:date_no_longer_in_force*.

Muuteutuissa säädöksissä tai lainkohdissa voimaantulo tarkoittaa viimeisen voimaantulleen muutoksen voimaantulopäivää. Se merkitään versiotasolle ominaisuudella *eli:version_date*. Lainkohtaa muuttanut säädös ilmaistaan ominaisuuksilla *eli:changed_by* ja *eli:changes*. Muutoksen voimaantulopäivä tunnistetaan ajantasaisen dokumentin lopusta löytyvistä muutossäädösten voimaantulotiedoista.

Versiotasolla ilmaistaan tieto säädöksen ja lainkohdan tämän hetkisestä voi-

massaolosta (*eli:in_force*). Säädöksen kumoaminen on lopullinen toimenpide, mutta kumotun lainkohdan tilalle voidaan lisätä uusi lainkohta [58]. Tieto mahdollisesta lainkohdan kumonnesta kumoamissäädöksestä ilmaistaan ominaisuudella *sft:repealedBy*. Linkitys muutos- tai kumoamissäädökseen tehdään aina muutos- tai kumoamissäädöksen alkuperäiseen versioon.

Säädöstekstissä oleva viittaus toiseen säädökseen tai lainkohtaan kuvataan ominaisuuksilla *eli:cites* ja *eli:cited_by*. Viittauksen ilmaisemisessa ei oteta kantaa viitattun säädöksen versioon, vaan linkitys tehdään tunnistetasolle. Viittauksen päivämäärän voi kuitenkin päätellä viittaavan lainkohdan version voimaantulopäivästä.

Säädöksen alkuperäisessä versiossa ilmaistaan myös antopäivä, julkaisupäivä sekä linkitykset säädöksen esityöviittauksissa mainittuihin hallituksen esitykseen sekä mahdollisiin EU-säädöksiin. Linkitykset luodaan käyttäen hallituksen esityksille ja EU-säädöksille ELI-standardin mukaisia tunnisteita.

Hallituksen esityksiä ei vielä ole julkaistu Semanttinen Finlex-palvelussa, mutta niihin viitataan silti ELI-muotoisilla tunnisteilla. Tunniste luodaan vastaavasti kuin säädöksille, paitsi että asiakirjatyypin elementin arvoksi tulee *he*. Säädöksen alkuperäisen version suhde hallituksen esitykseen ilmaistaan ELI-ontologian ominaisuudella *eli:related_to*.

EU-säädösten URI-tunnisteet muodostetaan kaavalla

http://data.europa.eu/eli/{asiakirjatyypin}/{antovuosi}/{säädösnumero}/oj

missä asiakirjatyypin on joko *reg* (asetus, engl. regulation) tai *dir* (direktiivi). Asetuksia sovelletaan EU:n jäsenmaiden oikeuskäytännössä sellaisenaan, kun taas direktiivit sallivat jäsenmaiden itse valita tavan, jolla direktiivin edellyttämät toimenpiteet toteutetaan kansallisessa lainsäädännössä [59]. URI-tunnisteen lopussa oleva *oj* viittaa Euroopan unionin viralliseen lehteen, jossa EU:n säädökset julkaistaan. EU-säädösten URI-tunnisteet ohjaavat EUR-Lex-palveluun, mistä säädökset ovat saatavissa sekä HTML-dokumentteina että RDF-datana. Säädöksen alkuperäisen version suhde EU-säädökseen ilmaistaan ELI-ontologian ominaisuudella *eli:transposes*.

4.6 Säädösten semanttinen metatieto

Semanttisella metatiedolla tarkoitetaan sisällön merkitystä kuvaavaa tietoa [66]. Semanttista metatietoa ovat esimerkiksi asiakirjan nimike, aihe, tiivistelmä sekä asiakirjaan liittyvät asiasanat. Säädösten semanttinen metatieto koostuu nimikkeestä, mahdollisesta lempinimestä sekä asiasanoituksesta.

Säädöksen nimike ja lempinimi ovat kielikohtaisia ja ilmaistaan ekspressiotasolla ominaisuuksilla *eli:title* ja *eli:title_alternative*. Lainkohdan tapauksessa *eli:title*-ominaisuudella ilmaistaan otsikko. Asiasanoitus voi muuttua versoiden välillä mutta ei ole kieliriippuvainen. Siksi asiasanat ilmaistaan versiotasolla ominaisuudella *eli:is_about*.

Säädösten asiasanoitus on sama kuin Finlex-palvelussa ja perustuu vanhaan Suomen säädöskokoelman asiasanastoon, jota Edita Publishing Oy ylläpitää [76]. Asiasanasto sisältää toisaalta hyvin yleisiä asiasanoja (kuten *metsä*) ja toisaalta

hyvin spesifejä (kuten *Tammisaaren saariston kansallispuisto*) eikä asiasanojen välille ole määritetty hierarkiaa. Jokaiselle asiasanalle löytyy myös ruotsinkielinen käännös.

Asiasanaston mallintamisessa linkitetyn datan teknologioita hyödyntäen voidaan käyttää SKOS-sanastokielen (Simple Knowledge Organization System) [52] määrittämiä. SKOS mahdollistaa ontologiaperustaisen tiedonhaun aineistosta [35]. SKOS-kielellä voidaan kuvata käsitteiden välisiä hierarkkisia ja assosiatiivisia suhteita RDF-perustaisesti.

Koska Finlex-palvelun asiasanastoon ei olla rakennettu hierarkiaa eikä merkitty assosiatiivisia suhteita, on asiasanaston SKOS-muunnoksessa ainoastaan kuvattu kaikki käsitteet yläkäsitteinä (ominaisuus *skos:topConceptOf*). Lopuksi SKOS-muunnettu asiasanasto on vielä validoitu Skosify-sovelluksella [73].

Käsitteiden välisten suhteiden puuttumisen lisäksi toinen säädösten asiasanoitukseen liittyvä ongelma on, että asiasanoitus ei ota kantaa siihen, missä lainkohdassa asiasanan mukaista teemaa käsitellään. Esimerkiksi rikoslaki on asiasanoitettu sanalla *metsä*, mutta metsärikoksesta puhutaan ainoastaan 48 a luvun 3 §:ssä. Siten tiedonhakua ei pystytä asiasanoituksen perusteella kohdentamaan yksittäiseen pykälään. Aineistojen automaattinen semanttinen annotointi [71] voisi jatkossa tarjota ratkaisun yksittäisten pykälien asiasanoittamiseksi.

5 Oikeuden ratkaisujen tietomalli

Tässä luvussa kuvataan Semanttinen Finlex -palvelun ECLI-standardiin perustuvat oikeuden ratkaisujen URI-käytännöt sekä RDF-tietomalli. Sen jälkeen tarkastellaan, miten nimentunnistusta käytetään oikeuden ratkaisujen metatietojen rikastamiseen. Lopuksi käsitellään vielä aineistoihin liittyvät asiasanastot.

5.1 Ratkaisujen URI-tunnisteet

Oikeuden ratkaisujen URI-tunnisteet ovat muotoa

$$/oikeus/{ECLI-tunniste}/{kieli}/{formaatti}$$

Esimerkiksi korkeimman oikeuden vuoden 2016 55:nmen ennakkopäätöksen suomenkielisen HTML-version URI on */oikeus/ECLI:FI:KKO:2016:55/fin/html*. URI-elementit on luetteloitu taulukossa 9.

Taulukko 9: Oikeuden ratkaisujen URI-elementit

Elementti	Muoto	Arvojoukko
ECLI	ECLI-tunniste	
Kieli	ISO 639-2	<i>fin, swe</i>
Formaatti	merkkijono	<i>html, txt, xml</i>

Oikeuden ratkaisujen URI-käytäntöjen yhtenäistämiseksi säädösasiakirjojen kanssa voisi olla perusteltua ECLI-tunnuksen käytön sijaan erottaa ECLI:n elementit toisistaan URI-tunnisteessa kauttaviivoilla. Silloin oikeuden ratkaisujen URI-tunnisteet olisivat muotoa:

$$/oikeus/{oikeuslaitos}/{antovuosi}/{ratkaisun_tunniste}/{kieli}/{formaatti}$$

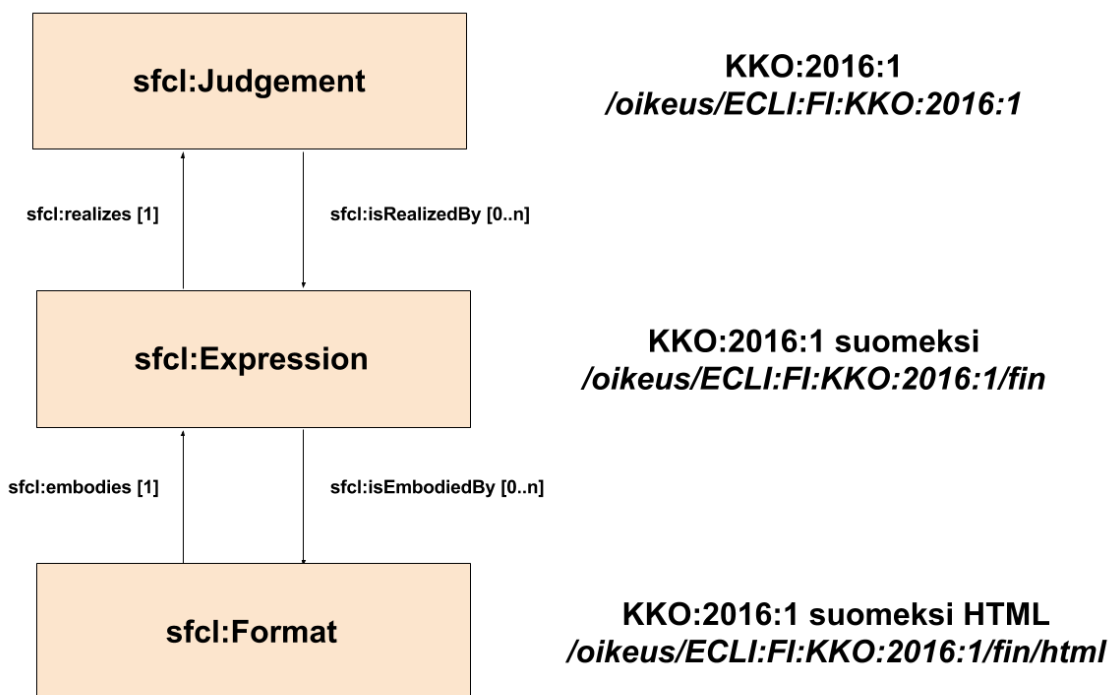
Semanttinen Finlex ei kuitenkaan toistaiseksi tue tätä URI-muotoa.

5.2 Tietomallin luokat ja ominaisuudet

Jotta oikeuden ratkaisujen tietomalli olisi yhtenäinen säädösten tietomallin kanssa, FRBR-kuvailumallia on sovellettu myös oikeuden ratkaisuihin. Yksittäisestä ratkaisusta ei kuitenkaan ole olemassa kuin yksi sisällöltään muuttumaton versio suomeksi ja ruotsiksi. Näin ollen neljättä kuvailutasoa ei oikeuden ratkaisujen mallintamiseen tarvita. Palvelua varten räätälöity SFCL-ontologia (Semantic Finlex Case Law) määrittelee kolme FRBR-kuvailumallin mukaista luokkaa, jotka ovat:

- *sfcl:Judgement* (Oikeuden ratkaisu) vastaa FRBR-mallissa teostasoa

- *sfcl:Expression* (Ekspressio), joka kuvaa oikeuden ratkaisun kieliversiota, vastaa FRBR-mallissa ekspressiotasoa
- *sfcl:Format* (Formaatti), joka kuvaa oikeuden ratkaisun kieliversiota tietyssä sisältömuodossa, vastaa FRBR-mallissa manifestaatiotasoa



Kuva 11: SFCL-ontologian mukainen FRBR-kuvailu ja oikeuden ratkaisu KKO:2016:1 esimerkkinä.

Kuva 11 havainnollistaa SFCL-ontologian FRBR-mallia sekä URI-käytäntöä eräällä korkeimman oikeuden ratkaisulla. Eri kieliversioiden ja sisältömuotojen URI-tunnisteet muodostetaan vastaavasti kuin ELI-perustaiset säädösten URI-tunnisteet.

ECLI:n määrittelemät Dublin Core -perustaiset metatietokentät on jaettu FRBR-kuvailumallin tasojen kesken siten, että kieliriippuvaiset metatiedot tulevat ekspressiotasolle ja kieliriippumattomat teostasolle. Suurin osa metatiedoista saadaan suoraan XML-rakenteesta, kuten ratkaisun ECLI-tunniste, esittelypäivä, antopäivä, tiivistelmä ja oikeuden ala. Eri SFCL-ontologiassa määriteltyjen luokkien ominaisuudet on esitetty taulukoissa 11, 12 ja 13. Ominaisuudet ja niiden kuvaama data käsitellään tarkemmin seuraavissa aliluvuissa. Käytetyt nimiavaruudet on listattu taulukossa 10.

Ratkaisuasiakirjan osiin ei viitata samalla tavalla kuin lainkohtiin, joten ratkaisujen sisäistä rakennetta ei pilkota RDF-tietomallissa. Koko ratkaisun teksti ilmaistaan manifestaatiotasolla XML-, HTML- ja tekstimuodoissa. XML-muoto vastaa täysin Edita Publishing Oy:ltä saatua XML-dokumenttia.

Taulukko 10: Oikeuden ratkaisujen mallinnuksessa käytetyt nimiavaruudet

Lyhenne	Nimiavaruus
dcterms	http://purl.org/dc/terms/
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#
xsd	http://www.w3.org/2001/XMLSchema#

Taulukko 11: *sfcl:Judgement*-luokan ominaisuudet

Ominaisuus	RDF-tunniste	Arvojoukko
ECLI-tunniste	<i>dcterms:isVersionOf</i>	<i>rdfs:Literal</i>
asian ratkaisijat	<i>dcterms:contributor</i>	<i>dcterms:Agent</i>
tuomioistuimen kotimaa	<i>dcterms:coverage</i>	<i>dcterms:Location</i>
tuomioistuin	<i>dcterms:creator</i>	<i>dcterms:Agent</i>
asiasanat	<i>dcterms:description</i>	<i>skos:Concept</i>
oikeuden ala	<i>dcterms:subject</i>	<i>skos:Concept</i>
ratkaisun päivämäärä	<i>dcterms:date</i>	<i>xsd:date</i>
ratkaisun julkaisupäivä	<i>dcterms:issued</i>	<i>xsd:date</i>
viittaus lainsäädäntöön	<i>sfcl:refToLegislation</i>	<i>sfl:Statute</i>
ekspressio	<i>sfcl:isRealizedBy</i>	<i>sfcl:Expression</i>

Taulukko 12: *sfcl:Expression*-luokan ominaisuudet

Ominaisuus	RDF-tunniste	Arvojoukko
ratkaisun nimi	<i>dcterms:title</i>	<i>rdfs:Literal</i>
ratkaisun kieli	<i>dcterms:language</i>	<i>rdfs:Literal</i>
tiivistelmä	<i>dcterms:abstract</i>	<i>rdfs:Literal</i>
teos	<i>sfcl:realizes</i>	<i>sfcl:Judgement</i>
manifestaatio	<i>sfcl:isEmbodiedBy</i>	<i>sfcl:Format</i>

Taulukko 13: *sfcl:Format*-luokan ominaisuudet

Ominaisuus	RDF-tunniste	Arvojoukko
sisältömuoto	<i>dcterms:format</i>	URI
tekstisisältö	<i>sfcl:text</i>	<i>rdfs:Literal</i>
HTML-sisältö	<i>sfcl:html</i>	<i>rdfs:Literal</i>
XML-sisältö	<i>sfcl:xml</i>	<i>rdf:XMLLiteral</i>
ekspressio	<i>sfcl:embodies</i>	<i>sfcl:Expression</i>

5.3 Nimien tunnistaminen ratkaisun tekstistä

Ratkaisujen metatietomalli määrittelee joitakin metatietoja, joita oikeuden ratkaisujen XML-dokumenteista ei löydy suoraan, vaan ne löytyvät ainoastaan ratkaisun sisältötekstin seasta. Tällaisia metatietoja ovat viittaukset lainsäädäntöön sekä KKO:n ennakkopäätöksissä asian ratkaisseet oikeusneuvokset.

Nimentunnistuksella (NER, Named Entity Recognition) tarkoitetaan nimettyjen entiteettien, kuten henkilöiden, paikkojen ja päivämäärien, tunnistamista tekstistä. Dozier et al. [20] esittävät kolme menetelmää tunnistaa nimettyjä entiteettejä oikeudellisista teksteistä:

1. **Nimien listaaminen ennalta.** Tekstistä tunnistetaan ennalta listatut nimet. Menetelmä on yksinkertainen toteuttaa, mutta tuottaa helposti vääriä positiivisia tuloksia, jos listatuilla nimillä on useita merkityksiä. Jos taas lista ei ole riittävän kattava, menetelmä tuottaa vääriä negatiivisia tuloksia.
2. **Asiayhteydelliset säännöt.** Asiayhteydestä tekstissä voi usein päätellä nimen esiintymisen siinä. Esimerkiksi lain voimaantulosta säädettyä voimaantulosäännöksessä lukee ”Tämä laki tulee voimaan” ennen päivämäärää, minkä perusteella voidaan olettaa kyseisen virkkeen sisältävän voimaantulopäivän. Menetelmällä voidaan usein saavuttaa parempi tarkkuus kuin pelkällä nimien ennalta listaamisella. Menetelmän haittana on, että sääntöjen laatiminen ja ylläpitäminen voivat edellyttää runsaasti käsityötä.
3. **Tilastolliset menetelmät.** Menetelmissä valitaan jokin tilastollinen malli, jonka parametrit koneoppimissovellus oppii ennalta käsin annotoidun opetusdatan perusteella. Myös tilastollisilla menetelmillä voidaan saavuttaa hyvä tarkkuus, mutta menetelmien käyttö edellyttää käsin annotoitua opetusaineistoa ja usein myös vahvaa asiantuntemusta koneoppimisesta.

Asiakirjaviitteiden ja oikeusneuvosten nimien tunnistamiseksi KKO:n ja KHO:n ratkaisuista käytetään menetelmää numero 2. Asian ratkaisijat on KKO ratkaisuissa lueteltu asiakirjan lopussa erillisessä kappaleessa, joka alkaa sanoilla ”Asian ovat ratkaisseet”. Tätä seuraavat asian ratkaisseiden oikeusneuvosten nimet, jotka tunnistetaan. Vanhemmissa ratkaisuissa kappale alkaa sanoilla ”Ratkaisuun osallistuneet” ja oikeusneuvokset on kirjoitettu tekstiin pelkillä sukunimillä. Kappaleen lopussa esiintyvä asian esittelijä jätetään tunnistamatta. Esittelijän esiintyminen voidaan päätellä siitä, että ennen esittelijän nimeä tekstissä lukee joko ”esittelijä” tai ”esittelijänä”.

Nimen linkityksellä (NEL, Named Entity Linking) tarkoitetaan nimen liittämistä solmuun tietämuskannassa [33]. Jokaisesta oikeusneuvoksesta luodaan oma RDF-resurssi, johon oikeuden ratkaisun *dcterms:contributor*-ominaisuus osoittaa. KKO:n oikeusneuvosten nimien oletetaan olevan yksikäsitteisiä heidän vähäisestä määrästäan johtuen.

Viittaukset säädöksiin tunnistetaan ratkaisuista eristämällä säädösnumeroita ratkaisutekstistä. Säädösviittauksista poimitaan tällä hetkellä vain muotoa (*{säädösnumero}/{antovuosi}*) olevat. Viittausten asiayhteydestä ei päätellä mitään, joten

kaikki viittaukset merkitään yleisellä ominaisuudella *sfcl:refToLegislation* (viittaus lainsäädäntöön). Myöskään säädöksen versiota ei voida määrittää, koska rikoksen tekohetkestä ei ole tietoa. Tämän vuoksi kaikki viittaukset on tehty säädösten tunnistetasolle (luokka *sfl:Statute*).

Myös säädösten nimikkeitä, nimikkeiden lyhenteitä sekä viittauksia lainkohtiin tullaan jatkossa tunnistamaan ratkaisujen teksteistä. Säädösten nimikkeiden tunnistamiseksi ja linkittämiseksi voidaan käyttää RDF-perustaisten aineistojen automaattiseen annotointiin kehitettyä ARPA-palvelua [54]. ARPA mahdollistaa ratkaisutekstin perusmuotoistamisen ja yhteensopivien käsitteiden hakemisen SPARQL-palvelupisteestä perusmuotoistamisella tuotetuille n-grammeille.

5.4 Asiasanastot

Sekä KKO:n että KHO:n ratkaisuihin on käytössä omat asiasanastonsa. ECLIMetatietomallin mukaisesti oikeuden ratkaisujen asiasanoitus ilmaistaan Dublin Core-sanaston ominaisuudella *dcterms:description*.

KHO:n asiasanastoa ei ole erillisenä sanastorakenteena, joten ratkaisujen XML-dokumenteissa kuvatuista asiasanoista luodaan SKOS-käsitteitä ratkaisujen RDF-muunnoksen yhteydessä. Ratkaisujen asiasanoituksesta ei ilmene, että sanastossa olisi hierarkiaa, joten kaikki käsitteet kuvataan yläkäsitteinä. KKO:n sanasto sen sijaan on hierarkkinen ja se käännetään SKOS-sanastoksi erillisestä XML-dokumentista, jossa asiasanaston hierarkia on kuvattu XML-puurakenteena.

6 Datamuunnos ja tiedonhallinta

Tässä luvussa esitellään Semanttinen Finlex -hankeessa toteutettu sovellus, joka muuntaa XML-muotoiset säädökset sekä oikeuden ratkaisut RDF-muotoon ja päivittää julkaisun aineistoa. XML-dokumenttien sisältämistä virheistä sekä konversioprosessin virhealttiudesta johtuen muunnettua dataa pyritään validoimaan ennen julkaisua. Luvun lopussa esitellään datamuunnoksen ja tiedonhallinnan ratkaisemattomat haasteet.

6.1 XML-RDF-muunnin

XML-RDF-muunnin muuntaa Edita Publishing Oy:n toimittamat säädös- ja oikeustapauskirjoitukset XML-asiakirjoista RDF-perustaiseksi dataksi. Muunnin on kirjoitettu Scala-ohjelmointikielellä [21] käyttäen Xpath-kieltä [12] XML-dokumenttien käsittelyyn sekä Apache Jena -kirjastoa [77] RDF-datan tuottamiseen. Muunnin tukee ajantasaisten säädösten XML-rakennetta, alkuperäisten säädösten XML-rakennetta sekä oikeuden ratkaisujen ECLI-perusteista XML-rakennetta.

Ensiksi muunnin kääntää kaikki alkuperäiset säädökset RDF-muotoon. Tämän jälkeen muunnin kääntää kaikki ajantasaisten säädösten versiot aikajärjestyksessä vanhimmasta uusimpaan. Ajantasaista lakia pystytään näin keräämään alkuperäisen päälle siten, että ajantasaistetuista versiosta muunnetaan vain ne lainkohdat, joiden sisältö on edellisen version jälkeen muuttunut. Oikeuden ratkaisujen käänno-prosessi ei ole riippuvainen lainsäädäntöaineistosta, joten se voidaan tehdä missä vaiheessa tahansa.

Yksittäistä dokumenttia kääntäessään muunnin etenee dokumenttipuussa sen alusta loppuun kutsuen omaa muunnosmetodiaan rekursiivisesti siten, että myös dokumentin sisäkkäiset rakenteet tulevat erikseen käännettyiksi RDF-muotoon. Jokaisen lainkohdan sisältö tallennetaan sekä HTML- että tekstimuodossa myös osaksi ylempien lainkohtien sisältöä. Ajantasainen säädös saattaa sisältää useamman version samasta lainkohdasta, mutta ainoastaan voimassaoleva versio tallennetaan ylemmän lainkohdan version osaksi. Hylätyt versiot voivat olla aiemmin voimassaolleita tai myöhemmin voimaantulevia. Lainkohdan voimassaolo määritetään muutossäädöksen voimaantulopäivän sekä ajantasaistetun dokumentin julkaisupäivän perusteella.

Tarkastellaan esimerkkinä kuvitteellista säädöstä, jonka säädösnumero on 123/2015 ja joka sisältää johtolauseen sekä kaksi pykälää, joiden kummankin sisällä on kaksi momenttia. Johtolause on alkuperäinen, kuten 2 § 1 momenttikin. 1 § 1 momentin viimeisin muutos on tullut voimaan 20.6.2005. 1 § 2 momentin on tullut voimaan jo aiemmin 15.3.2003. 2 § 2 momentin muutos on tullut voimaan molempia myöhemmin 21.9.2010. Taulukossa 14 on luetteloitu järjestys, jossa muunnin muuntaisi kyseisen säädöksen lainkohdat RDF-muotoon. Taulukon kolmanteen sarakkeeseen on määritelty säädökselle ja jokaiselle lainkohdalle ELI-muotoinen URI, jonka muunnin luo säädöksen tai lainkohdan versiota kuvaavalle resurssille.

Ajantasaistettua säädöstä muunnettaessa lainkohtien uusin voimaantulopäivä kopioidaan aina ylöspäin dokumentin rakenteessa versiopäiväksi (*eli:version_date*) myös ylemmille lainkohdille. Koska taulukon esimerkin säädöksen 1 § 1 momentti on

Taulukko 14: XML-RDF muuntimen toiminta

Järj.	Lainkohta	URI-tunniste
1	säädös	-
2	johtolause	/eli/sd/2015/123/johtolause/alkup
3	1. §	-
4	1. § 1. mom	/eli/sd/2015/123/pykala/1/momentti/1/ajantasa/20050620
5	1. § 2. mom	/eli/sd/2015/123/pykala/1/momentti/2/ajantasa/20030315
6	1. §	/eli/sd/2015/123/pykala/1/ajantasa/20050620
7	2. §	-
8	2. § 1. mom	/eli/sd/2015/123/pykala/2/momentti/1/alkup
9	2. § 2. mom	/eli/sd/2015/123/pykala/2/momentti/2/ajantasa/20100921
10	2. §	/eli/sd/2015/123/pykala/2/ajantasa/20100921
11	säädös	/eli/sd/2015/123/ajantasa/20100921

muuttunut myöhemmin kuin 1 § 2 momentti, merkitään myös 1 §:n voimaantulopäiväksi 1 § 1 momentin voimaantulopäivä. Koska 2 §:n sisällä on vielä myöhemmin voimaantullut momentti kuin 1 § 1 momentti, tulee itse säädöksen versiopäiväksi kuitenkin 2 § 2 momentin voimaantulopäivä. Pykälien tai pykälän sisällä olevien momenttien voimaantulopäivillä ei ole vaikutusta toisiinsa.

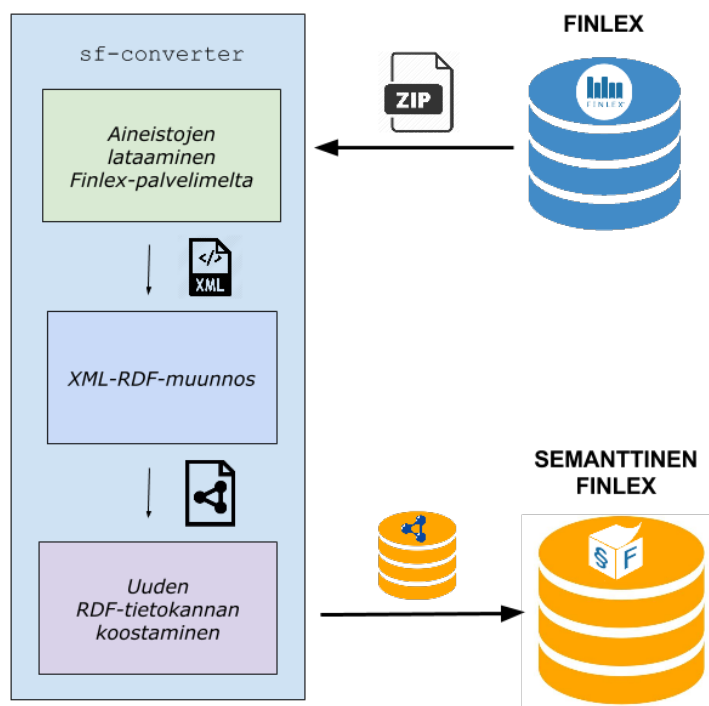
XML-RDF-muunnin on julkaistu avoimena lähdekoodina²⁰. Ajantasaisen lainsäädäntöaineiston käyttöehdot eivät kuitenkaan mahdollista ajantasaisten säädösten XML-dokumenttien julkaisemista, joten sovelluskoodin käyttäjä joutuu ajamaan XML-RDF-muunnokseen ilman ajantasaisen lainsäädännön kokoelman aineistoja.

6.2 Aineiston päivitys

Finlex-palvelun aineistoa päivitetään viikottain uusilla säädöksillä ja oikeuden ratkaisuilla. Siksi myös Semanttinen Finlex -palvelun RDF-aineistoa päivitetään säännöllisesti uusilla aineistoilla. Päivitysprosessi jakautuu neljään päävaiheeseen:

- **Aineiston lataaminen Finlex-palvelimelta.** Ensimmäiseksi ajetaan latausskripti. Se lataa Finlex-palvelimelta kaikki uudet XML-dokumentit, jotka eivät jo löydy paikalliselta levyltä, ja järjestää ne aineistokohtaisesti hakemistoihin julkaisupäivämäärän mukaan.
- **XML-RDF-muunnos.** Kaikki paikalliset XML-tiedostot käännetään RDF-dokumenteiksi julkaisupäivän mukaisessa järjestyksessä XML-RDF-muuntimella.
- **RDF-dokumenttien kokoaminen tietokannaksi.** Uusi RDF-tietokanta kootaan muuntimella luoduista RDF-dokumenteista *tdbloader*-työkalulla. *tdbloader* on Apache Jena -paketin mukana tuleva apusovellus, jonka avulla suuri määrä RDF-aineistoja voidaan ladata RDF-tietokantaan kerralla.

²⁰<http://github.com/apoksane/sf-converter>



Kuva 12: Semanttinen Finlex -datajulkaisun päivitysprosessi

- **Aineiston vieminen julkaisuun.** Julkaisussa oleva vanha RDF-tietokanta korvataan päivitetyllä versiolla.

Päivitysprosessia on havainnollistettu kuvassa 12. Toistaiseksi päivitys tehdään käsitönnä, mutta jatkossa se on tarkoitus automatisoida täysin. Jatkossa koko aineistoa ei tulla ajamaan joka kerta muuntimen läpi, vaan ainoastaan uudet säädösasiakirjat ja oikeuden ratkaisut käännetään RDF-muotoon ja ladataan julkaisussa olevaan RDF-tietokantaan.

6.3 Validointi

XML-RDF-muunnoksen onnistumista validoidaan testiaineistolla sekä ohjelmallisilla testeillä. Testeissä pieni määrä XML-asiakirjoja käännetään RDF-dokumenteiksi, joihin kohdistetaan SPARQL-kyselyitä. Kyselyssä voidaan esimerkiksi tarkistaa yksittäisen lainkohdan voimaantulopäivä tai voimassaolo. Testit rajoittuvat hyvin pieneen osaan XML-aineistoa ja noin kolmeenkymmeneen eri SPARQL-kyselyyn. Testien laajentaminen onkin yksi hankkeen seuraavista kehitystarpeista. Jatkossa konversioprosessin yhteyteen tulisi kehittää nykyistä laajempi sääntöpohjainen validointijärjestelmä.

Myös kokonaista käännettyä aineistoa voidaan validoida kohdistamalla siihen SPARQL-kyselyitä, jotka poimivat datasta mahdolliset virheet. Näitä SPARQL-kyselyitä voidaan muodostaa eri sääntöjen perusteella. Esimerkki validointisäännöstä on, että voimaantulopäivä ei voi edeltää julkaisupäivää, koska lain mukaan julkaiseminen on edellytys lain voimaantulolle (Perustuslaki 79 §:n 3 momentti).

Kaikkien erikoistapausten löytäminen ja niistä syntyvien virheiden korjaaminen on dataa käsiteltäessä osoittautunut haastavaksi. Aineiston validoimiseksi tulisi kehittää yleispätevien sääntöjen kokoelma, jonka avulla voitaisiin automaattisesti validoida jokainen säädös ja oikeuden ratkaisu ennen RDF-tietokannan päivitystä. RDF-muotoisen tiedon validoimiseksi kehitettyjä sääntö- ja rajoitekieliä, kuten SHACL (Shapes Constraint Language) [45] tai SPIN (SPARQL Inferencing Notation) [44] on myös syytä tutkia aineiston validoimiseksi tulevaisuudessa.

6.4 Datamuunnoksen haasteet ja puutteet

Kuten aiemmin on todettu, liittyy datamuunnoksen tekemiseen tiedonhallinnallisia haasteita. Osa haasteista johtuu virheistä ja puutteellisista metatiedoista alkuperäisessä datassa, osa lainsäädännön laatimis- ja julkaisuprosessien monimutkaisuudesta ja osa asioista, joita ei olla vielä huomioitu täysin tietomallissa ja datamuunnoksessa, kuten jotkin versioiden välillä liikkuneet lainkohdat.

Alkuperäisten säädösten XML-rakenteessa ei juuri ole tarjolla metadataa. Tämä vaikeuttaa olennaisesti esimerkiksi muutossäädösten käsittelyä, koska johtolauseen tekstistä pitäisi pystyä jäsentämään muutettavien lainkohtien tunnistet, mitä ei olla tehty ainakaan toistaiseksi. Siksi tiedot muutoksista tunnistetaan ainoastaan ajantasaisten säädösten dokumenteista. Myös voimaantulopäivän jäsentäminen voimaantulosäännöksestä aiheuttaa vaikeuksia, koska saman säädöksen eri lainkohdat voivat tulla voimaan eri aikoina, kuten luvussa 2 esitettyssä kuvassa 3 ilmenee.

Myös asiakirjojen XML-rakenteissa esiintyvät virheet aiheuttavat haasteita muunnoksen tekemisessä. Esimerkiksi kohdat on joissakin tapauksissa merkitty virheellisesti alakohdiksi. Lainkohdan tyyppi pitää silloin tarkistaa sisällön perusteella. Jos alakohdaksi merkityn lainkohdan teksti alkaa numerolla, kyseessä on kohta.

III OSA
Otsikko on kumottu L:lla [20.3.2015/303](#). ([20.3.2015/303](#))
9 luku
Vastuuvelka kirjanpidossa ([20.3.2015/303](#))

Kuva 13: Vakuutusyhtiölain kolmannen osan kumottu otsikko (kuva: Finlex)

III OSA
VAKAVARAIJUUS ([20.3.2015/303](#))
10 luku ([20.3.2015/303](#))
Varojen ja velkojen arvostaminen ja vastuuvelka
vakavaraisuuslaskennassa

Kuva 14: Vakuutusyhtiölain kolmannen osan uusi otsikko (kuva: Finlex)

Myös ajantasaisen säädöskokoelman XML-rakenteen ja toimituskäytäntöjen dokumentoinnin puuttuminen tekee datamuunnoksen onnistuneesta tuottamisesta haas-

teellista. Tarkastellaan esimerkkinä Vakuutusyhtiölain (521/2008) voimassaolevaa ajantasaista versiota päivältä 1.1.2016. Kyseisessä versiossa on kolmannen osan otsikko kumottu ja siirretty sen paikkaan 9 luvun edestä 10 luvun eteen (kuvat 13 ja 14). Paikan siirtymisestä ei kuitenkaan olla merkitty tietoa III osaan, vaan ainoastaan sen otsikkoon on merkitty tieto otsikon kumoamisesta. 10 luvun edelle on merkitty toisen kerran III osa ja sille otsikko, joka tulee voimaan samalla säädöksellä kuin sitä edellinen on kumottu. Tässä tilanteessa tulee tietää, että osan otsikon kumoaminen tarkoittaa, että osan paikka muuttuu 9 luvun edeltä 10 luvun edelle ja 9 luku siirtyy siten II osan alle. Tätä testataan XML-RDF-muunnoksen jälkeen ohjelmallisella testillä, joka hakee SPARQL-kyselyllä molemmista versioista 9 luvun ja tarkistaa, mihin osiin ne kuuluvat.

58 b § (17.12.2015/1546)

Asuntovelan koron vähennyskelpoisuus verovuosina 2016–2018

Poiketen siitä, mitä 58 §:n 2 momentissa säädetään, verovelvollisella on oikeus vähentää pääomatuloistaan 55 prosenttia asuntovelan koroista verovuonna 2016, 45 prosenttia verovuonna 2017 ja 35 prosenttia verovuonna 2018.

L:lla [1546/2015](#) muutettu 58 b § on voimassa 1.1.2016–31.12.2018. Aiempi sanamuoto kuuluu:

58 b § (12.12.2014/1086)

Asuntovelan koron vähennyskelpoisuus verovuosina 2015–2017

Poiketen siitä, mitä 58 §:n 2 momentissa säädetään, verovelvollisella on oikeus vähentää pääomatuloistaan 65 prosenttia asuntovelan koroista verovuonna 2015, 60 prosenttia verovuonna 2016 ja 55 prosenttia verovuonna 2017.

L:lla [1086/2014](#) lisätty 58 b § on väliaikaisesti voimassa 1.1.2015–31.12.2017.

58 b§ (29.12.2011/1515)

Asuntovelan koron vähennyskelpoisuus verovuosina 2012 ja 2013

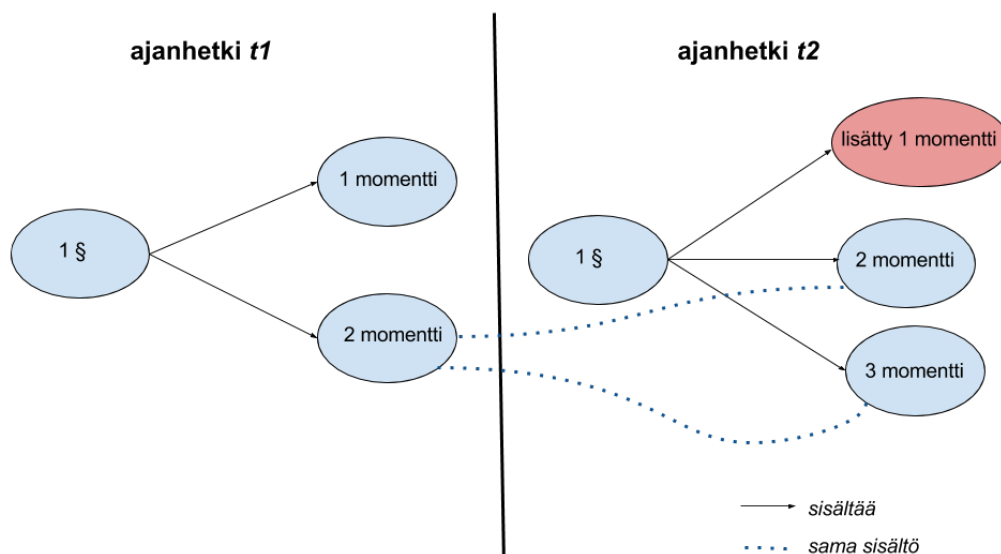
Poiketen siitä, mitä 58 §:n 2 momentissa säädetään, verovelvollisella on oikeus vähentää pääomatuloistaan 85 prosenttia asuntovelan koroista verovuonna 2012 ja 80 prosenttia verovuonna 2013.

L:lla [1515/2011](#) lisätty 58 b § oli väliaikaisesti voimassa 1.1.2012–31.12.2013.

Kuva 15: Tuloverolain (1535/1992) 3 luvun 58 b §:n kolme eri versiota (kuva: Finlex)

Toinen tiedonhallinnollinen haaste on, että ajantasainen dokumentti saattaa sisältää useita eri versioita samasta lainkohdasta. Kuvassa 15 on esimerkki Tuloverolain ajantasaisesta versiosta, joka sisältää kolme eri aikoina väliaikaisesti voimassa olevaa versiota samasta pykälästä. Koska dokumenttia muunnettaessa tallennetaan lainkohdasta vain julkaisuhetkellä voimassa ollut versio dokumentin osaksi, saattavat edellinen ja tuleva versio jäädä kokonaan julkaisusta pois. Vanhempaa versiota ei

voida liittää osaksi vanhempaa versiota ylemmästä rakenteesta, koska ei välttämättä tiedetä, mitä ylempi rakenne sisälsi aiemmin. Toisaalta tulevaisuuttakaan ei voida ennustaa, koska säädös saattaa muuttua milloin tahansa, joten tulevaakaan versiota ei tallenneta. Jos dokumenttia ei päivitetä Edita Publishing Oy:n toimesta ennen lainkohdan uuden version voimaantuloa, jää uusi versio lainkohdasta kokonaan pois julkaisusta. Tämä voitaisiin ratkaista esimerkiksi siten, että muunninsovellus kerää listaa myöhemmin voimaantulevista lainkohdista ja muuntaa kyseisestä säädöksestä lainkohdan voimaantulopäivänä uuden version, jos uutta ajantasaista dokumenttia ei olla julkaistu siihen mennessä.



Kuva 16: Pykälän alkuun lisätty momentti

Myös lainkohtien väliin lisätyt uudet lainkohdat aiheuttavat haasteita. Tarkastellaan esimerkkinä kuvitteellista tilannetta, jossa ajantetkellä t_1 1 § sisältää kaksi momenttia ja ajantetkellä t_2 1 §:ään on lisätty uusi 1 momentti. Kuva 16 havainnollistaa esimerkkiä. Tällä hetkellä XML-RDF-muunnin ei huomioi kahden momentin liikkumista rakenteessa, koska liikkumista ei olla merkitty muutoksena ajantasaisen dokumentin XML-rakenteeseen. Ajantetkellä t_2 XML-RDF-muunnin merkitsee 2 momentin sekä 3 momentin versiot virheellisesti samalla versiopäivämäärällä kuin ajanhetkellä t_1 . Siten ajanhetken t_2 2 momentti merkitään virheellisesti samaksi kuin ajanhetken t_1 2 momentti. 3 momentista tallennetaan kokonaan uusi versio samalla versiopäivällä ja samalla sisällöllä kuin 2 momentti ajanhetkellä t_1 . Lopputuloksena on, että 1 § ajanhetkellä t_2 sisältää lisätyn 1 momentin lisäksi kaksi sisällöltään ajanhetken t_1 2 momenttia vastaavaa momenttia.

Ratkaisuksi tähän ongelmaan voisi soveltaa Metalex Document Server-palvelussa [38] käytössä olevaa menetelmää, jossa lainkohtien liikkuminen havaitaan datamuunnosvaiheessa vertailemalla lainkohtien sisältöjä toisiinsa eri versioiden välillä. Palvelussa lainkohdasta luodaan kaksi resurssia, joista toinen ilmaisee lainkohdan paikan rakenteessa tietyllä ajanhetkellä ja toinen sen sisällön ottamatta kantaa paik-

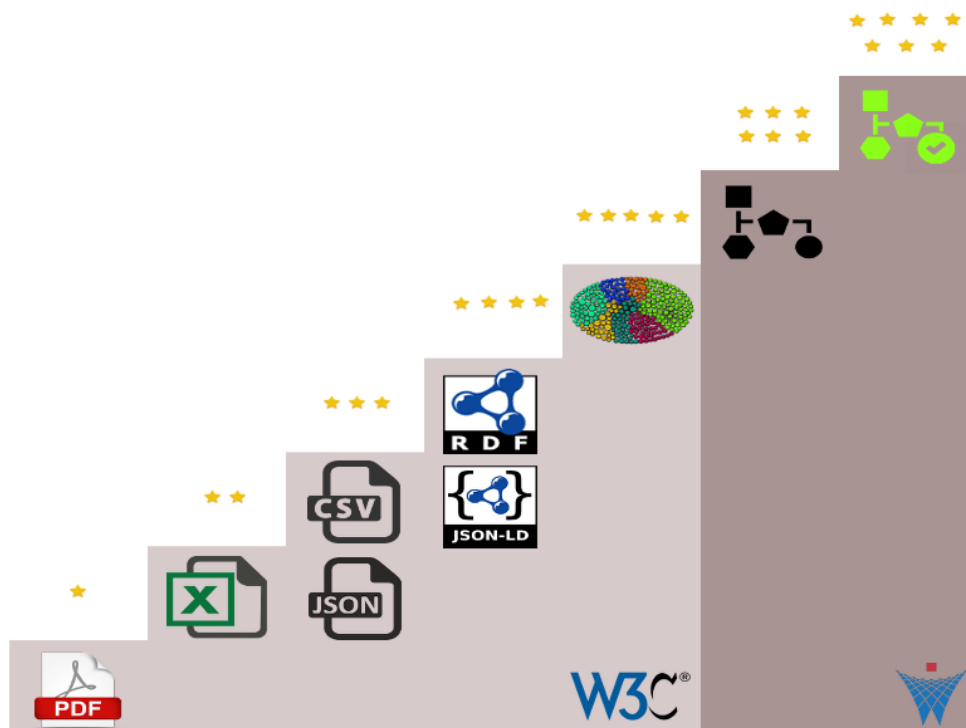
kaan rakenteessa. Tietyssä paikassa tietyllä ajanhetkellä oleva lainkohta voidaan siten liittää tiettyyn sisältöön.

Koska kaikkia versioita ajantasaisesta dokumentista ei ole saatavilla, ei ajanhetken t_2 2 ja 3 momenttien URI-tunnisteille osattaisi välttämättä muodostaa 1 momentin muutoksen voimaantulopäivän perusteella oikeaa versiopäivää. 1 momentti voi olla muuttunut jo useasti lisäyksensä jälkeen, ja momentista poimittava viite viittaa vain uusimpaan muutossäädökseen ja sen voimaantulopäivään. Vanhoissa muutossäädöksissä voimaantulopäivä ei aina ole edes tiedossa, jolloin version URI-tunniste muodostetaan muutossäädöksen antopäivän perusteella. Tämä perustuu ainoastaan oletukseen siitä, että antopäivä todennäköisesti on ollut lähellä voimaantulopäivää. Ongelman täydellinen ratkaiseminen vaatisi, että kaikkien muutosten voimaantulopäivät olisivat selvillä ja kaikki versiot ajantasaistetusta säädöksestä olisivat saatavilla.

7 Datajulkaisu

Tässä luvussa arvioidaan Semanttinen Finlex -palvelun lopullisen datajulkaisun laatua linkitetyn datan seitsemän tähden julkaisumallin kriteereillä sekä käydään läpi palvelun käyttöehdot. Lopuksi esitellään datapalvelun tarjoamat ohjelmointirajapinnat.

7.1 Kuuden tähden datajulkaisu



Kuva 17: Seitsemän tähden malli linkitetyn avoimen datan julkaisulle

Berners-Lee [4] on esittänyt viiden tähden laadunarviointimallin avoimen datan julkaisulle. Mallin perusteella datajulkaisu sijoitetaan viisitasoiseen arvoasteikkoon seuraavasti:

- Yhden tähden julkaisussa data on saatavilla avoimesti verkossa jossakin tiedostomuodossa (esim. PDF).
- Kahden tähden julkaisussa data on saatavilla jossakin koneluettavassa, rakenteisessa formaatissa (esim. XLS).
- Kolmen tähden julkaisussa data on tarjottu jossakin avoimessa, rakenteisessa formaatissa (esim. CSV).
- Neljän tähden julkaisussa on käytetty URI-tunnisteita niin, että julkaisun sisältämään aineistoon voidaan viitata sisäisesti tai muista aineistoista (RDF)

- Viiden tähden julkaisussa dataan on luotu linkkejä muihin aineistoihin (Linkitetty data)

Hyvönen et al. [42] ehdottavat viiden tähden mallin laajentamista seitsemän tähden malliksi:

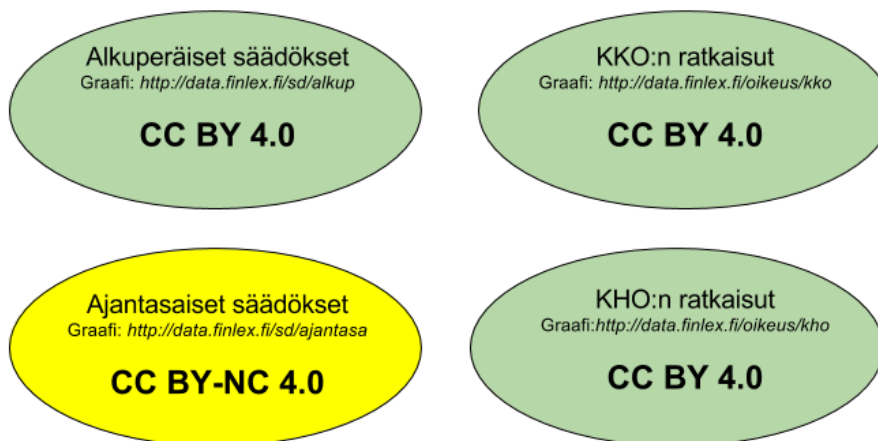
- kuuden tähden julkaisussa datan esittämiseen käytetyt tietomallit on julkaistu ja dokumentoitu
- seitsemän tähden julkaisussa data on vielä validoitu julkaistuja tietomalleja vasten

Kuva 17 havainnollistaa seitsemän tähden mallia. Malli painottaa datan uudelleenkäyttöön ja luottamukseen liittyviä kysymyksiä. Dokumentaation laatiminen, skeemamäärittelyjen julkaiseminen ja datan laadun varmistaminen helpottavat datajulkaisun käyttämistä sovelluksissa.

Semanttinen Finlex täyttää kuuden tähden julkaisun kriteerit. Datan mallintamiseen käytetyt skeemat on julkaistu avoimesti samassa RDF-tietokannassa kuin oikeudelliset aineistotkin. Skeemojen dokumentaatio löytyy datapalvelun dokumentaation osoitteesta <http://data.finlex.fi>. Datajulkaisusta tarjotaan myös VoID-kuvaus²¹ (Vocabulary of Interlinked Datasets) [2]. VoID on RDF-perustainen sanasto, jonka avulla linkitetyn datan julkaisun metatietoja tyyppillisesti esitetään.

Semanttinen Finlex ei kuitenkaan täytä seitsemän tähden julkaisun kriteereitä, koska palvelun tarjoamaa dataa ei olla kyetty täysin validoimaan.

7.2 Käyttöehdot



Kuva 18: RDF-aineistojen käyttöehdot.

Open Knowledge International [62] määrittelee avoimuuden käsitteen suhteessa dataan ja sisältöön. Avoimen datan tulisi olla julkaistu verkossa avoimella lisenssillä

²¹<http://data.finlex.fi/void>

avoimessa tiedostomuodossa. Datajulkaisun lisenssiä valittaessa sisällön tuottajan ja datan julkaisijan intressit eivät kuitenkaan aina kohtaa [40].

Vaikka lakeihin ja asetuksiin ei kohdistukaan tekijänoikeussuojaa (Tekijänoikeuslaki 404/1996 9 §), aiheuttaa Edita Publishing Oy:n tekemä toimitustyö kuitenkin sen, että ajantasaisen lainsäädännön kokoelmaan sekä säädösmuutosten hakemistoon sisältyy aineistojen kaupallista käyttöä rajoittava tekijänoikeus. Näistä tietokannoista kopioituja dokumentteja ei saa vapaasti käyttää uuden tietokannan tai palvelun tuottamiseen [57].

Tämän vuoksi Semanttinen Finlex -palvelun ajantasaisen säädösten kannasta tuotettu data on julkaistu Creative Commons -lisenssillä CC BY-NC 4.0, joka rajoittaa datan käytön ainoastaan epäkaupalliseen tarkoitukseen. Muilta osin Semanttinen Finlex -julkaisun sisältämät aineistot ovat tekijänoikeuksista vapaita ja julkaistu lisenssillä CC BY 4.0.

Aineistot on jaettu omiin RDF-graafeihinsa, joten SPARQL-kysely on helppo kohdistaa ainoastaan avoimella lisenssillä julkaistuihin aineistoihin. Kuvassa 18 on kuvattu eri aineistokokonaisuuksien lisenssit ja niitä vastaavien RDF-graafien URI-tunnisteet.

7.3 Ohjelmointirajapinnat

SPARQL-palvelupiste

SPARQL-kyselyitä voi kohdistaa suoraan Semanttinen Finlex -palvelun RDF-tietokantaan käyttämällä SPARQL-palvelupistettä²². SPARQL-palvelimena toimii Apache Jena Fuseki [78], joka tarjoilee RDF-aineistoja tietokannasta. Datapalvelun verkossa olevaan dokumentaatioon on laadittu joitakin esimerkkikyselyitä.

REST-rajapinta

Aikaisemmassa tutkimuksessa [70] [17] on havaittu, että verkkosovelluskehittäjille RDF ja SPARQL ovat vielä melko vieraita teknologioita. Siksi myös Semanttinen Finlex-palvelun SPARQL-palvelupisteen päälle on rakennettu yksinkertaisiin HTTP-kutsuihin perustuva REST-perustainen (Representational State Transfer) [26] ohjelmointirajapinta²³.

Rajapinnan kautta on mahdollista hakea useita tai yksittäisiä säädöksiä, lainkohtia ja oikeuden ratkaisuja. Rajapinta tarjoaa datan JSON-LD-muodossa (JavaScript Object Notation for Linked Data) [50]. JSON-LD mahdollistaa linkitetyn datan esittämisen sovelluskehittäjille tutussa JSON-tiedostomuodossa. REST-rajapinta palauttaa tietoa yhdessä vastauksessa useasta eri RDF-resurssista upotettuna JSON-puurakenteeseen. Tämä mahdollistaa asiakirjan sisällön sekä RDF-perustaisten metatietojen hierarkkisen esittämisen yhtäaikaaisesti.

Useita säädöksiä voi hakea rajapinnasta HTTP-kutsuilla

<http://data.finlex.fi/api/v1/statutes>

²²<http://data.finlex.fi/sparql>

²³<http://data.finlex.fi/api/v1>

http://data.finlex.fi/api/v1/statutes/{antovuosi}

joista jälkimmäinen tarjoaa säädökset antovuoden perusteella. Kutsut palauttavat säädöksistä listan, joka sisältää säädösten nimikkeet sekä URI-tunnisteet. Yksittäisen säädöksen voi hakea rajapinnasta HTTP-kutsulla:

http://data.finlex.fi/api/v1/statutes/{antovuosi}/{säädösnumero}

Yksittäisen lainkohdan voi hakea hieman vastaavasti kuin ELI-muotoisella URI:lla. Esimerkiksi Rikoslain ensimmäisen luvun toinen pykälä haetaan kutsulla

http://data.finlex.fi/api/v1/statutes/1889/39/luku/1/pykala/2

Oikeuden ratkaisuja voi hakea samaan tapaan. Useita oikeuden ratkaisuja voi hakea kerralla kutsulla

http://data.finlex.fi/api/v1/judgements

ja yksittäisen ratkaisun kutsulla

http://data.finlex.fi/api/v1/judgements/{ECLI-tunniste}

Kaikille kutsuille voi antaa lisäksi kyselyparametrejä, jotka muokkaavat tai rajaavat eri tavoin rajapinnan palauttamaa vastausta. Kyselyparametrillä *tree* yksittäisen säädöksen tai lainkohdan voi hakea siten, että kaikki sen sisältämät lainkohdat on myös esitetty metatietoineen hierarkkisenä JSON-LD-puurakenteena vastauksessa. Ilman parametria vastaus sisältää vain ylimmän lainkohdan metatiedot ja sisällön ilman hierarkkista JSON-LD-rakennetta. Lainkohdan voimassaolevan sisällön voi myös hakea eri ajanhetkillä, eri kielillä ja eri sisältömuodoissa. Kyselyparametrillä *free* rajapinta palauttaa ainoastaan täysin avoimella lisenssillä julkaistua dataa, jonka kaupallista käyttöä ei ole rajoitettu.

Rajapinta on toteutettu Node.js-teknologialla [56], joka on palvelinpään JavaScript-kirjasto ja ajoympäristö. Sovellus on julkaistu avoimena lähdekoodina²⁴ ja se sisältää myös seuraavassa luvussa esiteltävän linkitetyn datan selaimen. Kehittäjäyhteisö voi vapaasti osallistua rajapinnan edelleenkehittämiseen. SPARQL-palvelupisteen päälle voi kuka tahansa toteuttaa myös oman ohjelmointirajapintansa.

Rajapinnan yksityiskohtaisempi tekninen kuvaus ja käyttöohjeet löytyvät Semanttinen Finlex -palvelun dokumentaatiosta²⁵.

Aineiston kertalataus

SPARQL- ja REST-rajapintojen lisäksi Semanttinen Finlex -palvelun käyttäjä pystyy lataamaan RDF-aineistot itselleen kerralla. Tietokannasta voi ladata yksittäisen kokonaisen graafin eri RDF-sarjallistamismuodoissa. RDF-aineistojen lisäksi myös alkuperäisten säädösten sekä oikeuden ratkaisujen alkuperäiset XML-tiedostot ovat ladattavissa palvelusta.

²⁴<https://github.com/apoksane/sf-browser>

²⁵<http://data.finlex.fi/#/doc/rest-api>

8 Sovellukset

Linkitettyä avointa oikeudellista dataa voidaan käyttää monin tavoin eri sovelluksissa. Tässä luvussa esitellään tämän diplomityön osana toteutetut Semanttinen Finlex -datajulkaisua hyödyntävät sovellusprototyypit.

8.1 Linkitetyn datan selaus ja semanttinen suosittelu

Semanttinen Finlex -datajulkaisun URI-tunnisteiden toiminta perustuu sisältöneuvotteluun. Kun HTTP-pyyynnön ylätunnisteessa on määritetty haluttu sisältömuoto RDF-tyyppiseksi, palautetaan vastauksena kysytyn resurssin RDF-metadatan. Sen sijaan HTML-muotoista sisältöä pyytävä verkkoselain ohjautuu ihmisluettavaan linkitetyn datan selaimeseen.

Linkitetyn datan selaimessa resurssin RDF-metadatan on esitetty taulukoina, joissa RDF-perustainen tieto näytetään linkkeinä. Linkkejä seuraamalla pääsee tarkastelemaan edelleen uusia resursseja linkitetyn datan selaimessa. RDF-metadatan on myös upotettu selaimen palauttamaan HTML-sisältöön koneluettavaksi RDF-standardin mukaisesti.

Säädösten URI-tunnisteet toteuttavat ELI-työryhmän suosittemat toiminnallisuudet [24]. Säädöksen hakeminen pelkällä säädösnumerolla varustetulla ELI-muotoisella URI-tunnisteella palauttaa asiakirjan versioiden välillä muuttumattoman metadatan, joka on ilmaistu abstraktilla tunnistetasolla. Sen lisäksi URI palauttaa selaimessa voimassaolevan version säädöksen sisällöstä. Säädöksen voimassaolevaa sisältöä voi hakea tietyllä ajanhetkeltä, jolloin selain palauttaa kyseisellä ajanhetkellä voimassaolleen version sisällön sekä metadatan.

Samat kutsut toimivat myös yksittäisille lainkohdille. Esimerkiksi:

<http://data.finlex.fi/eli/sd/1889/39/luku/1/pykala/2>

palauttaa rikoslain 1 luvun 2 §:n uusimman version sisällön sekä ylimmän FRBR-tason metadatan. Sen sijaan

<http://data.finlex.fi/eli/sd/1889/39/luku/1/pykala/2/ajantasa/20160115>

palauttaa rikoslain 1 luvun 2 §:n uusimman version sisällön sellaisena kuin se oli voimassa 15.1.2016 sekä kyseisen version metadatan. Kaupallisessa Suomen lakipalvelussa²⁶ on myös mahdollisuus tarkastella lainkohtia sellaisina kuin ne ovat olleet voimassa eri ajanhetkinä, mutta Finlex-palvelussa tätä mahdollisuutta ei ole. Kuvassa 19 on esitetty sekä selaimen palauttama näkymä että komentorivillä suoritettun RDF-sisältöä pyytävän HTTP-kutsun vastauksen sisältö edellisen esimerkin URI-tunnisteelle.

Yksittäisten dokumenttien lisäksi listan kaikista säädöksistä voi hakea URI-tunnisteella

<http://data.finlex.fi/eli/sd>

²⁶<http://www.suomenlaki.com>

<<http://data.finlex.fi/eli/sd/1889/39/luku/1/pykala/2/ajantasa/19960901>>

Resurssin ominaisuudet:

Ominaisuus	Resurssi
eli:changed_by	http://data.finlex.fi/eli/sd/1996/626/alkup
eli:has_part	http://data.finlex.fi/eli/sd/1889/39/luku/1/pykala/2/momentti/1/ajantasa/19960901
eli:has_part	http://data.finlex.fi/eli/sd/1889/39/luku/1/pykala/2/momentti/2/ajantasa/19960901
eli:is_part_of	http://data.finlex.fi/eli/sd/1889/39/luku/1/ajantasa/20111001
eli:is_part_of	http://data.finlex.fi/eli/sd/1889/39/luku/1/ajantasa/20150801
eli:is_realized_by	http://data.finlex.fi/eli/sd/1889/39/luku/1/pykala/2/ajantasa/19960901/fin
eli:is_realized_by	http://data.finlex.fi/eli/sd/1889/39/luku/1/pykala/2/ajantasa/19960901/swe
eli:version	sfl:Consolidated

```

<http://data.finlex.fi/eli/sd/1889/39/luku/1/pykala/2/ajantasa/19960901>
<http://data.europa.eu/eli/ontology#version_date>
"1996-09-01"^^<http://www.w3.org/2001/XMLSchema#date> .

<http://data.finlex.fi/eli/sd/1889/39/luku/1/pykala/2/ajantasa/19960901>
<http://data.europa.eu/eli/ontology#is_part_of>
<http://data.finlex.fi/eli/sd/1889/39/luku/1/ajantasa/20150801> .

<http://data.finlex.fi/eli/sd/1889/39/luku/1/pykala/2/ajantasa/19960901>
<http://data.europa.eu/eli/ontology#version>
<http://data.finlex.fi/schema/sfl/Consolidated> .

<http://data.finlex.fi/eli/sd/1889/39/luku/1/pykala/2/ajantasa/19960901>
<http://data.europa.eu/eli/ontology#changed_by>
<http://data.finlex.fi/eli/sd/1996/626/alkup> .

<http://data.finlex.fi/eli/sd/1889/39/luku/1/pykala/2/ajantasa/19960901>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://data.finlex.fi/schema/sfl/SectionVersion> .

<http://data.finlex.fi/eli/sd/1889/39/luku/1/pykala/2/ajantasa/19960901>
<http://data.europa.eu/eli/ontology#is_part_of>
<http://data.finlex.fi/eli/sd/1889/39/luku/1/ajantasa/20111001> .

<http://data.finlex.fi/eli/sd/1889/39/luku/1/pykala/2/ajantasa/19960901>
<http://data.europa.eu/eli/ontology#is_realized_by>
<http://data.finlex.fi/eli/sd/1889/39/luku/1/pykala/2/ajantasa/19960901/fin> .

```

Kuva 19: Semanttinen Finlex -palvelun linkitetyn datan selaus selaimessa ja komentorivillä

Säädöksiä voi hakea myös antovuoden perusteella. Esimerkiksi vuonna 2000 annetut säädökset voi hakea URI-tunnisteella

<http://data.finlex.fi/eli/sd/2000>

Oikeuden ratkaisujen URI-tunnisteet toimivat vastaavasti kuin ELI-perustaiset säädösten tunnisteet. ECLI-tunnuksella varustettu URI tarjoaa selainkäyttäjälle sekä ratkaisun metadatan että sen sisällön.

Sisältö RDF-metadata

KKO:2016:1

Vankeusvanki A:lle oli määrätty huumausaineen käyttörikoksesta kurinpitorangeistukseksi yksinäisyysrangaistus A:n pidettyä vankilassa hallussaan vähäistä määrää huumausaineita. Syyttävä vaati myöhemmin vireille tullessa rikosasiassa A:lle saman menettelyn perusteella rangaistusta huumausaineen käyttörikoksesta.

A:lle rikoksen johdosta määrätyn kurinpitorangeistuksen katsottiin kuuluvan Euroopan ihmisoikeussopimuksen 7. lisäpöytäkirjan 4 artiklassa säädetyn niin sanotun ne bis in idem -säännön soveltamisalaan. Kurinpitorangeistus esti syytteen tutkimisen. Vrt. KKO:2012:46

RL 2 luku 13 § 1 mom

VankL 15 luku 3 § 1 mom 1 kohta

IhmisoikeusSop 7. lisäpöytäkirja 4 art

Asian käsittely alemmissa oikeuksissa
Muutoksenhaku Korkeimmassa oikeudessa
Korkeimman oikeuden ratkaisu

Viittaukset lainsäädäntöön

- Vankeuslaki

Samoja teemoja käsittelevät ratkaisut

Huumausaineen käyttörikos

- KKO:2003:62
- KKO:2009:73

Kurinpitorangeistus

- KKO:2012:46

Samoja teemoja laveasti käsittelevät ratkaisut

Huumausainerikos

- KKO:2001:91
- KKO:2002:111

A oli puhelimessa kehottanut B:tä noutamaan huoltoasemalle kätkeyn kokaiinierän ja toimittamaan sen eteenpäin myöhemmin ilmoitettavaan paikkaan. A oli menettelyllään pitänyt huumausainetta hallussaan ja yrittänyt levittää sitä.

Samoja käsittelevät

data.finlex.fi/oikeus/ECLI:FI:KKO:2001:91

Kuva 20: Semanttinen Finlex -palvelun semanttiset suosittelulinkit oikeuden ratkaisun yhteydessä

Linkitetyn datan selaimessa tarjotaan resurssin omien metatietojen ja sisällön lisäksi asiakirjoissa käsiteltyihin teemoihin ja asiakirjaviitteisiin perustuvia suosittelulinkkejä [85] muihin dokumentteihin. Oikeuden ratkaisujen yhteydessä tarjotaan linkkejä ratkaisun tekstissä viitattuihin säädöksiin. Alkuperäisten säädösten yhteydessä näytetään myös linkit esityöviittauksista poimituille EU-säädöksille, joille haetaan nimikkeet CELLAR-tietokannan SPARQL-palvelupisteestä.

CELLAR-tietokannan SPARQL-palvelupisteestä haetaan myös suosittelulinkkejä Eurovoc-asiasanojen perusteella. Eurovoc²⁷ on Euroopan unionin julkaisutoimiston ylläpitämä monikielinen asiasanasto. Vertaamalla Eurovoc-asiasanan nimikettä Finlex-asiasanaan tai KKO:n tai KHO:n asiasanaan haetaan säädösten ja oikeustapausten kanssa samoja teemoja käsitteleviä EU-säädöksiä.

Oikeuden ratkaisuille haetaan myös samoja teemoja käsitteleviä muita oikeuden ratkaisuja sekä samoja teemoja käsitteleviä säädöksiä. Samoja teemoja käsittelevät säädökset haetaan etsimällä Finlex-asiasanaston nimikkeistä vastineet KKO:n ja KHO:n sanastoille. Koska KKO:n sanastossa on hierarkia, voidaan asiasanan yläkäsitteen kautta etsiä KKO:n päätöksille myös samoja teemoja laveasti käsitteleviä säädöksiä ja oikeuden ratkaisuja.

Kuvassa 20 näkyvät suosittelulinkit ratkaisulle KKO:2016:1. Koska ratkaisun tunniste ei kerro ratkaisun sisällöstä mitään, näytetään ratkaisun tiivistelmä alapuolella

²⁷<http://eurovoc.europa.eu/>

vihjelaatikossa, kun käyttäjä vie kursorin suosittelulinkin päälle.

8.2 Haun tehostus



Kuva 21: Semantinen Finlex -palvelun täydentävä tekstihaku

Perinteisesti oikeudellisen tiedon haku on yritetty tehostaa hakuoperaattorien käytöllä tai tarjoamalla edistyneempiä hakulomakkeita, joiden avulla hakutulosta voi rajata perustuen dokumenttien metatietoihin tai asiasanastoon [68]. Samanaikaisesti nykyaikaiset yksinkertaiseen vapaatekstihakuun perustuvat hakukonesovellukset pystyvät tarjoamaan lyhyisiin, yksinkertaisiin tekstihakuihin vastauksena erittäin relevantteja hakutuloksia.

Sancho Ferrer et al. [67] mukaan oikeudellisen tiedonhaun haaste on saada hakusovellus tarjoamaan yhtä relevantteja hakutuloksia tavallisen käyttäjän kuin juridiikan asiantuntijan tiedonhauille. Asiantuntijat tekevät parempia hakuja, koska heillä on parempi käsitys alalla käytetystä sanastosta sekä haettavien asiakirjojen sisällöstä ja he osaavat valita oikeat hakutyökalut.

Tavallisten käyttäjien kyselyiden tehostamiseksi on Semantinen Finlex -palvelun linkitetyn datan selaimen yhteyteen kehitetty yksinkertainen vapaatekstihaku, jossa hyödynnetään hakutermin automaattista semanttista täydentämistä [41]. Kun käyttäjä näppäilee hakusanan tekstikenttään, tarjotaan alapuolella valikoissa eri vaihtoehtoja hakusanan tai hakusanojen täydentämiseksi.

Talo [75] tutki diplomityössään kaupallisen Edilex-palvelun tiedonhakutoimintojen käytettävyyden parantamista semanttisen webin teknologioilla. Edilex-palvelusta kerätyn käyttäjäpalautteen perusteella hakutoimintoja haluttiin yksinkertaistaa siten, että saman vapaatekstihaun voi kohdistaa sekä dokumentin sisältöön että viitetietoihin. Lisäksi haun haluttiin kohdistuvan myös dokumenttien asiasanoihin. Nykyisin

Edilex-palvelun vapaatekstihaku tarjoaakin hakutekstin täydennysvaihtoehtoina ensimmäiseksi asiasanoja.

Myös Semanttinen Finlex -palvelussa täydennysvaihtoehdot järjestetään siten, että ensimmäisenä näytetään asiasanoista poimitut täydennysvaihtoehdot. Jos käytössä olevien asiasanastojen käsitteille olisi määritetty ontologiassa synonyymejä, voitaisiin myös niitä tarjota käyttäjälle automaattisessa täydentämisessä. Seuraavana tarjotaan säädösten nimikkeitä sekä lainkohtien otsikoita ja lopuksi enintään viiden sanan mittaisia poimintoja säädöstekstien ja oikeuden ratkaisujen tekstisisällöistä. Hakutulokset päivittyvät selaimessa automaattisesti käyttäjän täydentäessä tekstihakuaan, mikä nopeuttaa haun tekemistä. Kuvassa 21 näkyy tekstihaun toiminta.

Myös hakutulokset järjestetään siten, että ensimmäisenä näytetään asiasanoituksen perusteella haetut dokumentit, sen jälkeen otsikon perusteella haetut ja lopuksi tekstisisällön perusteella haetut. Hakusovelluksen taustalla toimii RDF-tietokantaan kytketty Apache Lucene -teksti-indeksi [79]. Lucene pisteyttää nimekkeiden, otsikoiden ja sisällön perusteella poimitut hakutulokset perustuen vektoriavaruusmallin sekä Boolean mallin yhdistelmään. Pisteytyksen perusidea on, että mitä useammin hakutekstissä esiintyvä termi esiintyy yksittäisessä dokumentissa suhteessa termin esiintymien määrään koko dokumenttikokoelmassa, sitä relevantimpi hakutulos dokumentti on.

Vapaatekstihaun hakutulosten relevanssin arviointi vaatii kuitenkin vielä lisätutkimusta. Jatkossa hakua tullaan kehittämään tutkimalla eri mahdollisuuksia yhdistää tekstihakuun moninäkömahaku [53], jota käyttäen tekstihaun tarjoamaa hakutulosta voidaan kohdentaa ontologiaperustaisesti.

9 Johtopäätökset

Tässä luvussa esitetään vastaukset työn alussa esitettyihin tutkimuskysymyksiin sekä vertaillaan Semanttinen Finlex -julkaisua olemassa oleviin oikeudellisiin datapalveluihin. Lopuksi esitetään jatkokehitysideoita Semanttinen Finlex -hankkeen tulevaisuudelle.

9.1 Vastaukset tutkimuskysymyksiin

Ensimmäisessä tutkimuskysymyksessä kysyttiin, miten data kannattaa julkaista, jotta julkaisu palvelee mahdollisimman laajasti eri käyttötapauksia. Käyttötapaukset listattiin luvussa 3. Avointa dataa hyödyntävien sovellusten ja data-analyysin tarpeisiin vastataan siten, että data tarjotaan sekä kokonaisuudessaan ladattavaksi eri muodoissa että ohjelmallisesti käytettäväksi eri rajapintojen kautta. Kansainvälisten standardien käyttö tunnisteissa ja metatietomalleissa helpottaa kansainvälistä yhteentoimivuutta ja organisaatioiden välistä tiedonvaihtoa. Käytetyt tunnisteet mahdollistavat viittamisen yksittäisen lainkohdan eri versioihin yksikäsitteisesti. Myös lainkohtien sisältö on tunnistetta käyttämällä saatavissa verkon yli eri tiedostomuodoissa. Lainvalmistelussa voidaan hyödyntää datan sisältämiä asiakirjojen välisiä semanttisia suhteita.

Toisessa tutkimuskysymyksessä kysyttiin, miten lainsäädännölle ja oikeustapauksille luodaan metatietomäärittelyiltään riittävän kattava, laajennettavissa oleva ja muiden aineistojen kanssa yhteentoimiva tietomalli. Työssä vastattiin tähän kysymykseen ottamalla julkaisussa käyttöön linkitetyn datan teknologiat sekä jo yleisesti käyttöön otetut ELI- ja ECLI-standardien mukaiset tietomallit ja tunnisteet. Nämä standardit määrittelevät käytössä ollutta dataa varten riittävän laajan metatietomallin, joka on laajennettavissa muilla, yksityiskohtaisemmin dataa kuvaavilla malleilla.

Kolmannessa tutkimuskysymyksessä kysyttiin, minkälaisen rajapintojen kautta data tulee tarjota, jotta se on mahdollisimman helposti hyödynnettävissä eri sovelluksissa. Aikaisemmissa datan avaushankkeissa oltiin todettu, että sovelluskehittäjille SPARQL ja RDF ovat teknologioina vielä vieraita, minkä vuoksi dataa tarjotaan Semanttinen Finlex -palvelussa SPARQL-palvelupisteen lisäksi myös yksinkertaisilla HTTP-kutsuilla toimivan REST-rajapinnan kautta.

Neljännessä tutkimuskysymyksessä kysyttiin, miten aineistojen metadataa voidaan rikastaa automaattisesti. XML-RDF-muunnoksen yhteydessä aineistoja rikastetaan esimerkiksi tunnistamalla oikeusneuvosten nimiä sekä viittauksia säädöksiin oikeuden ratkaisuihin. Tulevaisuudessa hankkeessa tullaan kehittämään luonnollisen kielen käsittelyyn perustuvia sovelluksia aineistojen rikastamiseksi edelleen.

Viidennessä tutkimuskysymyksessä kysyttiin, miten aineistosta voidaan luoda yhteyksiä muihin datoihin. Datajulkaisussa on toistaiseksi luotu asiakirjoista yhteyksiä ainoastaan EU-säädöksiin hyödyntäen ELI-standardin mukaisia tunnisteita sekä Eurovoc-asiasanastoa. Koska Eurovoc-asiasanasto tarjoaa käsitteet kaikilla EU:n virallisilla kielillä, on sitä hyödyntäen mahdollista hakea suomalaisiin oikeudellisiin aineistoihin liittyviä aineistoja myös muista EU:n jäsenvaltioista. Aineistojen yhdis-

täminen SPARQL-kyselyitä käyttäen vaatii kuitenkin, että aineistot ovat tarjolla jonkin SPARQL-palvelupisteen kautta.

9.2 Vertailu olemassa oleviin datapalveluihin

Luvussa 3 esiteltiin olemassa olevia oikeudellisia datapalveluita. Yhdistyneessä kuningaskunnassa legislation.gov.uk ja Hollannissa Metalex Document Server tarjoavat lainsäädännöllisiä aineistoja linkitettynä avoimena datana. Molemmat palvelut tarjoavat RDF-perustaista metadataa ja XML-perustaista säädösasiakirjasisältöä avointen rajapintojen kautta. Metalex Document Server tarjoaa RDF-dataa myös SPARQL-palvelupisteen kautta. Kummankaan palvelun rajapinnat eivät kuitenkaan tarjoa dataa siten, että metadata ja sisältö tarjottaisiin sovelluskehittäjien suosimassa JSON-muodossa samanaikaisesti.

Sekä Metalex Document Server- että legislation.gov.uk-palvelun datan mallinnus perustuu MetaLex-standardiin ja eri RDF-sanastoihin. Semanttinen Finlex pyrkii toteuttamaan samat mallinnukselliset ominaisuudet mutta käyttäen aineistojen mallintamiseen MetaLex-standardin sijaan laajennettua ELI-ontologiaa. Samalla Semanttinen Finlex laajentaa aiemmin mainittujen palveluiden palvelumallia tarjoamalla lainsäädännöllisten aineistojen lisäksi myös oikeuden ratkaisuja linkitettynä avoimena datana sekä esittelemällä dataa hyödyntäviä sovellusprototyyppejä.

Luvussa 6.4 todettiin MetaLex Document Server -palvelun datamuunnoksen olevan Semanttinen Finlex -julkaisun muunnosta parempi siinä, että lainkohtien siirtymisen säädösrakenteen sisällä tunnistetaan vertailemalla lainkohtien sisältöjä versioiden välillä. Menetelmän soveltamista Semanttinen Finlex -palvelun datamuunnoksessa on syytä tutkia seuraavaksi.

9.3 Jatkokehitys

Hankkeeseen liittyviä prototyyppisovelluksia kehitettäessä ollaan havaittu, että datajulkaisun tietomallien monimutkaisuus vaikeuttaa SPARQL-kyselyiden muodostamista. Erityisesti lainsäädännön monitasoinen tietomalli monimutkaistaa kyselyitä. Jatkossa hankkeessa tullaan vielä tarkastelemaan, miten tietomallia voitaisiin yksinkertaistaa tai laajentaa siten, että SPARQL-kyselyistä saataisiin yksinkertaisempia. Tätä kirjoitettaessa myös ELI-ontologiasta on juuri julkaistu uusi versio 1.1, joka otetaan mahdollisesti käyttöön Semanttinen Finlex -palvelussa lähitulevaisuudessa.

Semanttinen Finlex -palvelun kehityksen parissa tähän mennessä tehty työ jättää avoimia kysymyksiä vastattavaksi tulevaisuuden tutkimukselle. Seuraavassa on listattu joitakin mahdollisia jatkotutkimuskohteita:

- **Sovellusprototyyppien jatkokehittäminen.** Hankkeen yhteydessä tähän mennessä rakennetuissa sovelluksissa on vielä jatkokehittävää. Semanttisia suosituksia voidaan edelleen laajentaa uusiin linkitetyn datan aineistoihin. Myös moninäkömahaun yhdistämistä vapaatekstihakuun on syytä tutkia ja evaluoida hakutuloksia. Asiakirjojen välisten yhteyksien sekä datasta koostettujen tilastojen visualisointiin on myös mahdollista kehittää uusia ratkaisuja.

- **Asiakirjaviittausten automaattinen tunnistaminen.** Sekä säädösten että oikeuden ratkaisujen teksteissä on paljon viittauksia eri aineistoihin. Näiden tunnistamiseen ei kuitenkaan ole kehitetty välineitä, jotka tunnistaisivat kattavasti eri viittauskäytäntöjä. Esimerkiksi lainkohtaviittausten koneellinen jäsentäminen muutossäädöksen johtolauseesta vähentäisi käsintehdyn toimistustyön tarvetta ajantasaisista säädöskokoelmia päivitettäessä.
- **Aineistojen automaattinen semanttinen annotointi.** Aineistojen olemassa oleva asiasanoitus on tehty hyvin kirjavin käytännön, yleisellä tasolla ja keskenään erilaisia asiasanastoja käyttäen. Asiasanastot on liitetty koko dokumentteihin eikä yksittäisillä lakipykälillä siten ole koneymmärrettävää semanttista metatietoa. Tämän vuoksi aineistojen automaattista asiasanoittamista olisi syytä jatkossa tutkia. Asiasanoituksessa olisi syytä ottaa käyttöön jokin yleinen oikeudellinen asiasanasto, jossa nykyisiä asiasanastoja olisi koottu yhteen ja harmonisoitu.
- **Julkaistujen aineistojen validointi.** Aiemmin työssä todettiin Semanttinen Finlex -palvelussa julkaistun datan sisältävän vielä virheitä. Datankäsittelyn todettiin olevan virheeltistä, koska kaikkia erikois- ja virhetilanteita ei olla kyetty huomaamaan. Virheiden tunnistamiseksi ja korjaamiseksi tulisi kehittää uusia sovelluksia. Toteutuksessa voitaisiin hyödyntää eri RDF-perustaisen tiedon validoimiseksi kehitettyjä sääntö- ja rajoitekieliä.
- **Oikeudellisten asiakirjojen rakenteinen laatiminen.** Suurin osa haasteista, joita Semanttinen Finlex -datajulkaisuhankkeessa on kohdattu, voitaisiin ottaa paremmin huomioon jo aiemmissa asiakirjojen käsittelyvaiheissa. Jos alkuperäisiin asiakirjoihin oltaisiin laadittu riittävät metatiedot, ei asiakirjojen muutostenhallinnassa tarvittaisi mittavaa ihmisvoimin tehtyä toimitustyötä, joka nyt tehdään Edita Publishing Oy:ssä. Asiakirjojen laadinnassa voitaisiin ottaa käyttöön jokin kansainvälinen päätösasiakirjojen laadintaan kehitetty XML-standardi, kuten Akomo Ntoso [65]. Vaihtoehtoisesti muutostenhallinnan yksinkertaistamiseksi voitaisiin uutta sääntelyä laadittaessa muutossäädöksen sijaan koota itse muutettavasta säädöksestä kokonaan uusi versio, kuten on käytäntö esimerkiksi Virossa.

10 Yhteenveto

Tässä työssä esiteltiin linkitetyn avoimen datan palvelu Semanttinen Finlex ja sen kehitystyö. Palvelussa on julkaistu Suomen ajantasainen lainsäädäntö keskeisiltä osin sekä kaikki alkuperäiset säädökset vuodesta 1987 eteenpäin. Lainsäädännöllisten aineistojen lisäksi datajulkaisu sisältää korkeimman oikeuden ennakkopäätökset vuodesta 1980 eteenpäin ja korkeimman hallinto-oikeuden ratkaisut alkaen vuodesta 1987. Aineistoja päivitetään säännöllisesti.

Lainsäädännöllisten aineistojen tietomallissa on otettu käyttöön eurooppalainen standardi ELI. ELI määrittelee säädösasiakirjoille URI-tunnistemallin sekä FRBR-perustaisen ontologian, jotka yhdessä mahdollistavat yksilöllisen viittaamisen sekä voimassa olevaan että aikaisempaan versioon säädöksestä tai lainkohdasta. Malli ottaa huomioon myös säädösten ja lainkohtien eri kieliversiot sekä eri sisältömuodot. Oikeuden ratkaisut on mallinnettu käyttäen eurooppalaista metatieto- ja tunnistestandardia nimeltään ECLI. Semanttinen Finlex -palvelussa oikeuden ratkaisujen ECLI-perustainen tietomalli on yhtenäistetty ELI-perustaisen lainsäädännön tietomallin kanssa.

Lopputuloksena on kuuden tähden linkitetyn avoimen datan julkaisu. Aineistot ovat ladattavissa palvelusta kertalatauksena sekä käytettävissä ohjelmallisesti SPARQL-palvelupisteen ja REST-rajapinnan kautta. Datajulkaisun toimivuutta on arvioitu eri sovellusprototyypeillä.

Palvelua kehitettäessä havaittiin, että alkuperäinen data on metatiedoiltaan vähäistä ja sitä on syytä rikastaa. Lisäksi olemassa olevat asiasanastot ja asiasanoitukset on laadittu hyvin vaihtelevin käytännöin. Eri asiasanastojen harmonisointi sekä aineistojen automaattinen annotointi voisivat tulevaisuudessa mahdollistaa kehittyneempien tiedonhakuovellusten rakentamisen. Metatietojen niukkuudesta ja datan sisältämistä virheistä johtuen myös tiedonhallintaan liittyy vielä ratkaisemattomia haasteita.

Viitteet

- [1] Adida, B., Birbeck, M., McCarron, S., Herman, I.: RDFa Core 1.1 - Third Edition, Syntax and processing rules for embedding RDF through attributes, W3C Recommendation 17 March 2015, <https://www.w3.org/TR/rdfa-syntax/>, viitattu 26.9.2016
- [2] Alexander, K., Cyganiak, R., Hausenblas, M., Zhao, J.: Describing Linked Datasets with the VOID Vocabulary, W3C Interest Group Note 03 March 2011, <https://www.w3.org/TR/void/>, viitattu 26.9.2016
- [3] Baader, F., Horrocks, I., Sattler, U.: Description Logics. Teoksessa: Staab, S., Studer, R. (toim.) Handbook on Ontologies, s. 3–28. International Handbooks in Information Systems, Springer–Verlag, Berlin, Germany (2004)
- [4] Berners-Lee, T.: Design Issues: Linked Data, <https://www.w3.org/DesignIssues/LinkedData.html>, viitattu 26.9.2016
- [5] Biasiotti, M.A., Tiscornia, D.: Legal ontologies: The linguistic perspective. Teoksessa: Sartor, G., Casanovas, P., Biasiotti, M., Fernández-Barrera, M. (toim.) Approaches to Legal Ontologies: Theories, Domains, Methodologies, s. 143–166. Springer Netherlands, Dordrecht (2011), http://dx.doi.org/10.1007/978-94-007-0120-5_9
- [6] Bizer, C., Heath, T., Berners-Lee, T.: Linked Data—The Story So Far. International Journal on Semantic Web and Information Systems (IJSWIS) 5(3), 1–22 (2009)
- [7] Boella, G., Di Caro, L., Graziadei, M., Cupi, L., Salaroglio, C.E., Humphreys, L., Konstantinov, H., Marko, K., Robaldo, L., Ruffini, C., Simov, K., Violato, A., Stroetmann, V.: Linking Legal Open Data: Breaking the Accessibility and Language Barrier in European Legislation and Case Law. Teoksessa: Proceedings of the 15th International Conference on Artificial Intelligence and Law. s. 171–175. ACM, New York, NY, USA (2015), <http://doi.acm.org/10.1145/2746090.2746106>
- [8] Boer, A., Hoekstra, R., Winkels, R., van Engers, T., Willaert, F.: METALex: Legislation in XML. Teoksessa: Bench-Capon, T., Daskalopulu, A., Winkels, R. (toim.) Legal Knowledge and Information Systems. Jurix 2002: The Fifteenth Annual Conference. s. 1–10. Frontiers in Artificial Intelligence and Applications, IOS Press, Amsterdam (2002)
- [9] Brickley, D., Guha, R.: RDF Schema 1.1, W3C Recommendation 25 February 2014, <https://www.w3.org/TR/rdf-schema/>, viitattu 26.9.2016
- [10] Casanovas, P., Sartor, G., Biasiotti, M.A., Fernández-Barrera, M.: Introduction: Theory and Methodology in Legal Ontology Engineering: Experiences and Future Directions. Teoksessa: Sartor, G., Casanovas, P., Biasiotti, M., Fernández-Barrera,

- M. (toim.) *Approaches to Legal Ontologies: Theories, Domains, Methodologies*, s. 1–14. Springer Netherlands, Dordrecht (2011), http://dx.doi.org/10.1007/978-94-007-0120-5_1
- [11] Casellas, N., Bruce, T.R., Frug, S.S., Bouwman, S., Dias, D., Lin, J., Marathe, S., Rai, K., Singh, A., Sinha, D., Venkataraman, S.: *Linked Legal Data: Improving Access to Regulations*. Teoksessa: *Proceedings of the 13th Annual International Conference on Digital Government Research*. s. 280–281. ACM, New York, NY, USA (2012), <http://doi.acm.org/10.1145/2307729.2307785>
- [12] Clark, J., DeRose, S.: *XML Path Language (XPath) Version 1.0*, W3C Recommendation 16 November 1999, revised 7 September 2015, <https://www.w3.org/TR/xpath/>, viitattu 29.9.2016
- [13] 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*. Teoksessa: *Official Journal of the European Union*, C 127, 29.4.2011. s. 1–7. Publications Office of the European Union (2011)
- [14] Council of the European Union: *Council conclusions inviting the introduction of the European Legislation Identifier (ELI)*. Teoksessa: *Official Journal of the European Union*, C 325, 26.10.2012. s. 3–11. Publications Office of the European Union (2012)
- [15] Cox, S., Little, C.: *Time Ontology in OWL*, W3C Working Draft 12 July 2016, <https://www.w3.org/TR/owl-time/>, viitattu 26.9.2016
- [16] Cyganiak, R., Wood, D., Lanthaler, M.: *RDF 1.1 Concepts and Abstract Syntax*, W3C Recommendation 25 February 2014, <http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>, viitattu 26.9.2016
- [17] Daga, E., Panziera, L., Pedrinaci, C.: *A BASILar Approach for Building Web APIs on top of SPARQL Endpoints*. Teoksessa: Maleshkova, M., Verborgh, R., Stadtmüller, S. (toim.) *Proceedings of the Third Workshop on Services and Applications over Linked APIs and Data*. s. 22–32 (2015), <http://oro.open.ac.uk/44026/>
- [18] 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)
- [19] DCMI Usage Board: *DCMI Metadata Terms*, <http://dublincore.org/documents/dcmi-terms/>, viitattu 26.9.2016
- [20] Dozier, C., Kondadadi, R., Light, M., Vachher, A., Veeramachaneni, S., Wudali, R.: *Named Entity Recognition and Resolution in Legal Text*. Teoksessa: Francesconi, E., Montemagni, S., Peters, W., Tiscornia, D. (toim.) *Semantic Processing of Legal Texts: Where the Language of Law Meets the Law of Language*, s. 27–43. Springer,

- Berlin, Heidelberg (2010), http://dx.doi.org/10.1007/978-3-642-12837-0_2
- [21] Ecole Polytechnique Fédérale de Lausanne: The Scala Programming Language, <http://www.scala-lang.org/>, viitattu 29.9.2016
- [22] Eduskunta: Lainvalmistelun vaiheet, https://www.eduskunta.fi/FI/lakiensaattaminen/lainvalmistelu_vaiheet/Sivut/default.aspx, viitattu 26.9.2016
- [23] ELI Task Force: ELI: A Technical Implementation Guide. Publications Office of the European Union (2015)
- [24] ELI Task Force: ELI Implementation Methodology: Good Practices and Guidelines. Publications Office of the European Union (2015)
- [25] Publications Office of the European Union: Common Data Model, <http://publications.europa.eu/mdr/cdm/>, viitattu 14.9.2016
- [26] Fielding, R.T.: Architectural Styles and the Design of Network-based Software Architectures. Väitöskirja, University of California, Irvine (2000)
- [27] Francesconi, E., Montemagni, S., Peters, W., Tiscornia, D.: Integrating a Bottom-Up and Top-Down Methodology for Building Semantic Resources for the Multilingual Legal Domain. Teoksessa: Francesconi, E., Montemagni, S., Peters, W., Tiscornia, D. (toim.) Semantic Processing of Legal Texts: Where the Language of Law Meets the Law of Language, s. 95–121. Springer Berlin Heidelberg, Berlin, Heidelberg (2010), http://dx.doi.org/10.1007/978-3-642-12837-0_6
- [28] Frosterus, M., Tuominen, J., Hyvönen, E.: Facilitating Re-use of Legal Data in Applications – Finnish Law as a Linked Open Data Service. Teoksessa: Proceedings of the Conference on Legal Knowledge and Information Systems JURIX 2014, Krakow, Poland. s. 115–124. IOS Press (2014)
- [29] Frosterus, M., Tuominen, J., Wahlroos, M., Hyvönen, E.: The Finnish Law as a Linked Data Service. Teoksessa: Cimiano, P., Fernández, M., Lopez, V., Schlobach, S., Völker, J. (toim.) The Semantic Web: ESWC 2013 Satellite Events: ESWC 2013 Satellite Events, Montpellier, France, May 26-30, 2013, Revised Selected Papers, s. 289–290. Springer, Berlin, Heidelberg (2013), http://dx.doi.org/10.1007/978-3-642-41242-4_46
- [30] Gandon, F.: Distributed Artificial Intelligence and Knowledge Management: Ontologies and Multi-Agent Systems for a Corporate Semantic Web. (Intelligence Artificielle Distribuée Et Gestion Des Connaissances : Ontologies Et Systèmes Multi-Agents Pour Un Web Sémantique Organisationnel). Väitöskirja, University of Nice Sophia Antipolis, France (2002)

- [31] Gangemi, A., Sagri, M.T., Tiscornia, D.: A Constructive Framework for Legal Ontologies. Teoksessa: Benjamins, V.R., Casanovas, P., Breuker, J., Gangemi, A. (toim.) *Law and the Semantic Web: Legal Ontologies, Methodologies, Legal Information Retrieval, and Applications*, s. 97–124. Springer Berlin Heidelberg, Berlin, Heidelberg (2005), http://dx.doi.org/10.1007/978-3-540-32253-5_7
- [32] Guarino, N., Oberle, D., Staab, S.: What is an Ontology? Teoksessa: Staab, S., Studer, R. (toim.) *Handbook on Ontologies*. Springer, 2. painos (2009)
- [33] Hachey, B., Radford, W., Nothman, J., Honnibal, M., Curran, J.R.: Evaluating Entity Linking with Wikipedia. *Artificial Intelligence* 194, 130–150 (2013), <http://www.sciencedirect.com/science/article/pii/S0004370212000446>
- [34] van Hage, W., Malaisé, V., Segers, R., Hollink, L., Schreiber, G.: The Simple Event Model Ontology, <http://semanticweb.cs.vu.nl/2009/11/sem/>, viitattu 26.9.2016
- [35] Haslhofer, B., Martins, F., Magalhães, J.a.: Using SKOS Vocabularies for Improving Web Search. Teoksessa: *Proceedings of the 22Nd International Conference on World Wide Web*. s. 1253–1258. ACM, New York, NY, USA (2013), <http://doi.acm.org/10.1145/2487788.2488159>
- [36] Heath, T., Bizer, C.: *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool, Palo Alto, California (2011), <http://linkeddatabook.com/editions/1.0/>
- [37] Hietanen, A.: Sähköinen, autenttinen ja ajantasainen: sähköisen säädösjulkaisemisen kehitysvaiheista. Teoksessa: *Oikeuskieli ja säädöstieto: suomenkielinen lakikirja 250 vuotta*, s. 277–287. Suomalainen lakimiesyhdistys (2010)
- [38] Hoekstra, R.: The MetaLex Document Server Legal Documents as Versioned Linked Data. Teoksessa: *Proceedings of the ISWC 2011, Bonn, Germany*. s. 128–143. Springer-Verlag (2011)
- [39] Hoekstra, R., Breuker, J., Bello, M.D., Boer, A.: The LKIF Core Ontology of Basic Legal Concepts. Teoksessa: Casanovas, P., Biasiotti, M.A., Francesconi, E., Sagri, M.T. (toim.) *Proceedings of the Workshop on Legal Ontologies and Artificial Intelligence Techniques (LOAIT 2007)* (2007)
- [40] Hyvönen, E.: *Publishing and Using Cultural Heritage Linked Data on the Semantic Web*. Morgan & Claypool, Palo Alto, California (2012)
- [41] Hyvönen, E., Mäkelä, E.: Semantic Autocompletion. Teoksessa: *Proceedings of the First Asia Semantic Web Conference (ASWC 2006)*, Beijing. Springer-Verlag (2006)

- [42] Hyvönen, E., Tuominen, J., Alonen, M., Mäkelä, E.: Linked Data Finland: A 7-star Model and Platform for Publishing and Re-using Linked Datasets. Teoksessa: Proceedings of ESWC 2014 Demo and Poster Papers, Springer-Verlag (2014)
- [43] IFLA: Functional Requirements for Bibliographic Records: Final Report. K. G. Saur (1998), <http://www.amazon.com/Functional-Requirements-Bibliographic-Records-Publications/dp/359811382X>
- [44] Knublauch, H., Hendler, J.A., Idehen, K.: SPIN - Overview and Motivation, W3C Member Submission 22 February 2011, <https://www.w3.org/Submission/spin-overview/>, viitattu 26.9.2016
- [45] Knublauch, H., Kontokostas, D.: Shapes Constraint Language (SHACL), W3C Working Draft 14 August 2016, <https://www.w3.org/TR/shacl/>, viitattu 26.9.2016
- [46] Koniaris, M., Anagnostopoulos, I., Vassiliou, Y.: Legislation as a complex network: Modelling and analysis of European Union legal sources. Teoksessa: Legal Knowledge and Information Systems - JURIX 2014: The Twenty-Seventh Annual Conference, Jagiellonian University, Krakow, Poland, 10-12 December 2014. s. 143–152 (2014), <http://dx.doi.org/10.3233/978-1-61499-468-8-143>
- [47] Lebo, T., Sahoo, S., McGuinness, D.: PROV-O: The PROV Ontology, W3C Recommendation 30 April 2013, <https://www.w3.org/TR/prov-o/>, viitattu 26.9.2016
- [48] Lima, J.A.O.: An Adaptation of the FRBR Model to Legal Norms. Teoksessa: Biagioli, C., Francesconi, E., Sartor, G. (toim.) Proceedings of the V Legislative XML Workshop, 2006. s. 53–65. EPAP (2007)
- [49] Linholm, J., Derlén, M.: Festina lente – Europarättens genomslag i svensk rättspraxis 1995–2015. Europarättslig tidskrift 1, 151–177 (2015)
- [50] Manu, S., Longley, D., Kellogg, G., Lanthaler, M., Lindström, N.: JSON-LD 1.0, A JSON-based Serialization for Linked Data, W3C Recommendation 16 January 2014, <https://www.w3.org/TR/json-ld/>, viitattu 29.9.2016
- [51] McGuinness, D.L., van Harmelen, F.: OWL Web Ontology Language Overview, W3C Recommendation 10 February 2004, <https://www.w3.org/TR/owl-features/>, viitattu 26.9.2016
- [52] Miles, A., Bechhofer, S.: SKOS Simple Knowledge Organization System Reference, W3C Recommendation 18 August 2009, <https://www.w3.org/TR/skos-reference/>, viitattu 26.9.2016
- [53] Mäkelä, E.: View-based user interfaces for the Semantic Web. G5 artikkeliväitöskirja, Aalto-yliopisto, Teknillinen korkeakoulu, Espoo (2010), <http://urn.fi/URN:ISBN:978-952-60-3447-8>

- [54] Mäkelä, E.: Combining a REST Lexical Analysis Web Service with SPARQL for Mashup Semantic Annotation from Text. Teoksessa: Proceedings of the ESWC 2014 demonstration track, Springer-Verlag (2014)
- [55] Mäkelä, E., Viljanen, K., Alm, O., Tuominen, J., Valkeapää, O., Kauppinen, T., Kurki, J., Sinkkilä, R., Käsälä, T., Lindroos, R., Suominen, O., Ruotsalo, T., Hyvönen, E.: Enabling the Semantic Web with Ready-to-Use Web Widgets. Teoksessa: Proceedings of the First Industrial Results of Semantic Technologies Workshop, ISWC2007 (2007)
- [56] Node.js Foundation: Node.js, <https://nodejs.org>, viitattu 29.9.2016
- [57] Oikeusministeriö: Finlex-käyttöehdot, <http://www.finlex.fi/fi/kayttoehdot/>, viitattu 26.9.2016
- [58] Oikeusministeriö: Lainkirjoittajan opas, <http://lainkirjoittaja.finlex.fi/>, viitattu 26.9.2016
- [59] Oikeusministeriö: Lainlaatijan EU-opas, <http://eu-opas.finlex.fi/>, viitattu 26.9.2016
- [60] Oikeusministeriö: Lainvalmistelun prosessiopas, <http://lainvalmistelu.finlex.fi/>, viitattu 26.9.2016
- [61] Oikeusministeriö: Oikeus.fi: tuomioistuimet, <http://www.oikeus.fi/fi/index/esitteet/oikeuslaitos/tuomioistuimet.html>, viitattu 26.9.2016
- [62] Open Knowledge International: Open Definition 2.1, <http://opendefinition.org/od/2.1/en/>, viitattu 26.9.2016
- [63] van Opijnen, M.: European Case Law Identifier: Indispensable Asset for Legal Information Retrieval. Teoksessa: Biasiotti, M.A., Faro, S. (toim.) From Information to Knowledge – Online Access to Legal Information: Methodologies, Trends and Perspectives. Frontiers in Artificial Intelligence and Applications, IOS Press (2011)
- [64] van Opijnen, M., Ivantchev, A.: Implementation of ECLI - State of Play. Teoksessa: JURIX 2015: The Twenty-Eighth Annual Conference on Legal Knowledge and Information Systems. s. 165–168. IOS Press (2015), <http://dblp.uni-trier.de/db/conf/jurix/jurix2015.html#OpijnenI15>
- [65] Palmirani, M.: Legislative Change Management with Akoma-Ntoso. Teoksessa: Sartor, G., Palmirani, M., Francesconi, E., Biasiotti, A.M. (toim.) Legislative XML for the Semantic Web: Principles, Models, Standards for Document Management, s. 101–130. Springer Netherlands, Dordrecht (2011), http://dx.doi.org/10.1007/978-94-007-1887-6_7

- [66] Salminen, A.: Metatiedot organisaatioiden sisällönhallinnassa. Teoksessa: Lehtinen, A., Salminen, A., Nurmeksela, R. (toim.) Metatiedot suomalaisen lain-säädäntöprosessin tiedonhallinnassa, RASKE2-projektin II väliraportti. s. 4–13. Eduskunnan kanslia (2005)
- [67] Sancho Ferrer, Á., Fernández Hernández, C., Mateo Rivero, J.M.: From Thesaurus Towards Ontologies in Large Legal Databases. Teoksessa: Sartor, G., Casanovas, P., Biasiotti, M., Fernández-Barrera, M. (toim.) Approaches to Legal Ontologies: Theories, Domains, Methodologies, s. 179–200. Springer Netherlands, Dordrecht (2011), http://dx.doi.org/10.1007/978-94-007-0120-5_11
- [68] Sancho Ferrer, Á., Mateo Rivero, J.M., Mesas García, A.: Improvements in Recall and Precision in Wolters Kluwer Spain Legal Search Engine. Teoksessa: Casanovas, P., Sartor, G., Casellas, N., Rubino, R. (toim.) Computable Models of the Law: Languages, Dialogues, Games, Ontologies, s. 130–145. Springer Berlin Heidelberg, Berlin, Heidelberg (2008), http://dx.doi.org/10.1007/978-3-540-85569-9_9
- [69] Scherp, A., Saathoff, C., Franz, T., Staab, S.: Designing Core Ontologies. Applied Ontology 6(3), 177–221 (2011), <http://dl.acm.org/citation.cfm?id=2351285.2351289>
- [70] Sheridan, J., Tennison, J.: Linking UK Government Data. Teoksessa: Bizer, C., Heath, T., Berners-Lee, T., Hausenblas, M. (toim.) Proceedings of the WWW2010 Workshop on Linked Data on the Web, LDOW 2010, Raleigh, USA, April 27, 2010. CEUR Workshop Proceedings, CEUR-WS.org (2010), <http://dblp.uni-trier.de/db/conf/www/ldow2010.html#SheridanT10>
- [71] Sinkkilä, R., Suominen, O., Hyvönen, E.: Automatic Semantic Subject Indexing of Web Documents in Highly Inflected Languages. Teoksessa: Proceedings of the ESWC 2011, Heraklion, Greece. s. 215–229. Springer-Verlag (2011)
- [72] Studer, R., Benjamins, V.R., Fensel, D.: Knowledge Engineering: Principles and Methods. Data & Knowledge Engineering 25(1-2), 161–197 (1998), [http://dx.doi.org/10.1016/S0169-023X\(97\)00056-6](http://dx.doi.org/10.1016/S0169-023X(97)00056-6)
- [73] Suominen, O., Hyvönen, E.: Improving the quality of SKOS vocabularies with Skosify. Teoksessa: Proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2012). Springer-Verlag (2012)
- [74] Szőke, Á., Főrhécz, A., Kőrösi, G., Strausz, G.: Versioned linking of semantic enrichment of legal documents. Artificial Intelligence and Law 21(4), 485–519 (2013), <http://dx.doi.org/10.1007/s10506-013-9145-z>
- [75] Taló, R.: Lakitietopalvelun käytettävyyden parantaminen semanttisen webin tekniikoilla. Diplomityö, Aalto-yliopisto, Teknillinen korkeakoulu, Espoo (2010)

- [76] Talo, Risto. Edita Publishing Oy:n kehityspäällikkö: Haastattelu puhelimitse 16.9.2016
- [77] The Apache Software Foundation: Apache Jena, <https://jena.apache.org/>, viitattu 29.9.2016
- [78] The Apache Software Foundation: Apache Jena Fuseki, <https://jena.apache.org/documentation/fuseki2/>, viitattu 29.9.2016
- [79] The Apache Software Foundation: Apache Lucene, <https://lucene.apache.org/>, viitattu 29.9.2016
- [80] Tieto Finland Oy: Valtioneuvoston päätösasiakirjojen skeemojen ja ulkoasujen kuvaukset, skeemaraportti. Tekninen raportti, Tieto Corporation, versio 1.0- 17, julkaistu 26.9.2013
- [81] Valente, A.: Types and Roles of Legal Ontologies. Teoksessa: Benjamins, V.R., Casanovas, P., Breuker, J., Gangemi, A. (toim.) Law and the Semantic Web: Legal Ontologies, Methodologies, Legal Information Retrieval, and Applications, s. 65–76. Springer Berlin Heidelberg, Berlin, Heidelberg (2005), http://dx.doi.org/10.1007/978-3-540-32253-5_5
- [82] Valtioneuvoston kanslia: Tietopyyntö koskien XML-editoria, <https://www.hankintailmoitukset.fi/fi/notice/view/2016-015560>, viitattu 26.9.2016
- [83] Venturi, G.: Legal language and legal knowledge management applications. Teoksessa: Francesconi, E., Montemagni, S., Peters, W., Tiscornia, D. (toim.) Semantic Processing of Legal Texts: Where the Language of Law Meets the Law of Language, s. 3–26. Springer Berlin Heidelberg, Berlin, Heidelberg (2010), http://dx.doi.org/10.1007/978-3-642-12837-0_1
- [84] W3C SPARQL Working Group: SPARQL 1.1 Overview, W3C Recommendation 21 March 2013, <https://www.w3.org/TR/sparql11-overview/>, viitattu 26.9.2016
- [85] Winkels, R., Boer, A., Vredereg, B., von Someren, A.: Towards a Legal Recommender System. Teoksessa: Proceedings of the Conference on Legal Knowledge and Information Systems JURIX 2014, Krakow, Poland. s. 169–178. IOS Press (2014)