



# ParliamentSampo – Finnish Parliament on the Semantic Web

Digital Humanities in Action:  
Sampo Model and Portals for Cultural Heritage, 29.10.2020

Jouni Tuominen

HELDIG – Helsinki Centre for Digital Humanities, University of Helsinki, <http://heldig.fi>

Semantic Computing Research Group (SeCo), Aalto University, <http://seco.cs.aalto.fi>

Päivikki Karhula

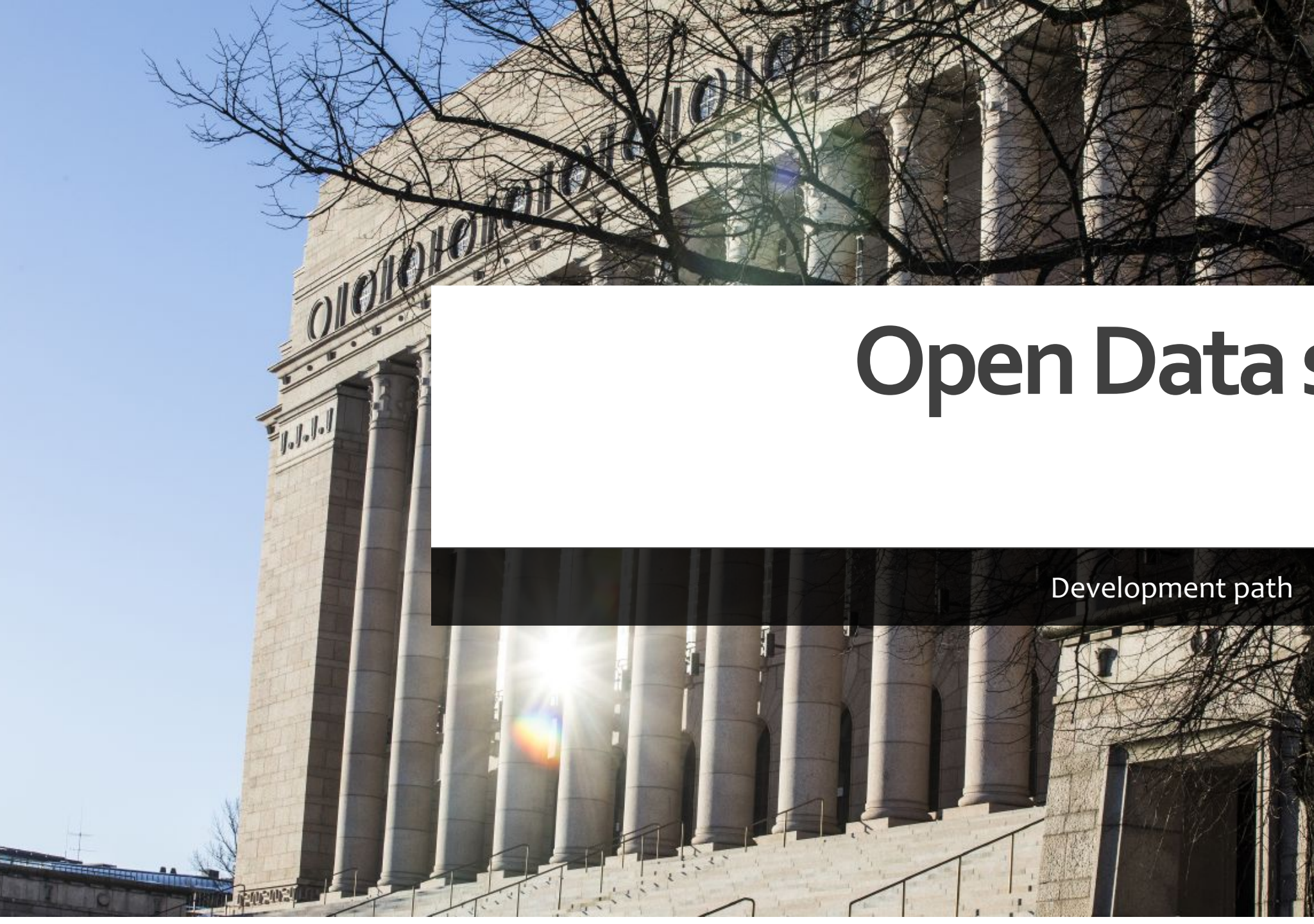
Library of Parliament, Finland, <https://www.eduskunta.fi>

# Outline

- **Open Data services** of the Parliament of Finland
  - Päivikki Karhula
- **ParliamentSampo vision and current status**

**Päivikki Karhula, Parliament of Finland**

**Open Data services**



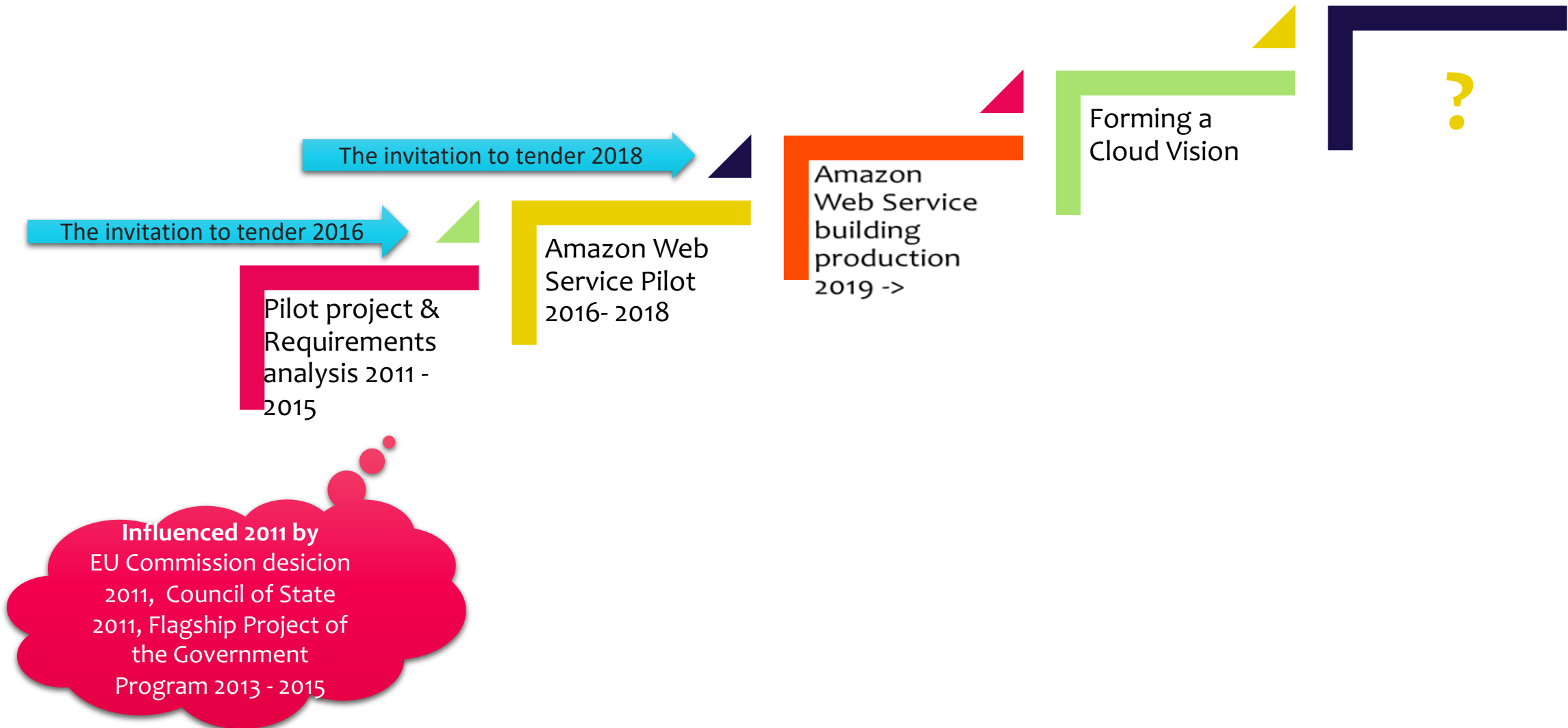
# Open Data services

Development path



EDUSKUNTA

# Open Data service





# Open Data service

Contents	Data format
Parliamentary documents (2015-)	XML
Voting results (1976-)	XML
Data on absence of plenary sessions and committee meetings (2015-)	PDF
MP's personal data and parliamentary action (2019-)	XML
Parliamentary documents (1907-2000)	PDF
Parliament's resolutions	JSON, CSV, XLS

# Customer Segments

**Citizens**



Want up-to-date information on current topics in an easy-to-understand format

**Experts**



More technically demanding, use larger data sets, know the content of documents and matters

**Data Professionals**



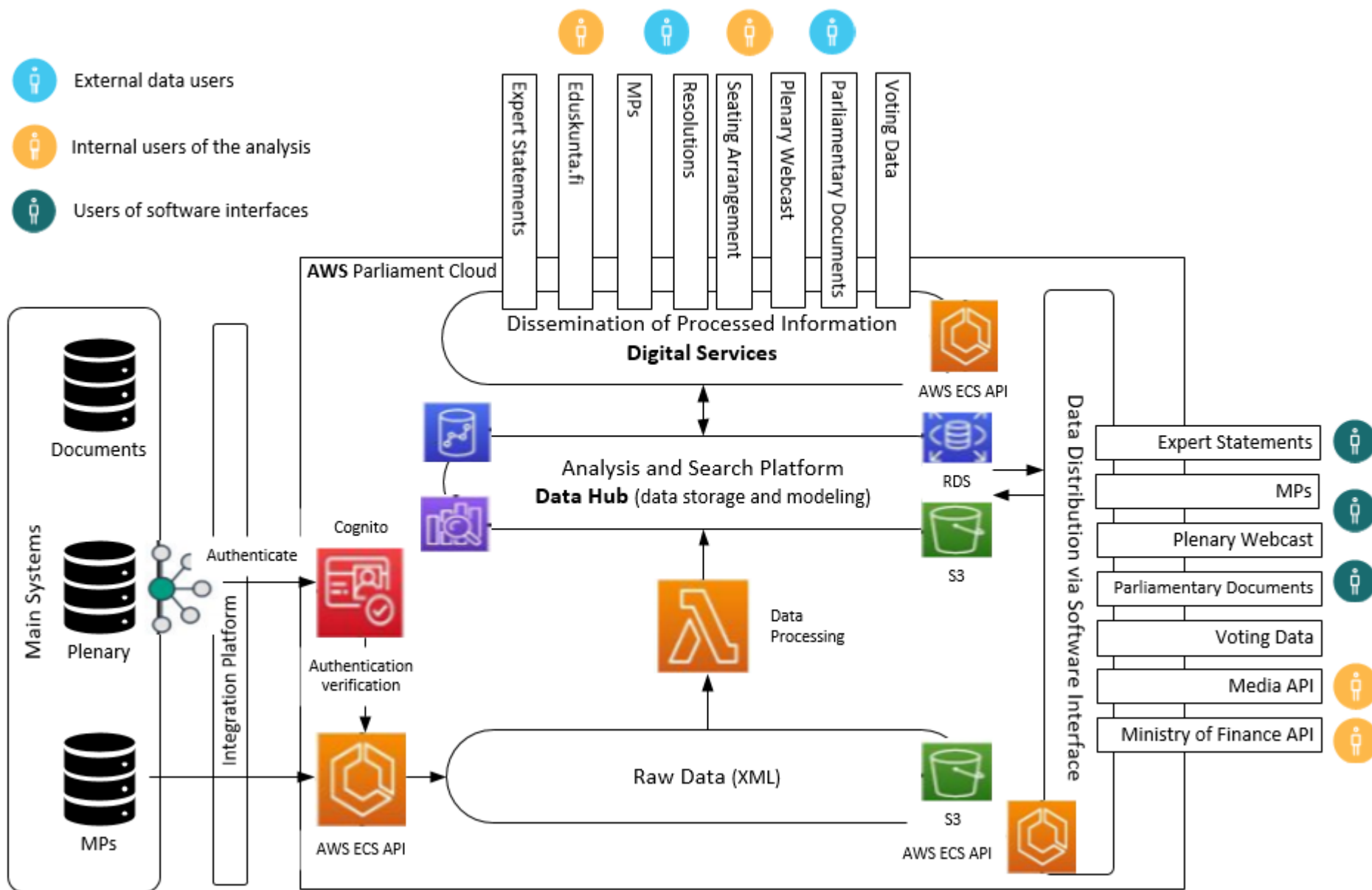
Use APIs, take advantage of automatic export functions

**Parliament Staff**



A demanding seeker of up-to-date and accurate factual and documentary information

# Architecture Plan





<http://avoindata.eduskunta.fi/>

# ParliamentSampo: Parliamentary Discussions, Texts, and Networks as Linked Open Data



Aalto University  
School of Science



UNIVERSITY OF HELSINKI



**HELDIG**  
Helsinki Centre for Digital Humanities

# Academy of Finland DIGIHUM project SEMPARL 2020–2022 consortium

- **University of Helsinki, Helsinki Centre for Digital Humanities (HELDIG)**
  - Consortium coordinator
  - Main focus: Language analysis and technologies
  - Eero Hyvönen (PI)
- **Aalto University, Department of Computer Science**
  - Main focus: Linked Data and Semantic Web technologies, AI
  - Data services and infrastructures
  - Jouni Tuominen (PI)
- **University of Turku, Centre for Parliamentary Studies**
  - Main focus: Political and Media Studies
  - Kimmo Elo (PI)



# **Example DH research questions/areas (special focus on time 1990–2020)**

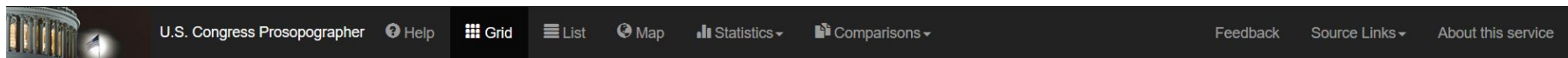
- **Changes in concepts: mainstream and marginalization**
- **Discursive networks over political party boundaries**
- **Parliamentary speech of MPs and its changes: language as a tool for politics**
- **Network analyses of politicians**

# Core datasets

- **Parliamentary debates at plenary sessions 1907–**  
+ enrichment with supporting documents:  
government proposals, committee reports, legislation, etc.
- **Members of Parliament**
  - Personal data, party, committee memberships, positions of trust, etc.
  - Absence, voting data



# US Congress dataset, our previous work in collaboration with Goki Miyakita, Keio University



## U.S. Congress Prosopographer

Biographical Repository of the U.S. Congress Legislators. This interface enables you to explore biographical and prosopographical data of the United States Congress members from the 1st through 115th Congresses (1789-2018).

Source: [Members of the United States Congress](#), [K7MOA Legacy Voteview Website](#)

Timeline: 1st (1789) - 115th (2018) Congresses

Remove all selections



<b>Search</b> <input type="text"/> <input type="button" value="X"/>	<b>Family Name</b> <input type="button" value="+"/>	<b>First Name</b> <input type="button" value="+"/>	<b>Gender</b> <input type="button" value="+"/>
<b>Chamber</b> <input type="button" value="Bar Chart"/> <div>-- No selection -- (11987) Representative (10001) Senator (1859)</div>	<b>Political Party</b> <input type="button" value="Bar Chart"/> <div>-- No selection -- (11987) Adams (85) Adams Democrat (1) American (43) American Labor (2)</div>	<b>Representing State</b> <input type="button" value="Bar Chart"/> <div>-- No selection -- (11987) AK (19) AL (206) AR (120) AS (3)</div>	<b>Links</b> <input type="button" value="Bar Chart"/> <div><input type="checkbox"/> Wikipedia (11862) <input type="checkbox"/> DBpedia (11860) <input type="checkbox"/> Twitter (524)</div>
<b>Occupation</b> <input type="button" value="+"/>	<b>Place of Birth</b> <input type="button" value="+"/>	<b>Serving Congress</b> <input type="button" value="+"/>	

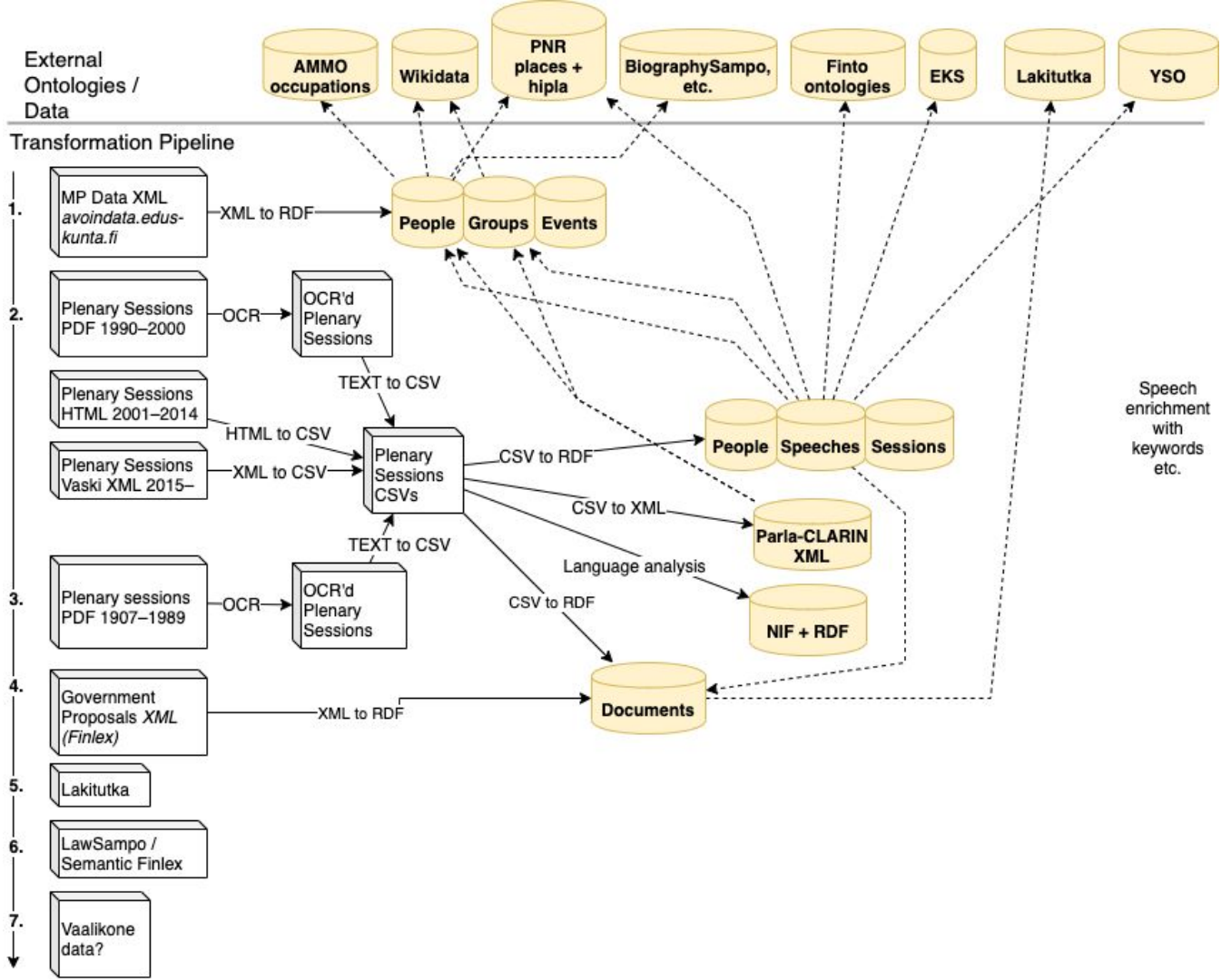
Sort by

Birthdate PAST PRESENT

Family Name

First Name





# Data production: status

- Re-OCR'd plenary session minutes (PDF) 1907–2000
  - Plenary session minutes 1940–2020 converted to harmonized structured format:
    - PDF (TXT) –2000
    - HTML 2001–2014
    - XML 2015–
- } → CSV, XML and Linked Data

10/1/16

PTK 10/2007 vp

7

osuutta, mitä selkeämmin palkkaratkaisu tukee näiden koulutettujen naisvaltaisten alojen, kunta-sektorin alojen palkkojen kehitystä.

Herra puhemies! Tämän paremmin ei hallitus-ohjelmassa, jossa ei päätetä palkankorotuseurois-ta, tämän paremmin ei hallitusohjelmassa olisi voitu kokoomuksen keskeistä vaalitavoitetta huomioida. Mitä tulee siihen väitteeseen, että me lupasimme hoitajille 500 euron palkankorotuk-sen kuukaudessa ja että siitä vaatimuksesta tai lu-pauksesta olisi livetty, kokoomus ei koskaan tien-tenkään missään ole luvannut 500 euroa kuukau-dessa (Välihuutoja — Ed. Tennilä: Kyllä lupasi kokoomuksen vaalimainoksissa!) hoitajille pal-kankorotusta. Te voitte lukea puheenjohtaja Ka-taisen tai ryhmänjohtaja Häkämiehen tai muiden

me tätä hallitusohjelman kohtaa lävitse. Siinä-hän ei ollut summista puhetta, niin kuin te viitta-sitte. Te viittasitte kunnallisiin palkkaratkaisui-hin, joita tullaan tekemään. Minä uskallan va-kuuttaa sen, että täällä oppositiossa me olemme valmiita tukemaan tätä prosessia kaikissa tilan-teissa. Kaikissa tilanteissa, ed. Zyskowitz. Me todellakin olemme tiukasti kiinni siinä, missä te-kin sanoitte, että äänestäjillä täytyy olla tietty ku-luttajansuoja suhteessa niihin lupauksiin, joita tehdään. Te olette puhuneet 500 eurosta. (Ed. Zyskowitz: Kaivakaa esiin se teksti, missä me olemme puhuneet!) Mekin olemme puolueena si-toutuneet siihen 500 euroon, ja me peräänkuulu-tamme täällä sitä asiaa. Me annamme kaiken mahdollisen tuen teille, kun te tulette sitten var-

alojen palkkojen kehitystä. Herra puhemies! Tämän paremmin ei hallitusohjelmassa, jossa ei päätetä palkankorotuseuroista, tämän paremmin ei hallitusohjelmassa olisi voitu kokoomuksen keskeistä vaalitavoitetta huomioida. Mitä tulee siihen väitteeseen, että me lupasimme hoitajille 500 euron palkankorotuksen kuukaudessa ja että siitä vaatimuksesta tai lupauksesta olisi livetty, kokoomus ei koskaan tietenkään missään ole luvannut 500 euroa kuukaudessa </u>


```
<vocal>  
  <desc>Välihuutoja — Ed. Tennilä: Kyllä lupasi kokoomuksen vaalimainoksissa!</desc>  
</vocal>
```

```
<u next="2007.10.16.7" prev="2007.10.16.5" who="#Ben_Zyskowitz" xml.id="2007.10.16.6"> hoitajille palkankorotusta. Te  
voitte lukea puheenjohtaja Kataisen tai ryhmänjohtaja Häkämiehen tai muiden kokoomuksen johtavien henkilöiden kaikki  
vaalien alla antamat haastattelut ja puheet, ja tällaista euromäärään, nimenomaan tähän euromäärään, yksittäiseen  
euromäärään, sidottua vaalilupausta ei mistään löydy. Iltalehti kysyi kaikilta puolueilta: 500 euroa hoitajille lisää  
kuukaudessa — kyllä vai ei? Siihen vastasi SDP, siihen vastasi vihreät, siihen vastasi kokoomus &quot;kyllä&quot;; mutta  
sen jälkeen siinä oli perustelut ja näistä perusteluista jokainen voi lukea myös Iltalehden toimituksessa, joka nyt tällä  
asialla riemuitsee, että me emme ole sitoutuneet tietenkään </u>
```

```
<vocal>  
  <desc>Välihuutoja</desc>
```

# Data modeling and ontology work

- Modeling parliamentary speeches
- Ontologies for subject indexing
  - Library of Parliament subject headings (EKS)
  - General Finnish ontologies (YSO et al.) from Finto service

 THE LIBRARY OF PARLIAMENT | **EKS subject headings**  
Instructions | Contents | Selma

suomi | svenska | English

Subject heading:

**ALPHABETICAL INDEX**

**CLASSIFIED INDEX**

- [06 Legal education](#)
- [07 Organizations of lawyers and legal profession](#)
- [09 Legal language and dictionaries](#)
- [11 Jurisprudence in general, philosophy of law, sociology of law,](#)
- [11 legal policy and comparative law](#)
- [12 Roman law](#)
- [13 History of law](#)
- [14 Civil law](#)
- [15 Commercial law](#)
- [16 Land-, water- and environmental law](#)
- [17 Labour law](#)
- [18 Criminal law](#)
- [19 Procedural law](#)
- [20 Constitutional law](#)



# Linked Data service



## 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 number 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. Our work on the [Linked Data](#) paradigm and stack of standards, which combines an expressive, semantic 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 consumers need more support in getting to know and access them, as well as a better grasp of the 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* 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

Query 1 x Query 1 x +

http://ldf.fi/warsa/sparql

```
8 # Events, photographs and articles that are situated in Vyborg
9 SELECT DISTINCT ?type ?resource ?prefLabel
10 WHERE {
11   {
12     # Events
13     BIND (<http://ldf.fi/warsa/places/municipalities/m_place_614> as ?vyborg)
14     ?resource crm:P7_took_place_at ?vyborg ;
15             a ?typeURI ;
16             crm:P4_has_time-span ?timespan .
17     FILTER(?typeURI != wet:Photography)
18     ?typeURI rdfs:subClassOf* crm:E5_Event .
19   }
20   UNION
21   {
```

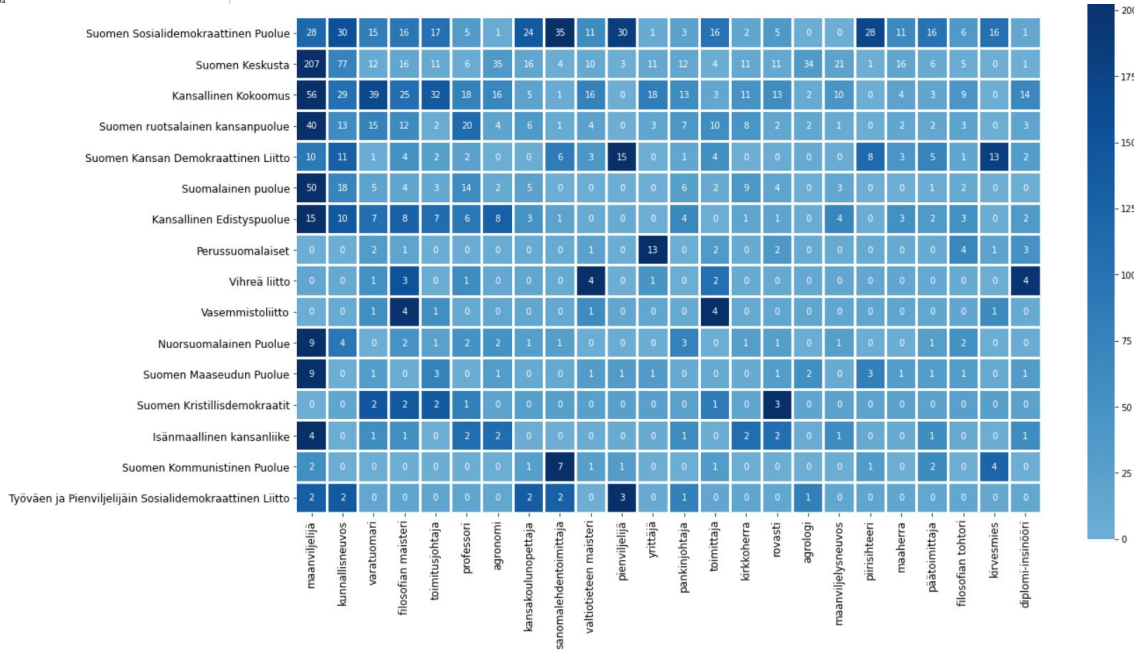
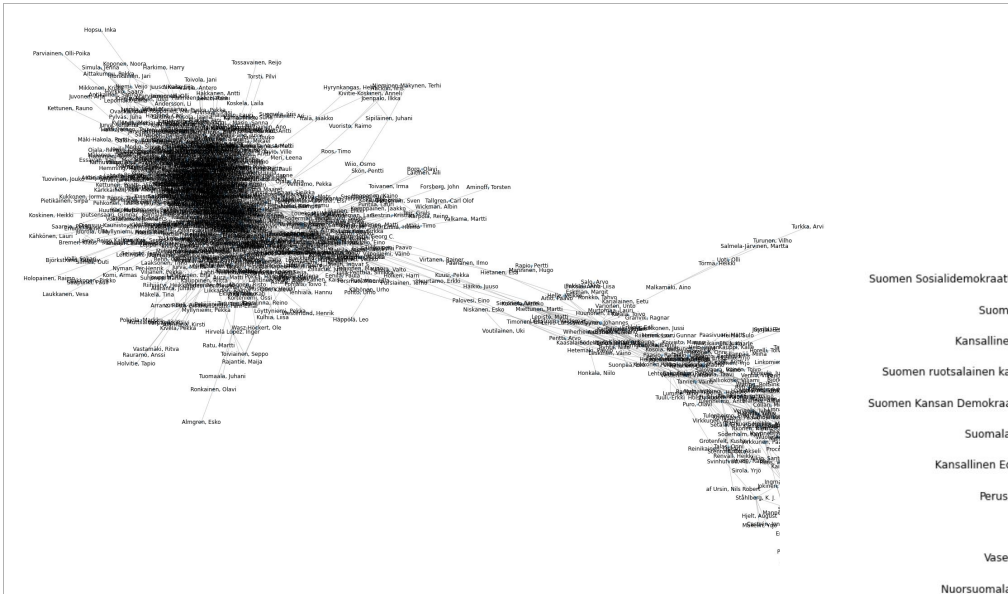
Table Response Pivot Table Google Chart Geo ↕ </>

Showing 1 to 50 of 11,147 entries (in 3.195 seconds)

type	resource	prefLabel
"Battle"@en	<a href="http://ldf.fi/warsa/events/event_1078">http://ldf.fi/warsa/events/event_1078</a>	"Defensive battles of Viipuri"@en
"Battle"@en	<a href="http://ldf.fi/warsa/events/event_1080">http://ldf.fi/warsa/events/event_1080</a>	"Viipuri"@en
"Battle"@en	<a href="http://ldf.fi/warsa/events/event_1092">http://ldf.fi/warsa/events/event_1092</a>	"Battle of Vyborg, Pappilanniemi and Kärenmäki sub-blocks"@en
"Battle"@en	<a href="http://ldf.fi/warsa/events/event_1092">http://ldf.fi/warsa/events/event_1092</a>	"Battle of Vyborg, Pappilanniemi and Huhtiala sub-blocks"@en



# Data analysis on Linked Data



# Sampo portal vision



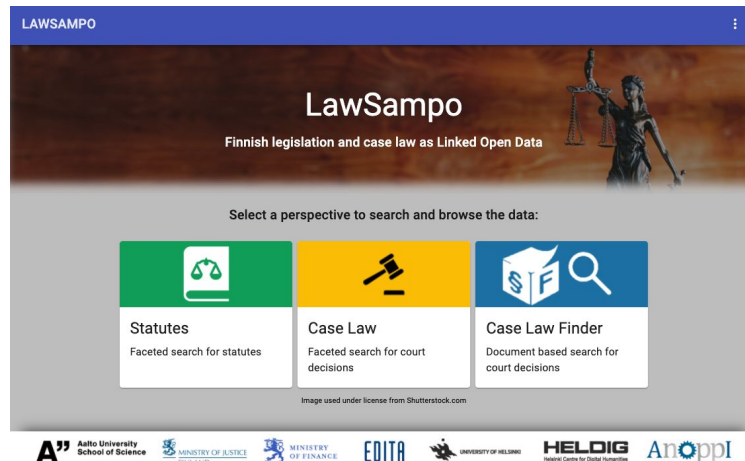
[ haku ] Täysistunnot Kansanedustajat Hallituksen esitykset | Palaute Info Ohjeet



## Parlamenttisampo

Eduskunta semanttisessa webissä

draft



Developing the Digital World Together  
**HELDIG**  
Helsinki Centre for Digital Humanities

**A"**  
Aalto-yliopisto

 **UNIVERSITY OF TURKU**

 **EDUSKUNTA RIKSDAGEN**