



Parlamenttisammon datapalvelut: miten lataat tietoa eri muodoissa ja käytät rajapintoja

Jouni Tuominen

Parlamenttisampo – eduskunta semanttisessa webissä, 14.2.2023

Parlamenttisampo-data käytettävissä eri muodoissa

- **CSV**
 - Yksinkertainen rakenteinen muoto: pvm, puhuja, puolue, sisältö, ...
 - Uudet puheet päivittyvät datajulkaisuun joka yö
- **XML (Parla-CLARIN ja ParlaMint)**
 - Rakenteisempi muoto
 - *mm. välihuudot erotettu puheesta*
 - ParlaMint: lingvistinen annotaatio (sanaluokat, nimetyt entiteetit, ...)
- **Linkitetty data (RDF)**
 - Linkitys ulkoisiin lähteisiin, semanttinen yhteentoimivuus, SPARQL-kyselyt

Datapalvelu Linked Data Finland

<http://rdf.fi>



Linked Data Finland

Living Laboratory Data Service for the Semantic Web

This site is the Living Laboratory of the [Linked Data Finland](#) research initiative, conducted by the [Semantic Computing Research Group](#) at [Aalto University](#) in collaboration with University of Helsinki and a large consortium of Finnish public organizations and companies.

Our goal is to make life easier for both publishers as well as consumers of structured data on the Web. We base our work on the [Linked Data](#) paradigm and stack of standards, which combines an expressive, semantic data model ([RDF](#)) with standardized access mechanisms ([SPARQL](#) and [live HTTP URIs](#)).

5-star Linked Data

The baseline of our work is the [5-star Linked Data model](#), proposed [originally](#) by Tim Berners-Lee.

- ★ Make data available on the Web in whatever format.
- ★★ Make data available as structured data (e.g., Excel instead of an image scan of a table).
- ★★★ Use non-proprietary formats (e.g., CSV instead of Excel format).
- ★★★★ Use URIs to denote things, so that people can point at your data.
- ★★★★★ Link your data to other data to provide context.

7-star Linked Data Service

However, in our opinion, providing 5-star Linked Data is just the beginning. To actually make use of the datasets, consumers need more support in getting to know and access them, as well as a better grasp of their quality and provenance. To this end, we extend the model with two additional stars:

- ★★★★★ Provide your data with a schema and documentation so that people can *understand and re-use* your data easily.
- ★★★★★★ Validate your data and denote its provenance so that people can *trust the quality* of your data.

This added support should come with as little extra work as possible to the data publisher. Our hypothesis is that a lot of this can be done automatically, basing on the Linked Data core. A data publisher needs only to provide

[Home](#)

[Project](#)

[Datasets](#)

[Schemas](#)

[Services](#)

[Policies](#)

[Documentation](#)

[Validation](#)

[Applications](#)

[Your Data?](#)

7 tähden datan julkaisumalli

Goals: enhance re-usability and data quality

7-star Linked Data Service

However, in our opinion, providing 5-star Linked Data is just the beginning. To actually make use of the datasets, consumers need more support in getting to know and access them, as well as a better grasp of their quality and provenance. To this end, we extend the model with two additional stars:

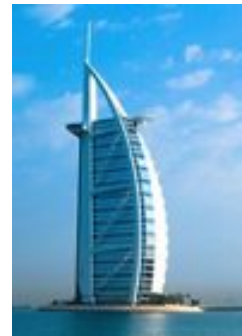


Provide your data with a schema and documentation so that people can *understand and re-use* your data easily.



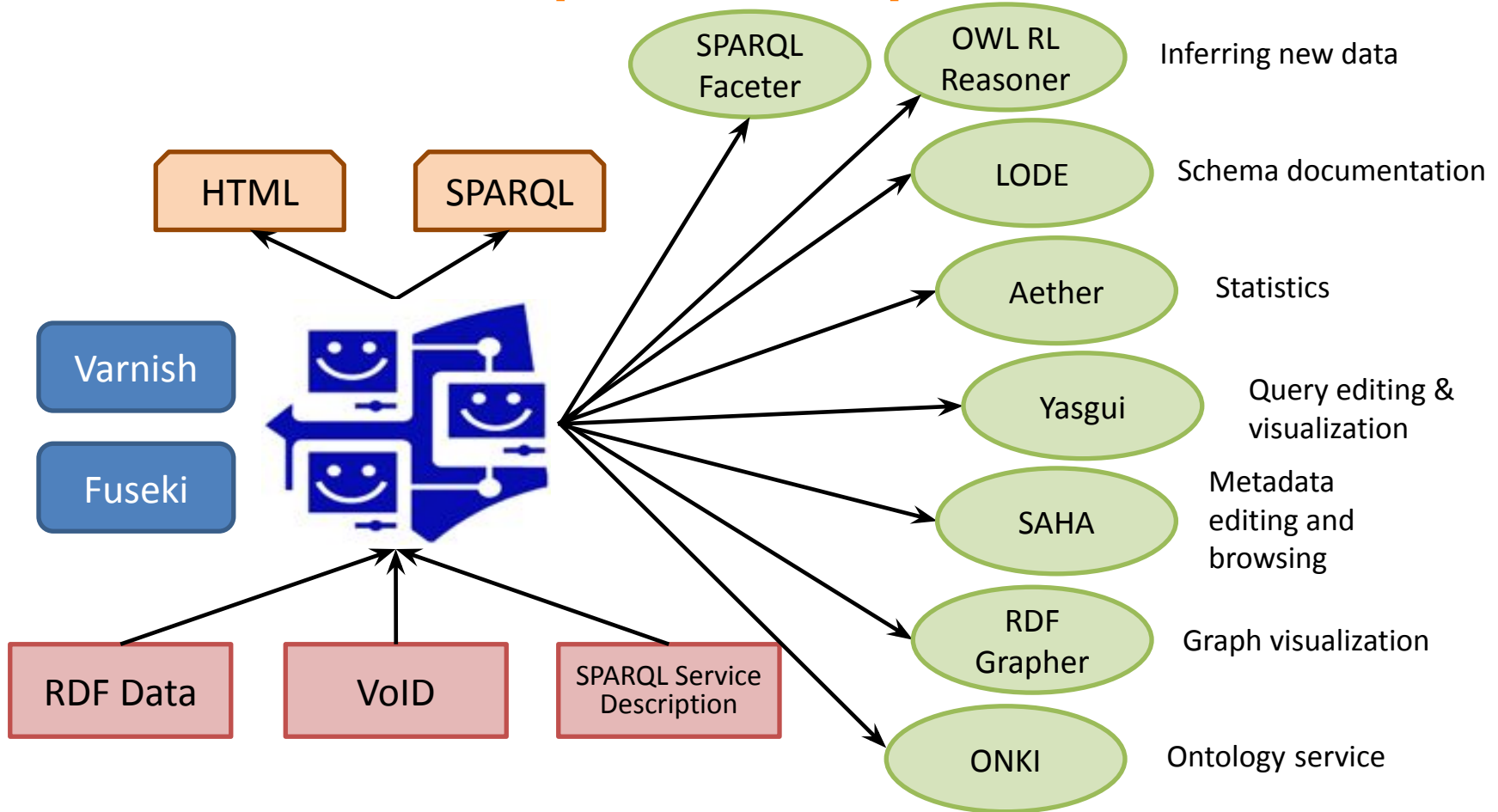
Validate your data and denote its provenance so that people can *trust the quality* of your data.

This added support should come with as little extra work as possible to the data publisher. Our hypothesis is that a lot of this can be done automatically, basing on the Linked Data core. A data publisher needs only to provide their data in the RDF format, and the LDF.fi portal will do the rest automatically. See the [overview paper](#) (in ESWC 2014 Proceedings, Springer-Verlag) for some more details about the underlying ideas.



Burj Al Arab

LDF.fi – Linked Data publication platform





ParliamentSampo

Linked Data Finland



This dataset includes data regarding Finnish Parliamentary debates and actors. The RDF data has been converted using data from the Parliament of Finland's data services and Wikidata; we call the new dataset ParliamentSampo.

The Knowledge Graph contains harmonized data of:

1. speeches from the plenary sessions of the Finnish Parliament 1907–2022, and
2. members and organizations of the Parliament.

The data model is designed for representing speeches, interruptions, items (on agenda), documents and other aspects related to plenary session speeches and minutes as well as member of the parliament and biographical information about them focusing on their political career.

To test and demonstrate its usefulness, this Knowledge Graph is in use in the semantic portal [ParliamentSampo](#), explained in more detail in the [project page](#).

The Knowledge Graph is published on [Zenodo](#) with a version history.

The Knowledge Graph can be downloaded also as CSV files ([Speeches](#), [Actors](#)) and [Parla-CLARIN XML files](#).

Example SPARQL queries for the data:

- [Speeches and interruptions by year](#)
- [Lifetime events of a Member of the Parliament](#)

- [Home](#)
- [Project](#)
- [Datasets](#)
- [Search Data](#)
- [Schemas](#)
- [Services](#)
- [Policies](#)
- [Documentation](#)
- [Validation](#)
- [Linked Data Science](#)
- [Applications](#)
- [Your Data?](#)
- [Linked Data School](#)

Data Download

The data can be downloaded at <https://zenodo.org/record/7636420/files/parliamentsampo-data-v1.0.0.zip>.

License



Licensors: <http://www.eduskunta.fi>, <http://seco.cs.aalto.fi>

See possible graph-specific licenses below.

Detailed Dataset Contents

Finnish Parliamentary Debates (URI: http://idf.fi/semparl/speeches_and_interruptions)

([Browse data](#))

Speeches and interruptions gathered from the minutes of the Parliament of Finland's plenary sessions.

Example resource URI: http://idf.fi/semparl/speeches/s2015_1_049_079

Plenary session items and referred documents (URI: http://idf.fi/semparl/items_and_documents)

([Browse data](#))

Plenary session items and referred parliamentary document details of the Parliament of Finland.

Example resource URI: <http://idf.fi/semparl/items/i20181581121>

CSV: puheet ja henkilöt

← → ↻ 🏠 a3s.fi/parliamentsampo/speeches/csv/index.html 📄 ☆ ⚙️ 🗄️ 👤

ParliamentSampo Speeches CSV data

The dataset contains speeches from the plenary sessions of the Finnish Parliament 1907-. The current parliamentary session's ('valtiopäivät' in Finnish) speeches are updated nightly.

The CSV files are named as speeches_PARLIAMENTARY_SESSION_NR.csv (Note: a parliamentary session does not fully correspond to a calendar year, e.g., parliamentary session 2022 continues in year 2023.).

The data has been updated on Tue Feb 14 02:25:45 EET 2023.

See [dataset documentation](#).

- [speeches-all.csv.zip](#) (2288750 lines, 589M)
- [speeches_1907.csv](#) (9710 lines, 9.0M)
- [speeches_1908.csv](#) (1930 lines, 1.7M)
- [speeches_1908_II.csv](#) (5826 lines, 5.9M)
- [speeches_1909.csv](#) (107 lines, 56K)
- [speeches_1909_II.csv](#) (2663 lines, 6.8M)
- [speeches_1910.csv](#) (1763 lines, 5.9M)
- [speeches_1910_II.csv](#) (122 lines, 228K)
- [speeches_1911.csv](#) (2046 lines, 4.7M)
- [speeches_1912.csv](#) (2261 lines, 4.8M)
- [speeches_1913.csv](#) (2351 lines, 4.9M)
- [speeches_1914.csv](#) (2237 lines, 5.4M)
- [speeches_1917.csv](#) (2695 lines, 5.7M)
- [speeches_1917_II.csv](#) (3888 lines, 9.6M)
- [speeches_1917_XX.csv](#) (71 lines, 104K)
- [speeches_1918.csv](#) (2081 lines, 4.1M)
- [speeches_1918_II.csv](#) (234 lines, 572K)
- [speeches_1919.csv](#) (6040 lines, 15M)
- [speeches_1920.csv](#) (6041 lines, 15M)
- [speeches_1921.csv](#) (6141 lines, 15M)
- [speeches_1922.csv](#) (2793 lines, 8.1M)
- [speeches_1923.csv](#) (3123 lines, 7.9M)
- [speeches_1924.csv](#) (3837 lines, 9.7M)
- [speeches_1925.csv](#) (4652 lines, 13M)
- [speeches_1926.csv](#) (6229 lines, 17M)
- [speeches_1927.csv](#) (3228 lines, 8.2M)
- [speeches_1928.csv](#) (4128 lines, 16M)
- [speeches_1929.csv](#) (1632 lines, 3.6M)
- [speeches_1929_II.csv](#) (3667 lines, 11M)
- [speeches_1930.csv](#) (2024 lines, 5.0M)
- [speeches_1930_II.csv](#) (2128 lines, 6.1M)
- [speeches_1931.csv](#) (3910 lines, 9.1M)
- [speeches_1932.csv](#) (5977 lines, 15M)

← → ↻ 🏠 a3s.fi/parliamentsampo/actors/csv/index.html

ParliamentSampo Actors CSV data

The dataset contains data of members and organizations of the Parliament of Finland.

See [dataset documentation](#).

- [people.csv](#) (10849 lines, 2.0M)

CSV: henkilöt, ryhmiin kuulumiset

id	label	family_name	given_name	other_family_names	other_given_names	gender	birth	death	group	role	start	end
http://ldf.fi/sempari/people/p910215	Aakula, Eemeli (1879-1955)	Aakula	Eemeli	Aakula	Mikkel Emil; Mikko Eemeli	Mies	1879	1955	Maalaisliiton eduskuntaryhmä (1930-1933)	jäsen	1930-10-21	1936-08-3
http://ldf.fi/sempari/people/p910215	Aakula, Eemeli (1879-1955)	Aakula	Eemeli	Aakula	Mikkel Emil; Mikko Eemeli	Mies	1879	1955	Maalaisliiton eduskuntaryhmä (1933-1936)	jäsen	1930-10-21	1936-08-3
http://ldf.fi/sempari/people/p910216	Aakula, Pekka (1866-1928)	Aakula	Pekka	Aakula	Petter	Mies	1866	1928	Sosialidemokraattinen eduskuntaryhmä (1909-1910)	jäsen	1909-10-01	1911-01-3
http://ldf.fi/sempari/people/p910216	Aakula, Pekka (1866-1928)	Aakula	Pekka	Aakula	Petter	Mies	1866	1928	Sosialidemokraattinen eduskuntaryhmä (1910-1911)	jäsen	1909-10-01	1911-01-3
http://ldf.fi/sempari/people/p910216	Aakula, Pekka (1866-1928)	Aakula	Pekka	Aakula	Petter	Mies	1866	1928	Sosialidemokraattinen eduskuntaryhmä (1914-1917)	jäsen	1914-02-02	1917-10-3
http://ldf.fi/sempari/people/p910216	Aakula, Pekka (1866-1928)	Aakula	Pekka	Aakula	Petter	Mies	1866	1928	Sosialidemokraattinen eduskuntaryhmä (1917-1917)	jäsen	1914-02-02	1917-10-3
http://ldf.fi/sempari/people/p910217	Aalle-Teljo, Ida (1875-1955)	Aalle-Teljo	Ida	Aalle-Teljo; Ahlstedt	Ida Sofia	Nainen	1875	1955	Sosialidemokraattinen eduskuntaryhmä (1907-1908)	jäsen	1907-05-22	1917-10-3
http://ldf.fi/sempari/people/p910217	Aalle-Teljo, Ida (1875-1955)	Aalle-Teljo	Ida	Aalle-Teljo; Ahlstedt	Ida Sofia	Nainen	1875	1955	Sosialidemokraattinen eduskuntaryhmä (1908-1909)	jäsen	1907-05-22	1917-10-3
http://ldf.fi/sempari/people/p910217	Aalle-Teljo, Ida (1875-1955)	Aalle-Teljo	Ida	Aalle-Teljo; Ahlstedt	Ida Sofia	Nainen	1875	1955	Sosialidemokraattinen eduskuntaryhmä (1909-1910)	jäsen	1907-05-22	1917-10-3
http://ldf.fi/sempari/people/p910217	Aalle-Teljo, Ida (1875-1955)	Aalle-Teljo	Ida	Aalle-Teljo; Ahlstedt	Ida Sofia	Nainen	1875	1955	Sosialidemokraattinen eduskuntaryhmä (1910-1911)	jäsen	1907-05-22	1917-10-3
http://ldf.fi/sempari/people/p910217	Aalle-Teljo, Ida (1875-1955)	Aalle-Teljo	Ida	Aalle-Teljo; Ahlstedt	Ida Sofia	Nainen	1875	1955	Sosialidemokraattinen eduskuntaryhmä (1911-1914)	jäsen	1907-05-22	1917-10-3
http://ldf.fi/sempari/people/p910217	Aalle-Teljo, Ida (1875-1955)	Aalle-Teljo	Ida	Aalle-Teljo; Ahlstedt	Ida Sofia	Nainen	1875	1955	Sosialidemokraattinen eduskuntaryhmä (1914-1917)	jäsen	1907-05-22	1917-10-3
http://ldf.fi/sempari/people/p910217	Aalle-Teljo, Ida (1875-1955)	Aalle-Teljo	Ida	Aalle-Teljo; Ahlstedt	Ida Sofia	Nainen	1875	1955	Sosialidemokraattinen eduskuntaryhmä (1917-1917)	jäsen	1907-05-22	1917-10-3
http://ldf.fi/sempari/people/p910218	Aaltio, Eino (1889-1974)	Aaltio	Eino	Andelin; Aaltio	Eino Abram; E.A.	Mies	1889	1974	Maalaisliiton eduskuntaryhmä (1927-1929)	jäsen	1929-02-01	1929-07-3
http://ldf.fi/sempari/people/p910219	Aaltio, Sampsa (1927-2013)	Aaltio	Sampsa	Aaltio	Sampsa Eino Urmas	Mies	1927	2013	Kansallisen kokoomuksen eduskuntaryhmä (1979-1983)	jäsen	1979-03-24	1987-03-2
http://ldf.fi/sempari/people/p910219	Aaltio, Sampsa (1927-2013)	Aaltio	Sampsa	Aaltio	Sampsa Eino Urmas	Mies	1927	2013	Kansallisen kokoomuksen eduskuntaryhmä (1983-1987)	jäsen	1979-03-24	1987-03-2
http://ldf.fi/sempari/people/p910207	Aalto, Anna-Kaarina (1920-1994)	Aalto	Anna-Kaarina			Nainen	1920	1994	Kansallisen kokoomuksen eduskuntaryhmä (1970-1972)	jäsen	1971-02-02	1972-01-2

XML: Parla-CLARIN

← → ↻ 🏠 a3s.fi/parliamentsampo/speeches/xml/index.html

ParliamentSampo Speeches Parla-CLARIN data

The dataset contains speeches from the plenary sessions of the Finnish Parliament 1907-2022.

See [dataset documentation](#), [Parla-CLARIN format documentation](#).

- [speeches-all.xml.zip](#) (2166289 lines, 560M)
- [Speeches 1907.xml](#) (17714 lines, 9.1M)
- [Speeches 1908.xml](#) (4478 lines, 1.7M)
- [Speeches 1908 II.xml](#) (13484 lines, 5.8M)
- [Speeches 1909.xml](#) (679 lines, 72K)
- [Speeches 1909 II.xml](#) (16627 lines, 6.6M)
- [Speeches 1910.xml](#) (12038 lines, 5.6M)
- [Speeches 1910 II.xml](#) (1192 lines, 248K)
- [Speeches 1911.xml](#) (13312 lines, 4.5M)
- [Speeches 1912.xml](#) (13715 lines, 4.5M)
- [Speeches 1913.xml](#) (14454 lines, 4.7M)
- [Speeches 1914.xml](#) (14115 lines, 5.1M)
- [Speeches 1917.xml](#) (17244 lines, 5.6M)
- [Speeches 1917 II.xml](#) (25824 lines, 9.1M)
- [Speeches 1917 XX.xml](#) (681 lines, 120K)
- [Speeches 1918.xml](#) (12348 lines, 3.9M)
- [Speeches 1918 II.xml](#) (1971 lines, 288K)
- [Speeches 1919.xml](#) (33289 lines, 13M)
- [Speeches 1920.xml](#) (33403 lines, 13M)
- [Speeches 1921.xml](#) (34726 lines, 14M)
- [Speeches 1922.xml](#) (18047 lines, 7.8M)
- [Speeches 1923.xml](#) (19955 lines, 7.8M)
- [Speeches 1924.xml](#) (23798 lines, 9.3M)
- [Speeches 1925.xml](#) (27204 lines, 11M)
- [Speeches 1926.xml](#) (37641 lines, 16M)
- [Speeches 1927.xml](#) (20677 lines, 7.9M)
- [Speeches 1928.xml](#) (29222 lines, 9.5M)
- [Speeches 1929.xml](#) (11536 lines, 3.5M)
- [Speeches 1929 II.xml](#) (27034 lines, 8.4M)
- [Speeches 1930.xml](#) (15801 lines, 4.6M)
- [Speeches 1930 II.xml](#) (16203 lines, 5.0M)
- [Speeches 1931.xml](#) (28312 lines, 8.6M)
- [Speeches 1932.xml](#) (45227 lines, 13M)

```
31794 <div>
31795 <note end="2015-10-14T14:24:12" link="https://www.eduskunta.fi/FI/Vaski/PoytakirjaAsiakohhta/Sivut/PTK_50+2015+2.aspx" multilingual="false" speechType="Puheenvuoro" start="2015-10-14T14:21:41"
31796 type="speaker" xml:id="2015_50_4" xml:lang="fi"/>
<u next="2015_50_4.2" who="#Anne_Kalmar" xml:id="2015_50_4.1">Arvoisa rouva puhemies! On erittäin tärkeää, että tänään keskustelemme täällä rasismista.Meissä kaikissa asuu se lämmin sydän, joka
haluaa tehdä hyvää. Mutta me kaikki saamme paljon palautetta ihmisiltä, jotka haluavat, että esimerkiksi maahanmuuttoon puuttaisiin, että meillä tehtäisiin asioita toisin kuin mitä nyt
tapahtuu. Siksi on tärkeää, että me kansanedustajat olemme vahvoja, kun perustelemme, miksi Suomessa toimitaan juuri niin kuin meillä nyt toimitaan ja miksi meillä on nollatoleranssi väkivallassa.
On tärkeää, että me jokainen teemme omilla kotikunnissamme tätä sen eteen, että lähiyhteisössä rasismi kiitetään pois. Se ei ole helppoa työtä, sillä tiedän, että kun meillä on erilaisia taustoja,
monesti sieltä tulee hyvin erilaisia palautetta vastaan. Tämä puheenvuoro, minkä edustaja Razmyar puhui, antaa meille kaikille voimia tehdä sitä työtä, kun me toivottavasti joka ainossa puolueessa
puolustamme sitä nollatoleranssia.On tärkeää, että olemme myös esimerkkeinä pienillä teoilla, kuten meidän pääministerimme oli antaessaan pakolaisille tai turvapaikanhakijoille oman kodin
käyttönsä. Ne esimerkit, että osallistumme vaatekeräyksiin tai vapaaehtoistyöhön, että lähdetään yhdessä </u>
</div>
<div>
<desc>Puhemies koputtaa</desc>
</vocal>
31797 </div>
31798 <div>
31799 <u next="2015_50_4.3" prev="2015_50_4.1" who="#Anne_Kalmar" xml:id="2015_50_4.2"> vaikkapa keräämään metsästä marjoja – kaikki tämäntyyppiset toimet luovat sitä kotoutumista, ja voimme itse olla
31800 edesauttamina myös </u>
</vocal>
<desc>Puhemies koputtaa</desc>
</vocal>
31801 </div>
31802 <u prev="2015_50_4.2" who="#Anne_Kalmar" xml:id="2015_50_4.3">tulijoiden tänne sopeutumista.</u>
31803 </div>
31804 <div>
31805 <note end="2015-10-14T14:26:17" link="https://www.eduskunta.fi/FI/Vaski/PoytakirjaAsiakohhta/Sivut/PTK_50+2015+2.aspx" multilingual="false" speechType="Puheenvuoro" start="2015-10-14T14:24:12"
31806 type="speaker" xml:id="2015_50_5" xml:lang="fi"/>
31807 <u who="#Arto_Satonen" xml:id="2015_50_5">Arvoisa puhemies! On monia asioita, joita voi ymmärtää, mutta rasismi ei ole sellainen. Rasismia kohtaan on oltava nollatoleranssi, ja sen takia on erittäin
31808 tärkeää, että myös poliittiset puolueet kautta linjan yhteisesti tuomitsevat rasismin ja myöskin allekirjoittavat tämän rasisminvastaisen julistuksen. Kokouksen lähtökohhta on aina ollut se, että
rasisimia emme voi hyväksyä missään muodossa.Täällä aiemmin edustaja Nasima Razmyar puheenvuorossa kävi varsin hyvin esille se, kuinka tärkeää on, että maahan tulijat pääsevät osaksi yhteiskuntaa.
Tämä integroituinen suomalaisen järjestelmään ja Suomeen ja onan paikan löytäminen tähän yhteiskunnassa on kaikkein tärkein asia. Siinä suhteessa se kaikki työ, millä tätä voidaan edistää niin
oppilaitoksissa, vapaaehtoistyössä kuin työelämässä on hyvin keskeistä, ja niitä kaikkia kysymyksiä meidän on tarkasteltava, kun sitten tämän viimeisen turvapaikanhakijoiden aallon kohdalta ollaan
siinä vaiheessa, että osa heistä saa oikeuden jäädä Suomeen ja varmaan myöskin yhdistää perheitään tänne.Luulen, että kaikkein tärkeintä on nyt se, että me emme aina minkäänlaista sijaa
käräilikkelle vaan olemme yhteisesti tukenassa matillista suhtautumista. Kysymys on nimenomaan se, miten voimme rakentaa asioita emmekä riko niitä.On erittäin hyvä, että eduskunta käy tämän
keskustelun hyvässä yhteisessä hengessä.</u>
</div>
31809 <div>
31810 <note end="2015-10-14T14:28:18" link="https://www.eduskunta.fi/FI/Vaski/PoytakirjaAsiakohhta/Sivut/PTK_50+2015+2.aspx" multilingual="false" speechType="Puheenvuoro" start="2015-10-14T14:26:17"
31811 type="speaker" xml:id="2015_50_6" xml:lang="fi"/>
31812 <u who="#Sampo_Terho" xml:id="2015_50_6">Arvoisa puhemies! Asia, josta puhume, on todellakin tärkeä, ja tässä kysymyksessä on syytä olla aloitteellinen. Maan hallituksella on suuri vastuu, kaikilla
eduskunnassa olevilla henkilöillä ja puolueilla on suuri vastuu, ja pidän myös hyvin tärkeänä, kuten edellinen puhuja totesi, että usimme Euroopan poliittisten puolueiden peruskirjaan rasismista
vapaan yhteiskunnan puolesta tekemään allekirjoitukset, jotka viimeksi 2011 on ryhmiin puolesta allekirjoitettu.Tässä kysymyksessä asiakirjassa todetaan rotusyrjinnästä seuraavaa: <u>Rotusyrjintä
on kaikkea rottuun, huonosti, syntyperään tai kansalliseen tai etniseen alkuperään perustuvaa erottelua, pöissuukemistä tai etuoikeutta.</u>Tämä on hyvin tärkeä ajatus, ja tämä on ajatus, jota
meidän täytyy aina korostaa. Ketään ei syntyperänsä eikä uskontonsa tai poliittisen mielipiteensä takia saa koskaan, missään yhteydessä tai tilanteessa, asettaa toisen ylä- tai alapuolelle, eteen
tai jälkeen. Mahdollisuuksien tasa-arvon on oltava ehdoton, samoin kuin keskinäisen kunnioituksen. Meidän on ehdottomasti luotava yhteiskunta, jossa jokainen voi omalla työllään, omalla
ahkeruudellaan saavuttaa sen, mitä ansaitsee. Se on jokaisen oikeus. Se on myös meidän kaikkien etu, mutta se ei ole etuoikeus, vaan Suomessa sen täytyy olla itsestäänselvyys, kuten ministeri sanoi,
itsestäänselvyys. Se on aivan oikea sana. Ja tästä itsestäänselvydestä voin taata, että oma ryhmäni ja hallitus, jota edustame, tunnustaa sen tinkimättömänä, jakamattomana ja kaikille kuuluvana. –
Kiitos.</u>
</div>
31813 </div>
```

XML: ParlaMint

- **Parlamenttisampo-datan julkaisu osana CLARIN-infrastruktuurin ParlaMint-aineistoa (2015–)**
- **Mahdollistaa eri maiden parlamenttiaineistojen vertailevan tutkimuksen**

The screenshot shows the CLARIN.SI repository page for the item 'Multilingual comparable corpora of parliamentary debates ParlaMint 2.1'. The page includes a search bar, navigation links (Repository, About, Contact), and a sidebar with options like Browse, My Account, Login, Statistics, and General Information. The main content area displays the item's metadata and description.

CLARIN.SI repository / View Item

Multilingual comparable corpora of parliamentary debates ParlaMint 2.1

Please use the following text to cite this item or export to a predefined format: [BIBTEX](#) [CMDI](#)

Erjavec, Tomaž; et al., 2021, *Multilingual comparable corpora of parliamentary debates ParlaMint 2.1*, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1432>.

Share: [f](#) [t](#)

CLARIN.SI Data & Tools

Authors Erjavec, Tomaž; et al.
▶ show everyone

Item Identifier <http://hdl.handle.net/11356/1432>

Project URL <https://www.clarin.eu/content/parlamint>

Demo URL <https://github.com/clarin-eric/ParlaMint/>

Referenced by <https://doi.org/10.1007/s10579-021-09574-0>

Date issued 2021-06-18

Type corpus, text

Size 3774204 utterances, 494949904 words

Language(s) Bulgarian, Croatian, Czech, Danish, Dutch, English, French, Hungarian, Icelandic, Italian, Latvian, Lithuanian, Polish, Slovenian, Spanish, Turkish

Description ParlaMint 2.1 is a multilingual set of 17 comparable corpora containing parliamentary debates mostly starting in 2015 and extending to mid-2020, with each corpus being about 20 million words in size. The sessions in the corpora are marked as belonging to the COVID-19 period (after November 1st 2019), or being "reference" (before that date).

The corpora have extensive metadata, including aspects of the parliament; the speakers (name, gender, MP status, party affiliation, party coalition/opposition); are structured into time-stamped terms, sessions and meetings; with speeches being marked by the speaker and their role (e.g. chair, regular speaker). The speeches also contain marked-up transcriber comments, such as gaps in the transcription, interruptions, applause, etc. Note that some corpora have further information, e.g. the year of birth of the speakers, links to their Wikipedia articles, their membership in various committees, etc.

The corpora are encoded according to the Parla-CLARIN TEI recommendation (<https://clarin-eric.github.io/parla-clarin/>), but have been validated against the compatible, but much stricter ParlaMint schemas.

This entry contains the ParlaMint TEI-encoded corpora with the derived plain text version of the corpus along with TSV metadata on the speeches. Also included is the 2.0 release of the data and scripts available at the GitHub repository of the ParlaMint project.

Note that there also exists the linguistically marked-up version of the corpus, which is available at <http://hdl.handle.net/11356/1431>.

Publisher CLARIN ERIC

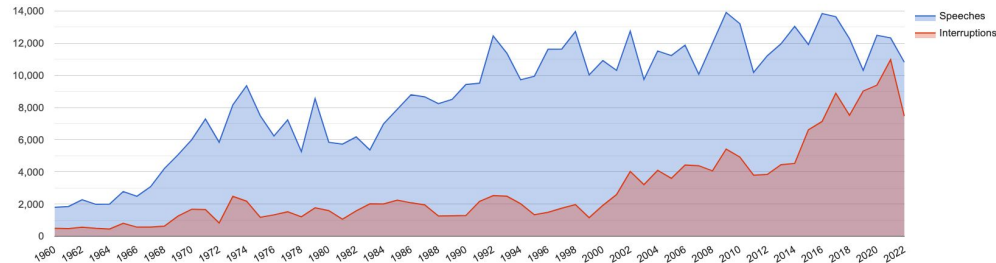
XML: ParlaMint, lingvistinen annotaatio

```
<ns0:s xml:id="ParlaMint-FI_2015-05-26-ps-8_seg4.13">
  <ns0:w xml:id="ParlaMint-FI_2015-05-26-ps-8_seg4.13.223"
    lemma="mitä"
    msd="UPosTag=CCONJ|">Mitä</ns0:w>
  <ns0:w xml:id="ParlaMint-FI_2015-05-26-ps-8_seg4.13.224"
    lemma="hyvä"
    msd="UPosTag=ADJ|Case=Nom|Degree=Cmp|Number=Plur">paremmat</ns0:w>
  <ns0:w xml:id="ParlaMint-FI_2015-05-26-ps-8_seg4.13.225"
    lemma="perustelu"
    msd="UPosTag=NOUN|Case=Nom|Derivation=U|Number=Plur">perustelut</ns0:w>
  <ns0:w xml:id="ParlaMint-FI_2015-05-26-ps-8_seg4.13.226"
    lemma="olla"
    msd="UPosTag=AUX|Mood=Ind|Number=Plur|Person=3|Tense=Pres|VerbForm=Fin|Voice=Act">ovat</ns0:w>
  <ns0:pc xml:id="ParlaMint-FI_2015-05-26-ps-8_seg4.13.227"
    lemma=","
    msd="UPosTag=PUNCT|">,</ns0:pc>
  <ns0:w xml:id="ParlaMint-FI_2015-05-26-ps-8_seg4.13.228"
    lemma="sitä"
    msd="UPosTag=ADV|">sitä</ns0:w>
  <ns0:w xml:id="ParlaMint-FI_2015-05-26-ps-8_seg4.13.229"
    lemma="vähemmän"
    msd="UPosTag=ADV|">vähemmän</ns0:w>
  <ns0:w xml:id="ParlaMint-FI_2015-05-26-ps-8_seg4.13.230"
    lemma="jäädä"
    msd="UPosTag=VERB|Mood=Ind|Number=Sing|Person=3|Tense=Pres|VerbForm=Fin|Voice=Act">jää</ns0:w>
  <ns0:w xml:id="ParlaMint-FI_2015-05-26-ps-8_seg4.13.231"
    lemma="tulkintaepäselvyys"
    msd="UPosTag=NOUN|Case=Par|Number=Plur">tulkintaepäselvyyksiä</ns0:w>
```


SPARQL-kyselyt

- Rikas ja monipuolinen tapa tehdä kyselyitä dataan
- Parlamenttisampo-portaali perustuu SPARQL:ään (avoin lähdekoodi GitHub-repositoriossa)

```
https://rdf.fi/semparl/sparql
1 PREFIX semparl: <http://ldf.fi/schema/semparl/>
2 PREFIX types: <http://ldf.fi/semparl/speechtypes/>
3
4 # Speeches and interruptions since 1960, grouped by year
5
6 SELECT (STR(?_year) AS ?Year) (COUNT(DISTINCT ?speech) AS ?Speeches) (COUNT(?interp) AS ?Interruptions)
7 WHERE {
8   VALUES ?type { # speech types, NB not including the speeches by the spokesman
9     types:Fuhsenvuoro types:Vastauspuheenvuoro types:Ryhmalpuheenvuoro
10    types:Esittelypuheenvuoro types:NopeastiLiisapuheenvuoro }
11
12   ?speech semparl:speechType ?type ;
13     semparl:endDate ?date .
14
```



Lataa koko aineisto linkitettyinä datana

zenodo Search Upload Communities Log in Sign up

February 14, 2023 Dataset Open Access

ParliamentSampo Knowledge Graph

Sinikallio, Laura; Leskinen, Petri; Drobac, Senka; Tamper, Minna; Leäl, Rafael; Tuominen, Jouni; Rantala, Heikki; La Mela, Matti; Koho, Mikko; Hyvönen, Eero

The ParliamentSampo Knowledge Graph includes data regarding Finnish Parliamentary debates and actors. The RDF data has been converted using data from the Parliament of Finland's open data services and Wikidata.

The Knowledge Graph contains harmonized data of

- speeches from the plenary sessions of the Finnish Parliament 1907–2022, and
- members and organizations of the Parliament.

The data model is designed for representing speeches, interruptions, items (on agenda), documents and other aspects related to plenary session speeches and minutes as well as member of the parliament and biographical information about them focusing on their political career.

This dataset is available on a public SPARQL endpoint (<http://rdf.fi/sempari/spargl>).

To test and demonstrate its usefulness, this Knowledge Graph is in use in the semantic portal ParliamentSampo, explained in more detail in the [project page](#).

The Knowledge Graph can be downloaded also as CSV and XML files. See the dataset page on LDF.fi for more details.

Preview

parliamentsampo-data-v1.0.0.zip

<input type="checkbox"/> actors.ttl	25.8 MB
<input type="checkbox"/> distances.ttl	5.6 MB
<input type="checkbox"/> eks.ttl	1.7 MB
<input type="checkbox"/> events.ttl	29.0 MB
<input type="checkbox"/> government_proposals.ttl	1.2 MB
<input type="checkbox"/> items_and_documents.ttl	94.6 MB
<input type="checkbox"/> in_judicial.ttl	877.5 MB
<input type="checkbox"/> occupations.ttl	231.5 kB
<input type="checkbox"/> places.ttl	3.0 MB
<input type="checkbox"/> publications.ttl	332.2 kB
<input type="checkbox"/> relations.ttl	83.7 kB
<input type="checkbox"/> schema.ttl	136.7 kB
<input type="checkbox"/> sessions_and_transcripts.ttl	10.4 MB
<input type="checkbox"/> shortcuts.ttl	2.2 GB
<input type="checkbox"/> sparql-service-description.ttl	23.8 kB
<input type="checkbox"/> speeches_and_interruptions.ttl	2.8 GB
<input type="checkbox"/> statuses_languages_speechtypes_and_parliamentary_roles.ttl	2.6 kB
<input type="checkbox"/> times.ttl	3.7 MB

Files (826.1 MB)

Name	Size	
parliamentsampo-data-v1.0.0.zip	826.1 MB	Preview Download
md5:d479f6328a905366ae91971c38b9893		

Citations 0

Show only: Literature (0) Dataset (0) Software (0) Unknown (0)

Citations to this version

Search

0 views 0 downloads

See more details...

Indexed in

OpenAIRE

Publication date: February 14, 2023

DOI: [10.5281/zenodo.7636420](https://doi.org/10.5281/zenodo.7636420)

Keyword(s): [parliamentary data](#) [linked data](#) [digital humanities](#) [RDF](#) [Semantic Web](#)

License (for files): [Creative Commons Attribution 4.0 International](#)

Versions

Version 1.0.0 Feb 14, 2023

[10.5281/zenodo.7636420](https://doi.org/10.5281/zenodo.7636420)

Cite all versions? You can cite all versions by using the DOI [10.5281/zenodo.7636419](https://doi.org/10.5281/zenodo.7636419). This DOI represents all versions, and will always resolve to the latest one. [Read more](#).

Share

[Home](#) [Twitter](#) [Facebook](#) [Email](#) [+](#)

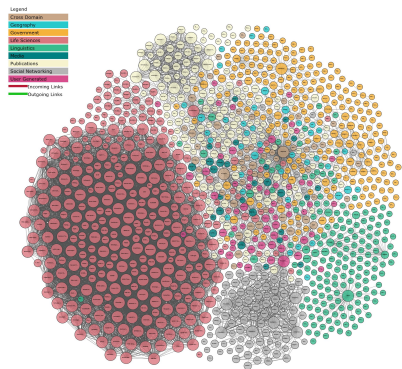
Cite as

Sinikallio, Laura, Leskinen, Petri, Drobac, Senka, Tamper, Minna, Leäl, Rafael, Tuominen, Jouni, Rantala, Heikki, La Mela, Matti, Koho, Mikko, & Hyvönen, Eero. (2023). ParliamentSampo Knowledge Graph (1.0.0) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.7636420>

Start typing a citation style...

Parlamenttisammon datapalvelu käytettävissä!

<http://ldf.fi/dataset/semparl>



ICT Solutions for Brilliant Minds

