Further with Knowledge Graphs. M. Alam et al. (Eds.)
AKA Verlag and IOS Press, 2021

doi:10.3233/SSW210049

Members of Parliament in Finland Knowledge Graph and Its Linked Open Data Service

Petri Leskinen[1][0000−0003−2327−6942], Eero Hyvönen[1,2][0000−0003−1695−5840], and Jouni Tuominen[1,2][0000−0003−4789−5676]

[1] Semantic Computing Research Group (SeCo), Aalto University, Finland
[2] HELDIG – Helsinki Centre for Digital Humanities, University of Helsinki, Finland
http://seco.cs.aalto.fi, http://heldig.fi, firstname.lastname@aalto.fi

Abstract. This paper presents a prosopographical knowledge graph describing the Members of Parliament in Finland and related actors in politics, extracted from the databases and textual descriptions of the Parliament of Finland. The data has been interlinked internally and enriched with data linking to external data sources according to the 5-star Linked Data model. The data has been published together with its schema for better re-usability and is validated using ShEx. The knowledge graph presented is integrated with another knowledge graph about over 900 000 parliamentary plenary speeches in Finland (1907–) to form a larger parliamentary LOD publication *FinnParla* of the Parliament of Finland. The data is being used for Digital Humanities research on parliamentary networks, culture, and language.

Keywords: Parliamentary data · Biographies · Linked Data · Digital Humanities · Entity linking

1 Introduction

A key idea of Linked Data [1] is to enrich datasets by integrating complementary local information sources in an interoperable way into global knowledge graphs [2] to be used in applications. This involves harmonization of the local data models used, as well as aligning the concepts and entities (resources) used in populating the local data models.

This paper reports first results of the Semantic Parliament (SEMPARL)[3] project that produces a Linked Open Data (LOD) and research infrastructure for Finnish parliamentary data, and develops novel semantic computing technologies and applications to study parliamentary political culture and language. The project is related to various similar efforts in other countries [3,4,5] and in EU [6]; parliamentary open data is an important asset for rendering political decision making transparent, and such data is widely used for research on political language and culture.

SEMPARL aims at three major contributions:

[3] https://seco.cs.aalto.fi/projects/semparl/en/

1. The project responds to the demand for an easy to use and "intelligent" access to the newly digitized Finnish parliamentary data by providing the data as a national Linked Open Data (LOD) infrastructure and service for researchers, citizens, government, media, and application developers.
2. The project studies long-term changes in the Finnish parliamentary and political culture and language. These use cases are pioneering studies using the Finnish digital parliamentary data.
3. The new LOD service enriches semantically content in other related Finnish LOD services, such as LawSampo for Finnish legislation and case law [7] and BiographySampo [8] for prosopographical data.

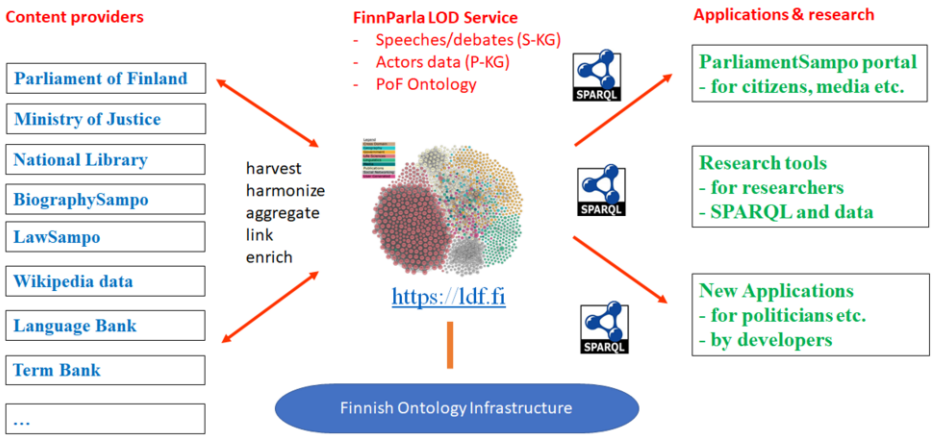

Fig. 1. Publishing model for Finnish parliamentary data in the SEMPARL project

The foundation of this work are two interlinked knowledge graphs (KG):

1. *S-KG* is a knowledge graph of all parliamentary debate speeches of Parliament of Finland (PoF) from 1907 to present time [9].
2. *P-KG* is a prosopographical knowledge graph of the Members of Parliament (MP), related other people, groups, and organizations, i.e., actors, pertaining to the parliamentary activities during the same period of time.

The two KGs are published as a LOD service called *FinnParla* about Parliament of Finland (PoF), based on an overarching ontology of PoF and the Finnish ontology infrastructure FinnONTO [10]. Fig. 1 illustrates the publishing model of SEMPARL. On the left, various content providing organizations and services are listed whose contents are transformed into, or linked with, the FinnParla LOD service in the middle. The data is used via SPARQL in research tasks and in developing applications on the right. These include the ParliamentSampo portal under development. The main data provider is PoF but also legislative data from

related sources are planned to be linked to FinnParla, such as the LawSampo data [7] from the Ministry of Justice. From the National Library, ontologies served at the Finto.fi[4] service are re-used and well as bibliographical data[5]. The Language Bank of Finland[6] contains, e.g., lots of videos of the debates, and The Helsinki Term Bank for the Arts and Sciences[7] terminological definitions pertaining to legislation and politics. The BiographySampo system contains 763 biographies of MPs as linked data as part of over 13 100 national biographies of the Finnish Literature Society. Wikipedia/Wikidata is used in various ways for enriching the FinnParla data. Possibly also media content from the Finnish Broadcasting Company Yle will be used in the project later on.

This paper introduces the prosopographical knowledge graph P-KG and addresses the following more general research question:

> *How to represent and publish prosopographical data about parliamentary actors and their activities so that the data can be used easily for Digital Humanities research?*

As an answer, the modeling principles of P-KG are presented and its transformation and publication processes are explained. It is also shown as a proof-of-concept how the LOD service can be used for Digital Humanities [11,12] research.

In the following, we first describe the original open XML data of PoF to be transformed into Linked Data. After this the RDF data model for representing parliamentary actors and their activities, as well as the transformation process are described. The produced linked data has been published as a data service using the 7-star model [13] of the Linked Data Finland platform. As a demonstration of using the data service in research, data analyses are presented. In conclusion, contributions of the work are summarized, related works are discussed, and directions for further research are outlined.

2 Parliament of Finland Actor Data

The main data source used for the P-KG is the Members of Finnish Parliament data publication available at the Parliament Open Data portal[8]. This data is regularly updated, and contains at this moment information about 2605 Members of Parliament (MP) since 1907. The person data entries are in XML format which is available in Finnish and Swedish for all the members, and in English for 202 cases.

An extract from the XML data is shown in Fig. 2. All the tags are in Finnish, and in the English version only the content is in English. A person data entry

[4] http://finto.fi/en/
[5] http://data.nationallibrary.fi
[6] https://www.kielipankki.fi/language-bank/
[7] https://tieteentermipankki.fi/wiki/Termipankki:Etusivu/en
[8] https://avoindata.eduskunta.fi/#/fi/dbsearch

```
<?xml version="1.0" ?>
<Henkilo kieliKoodi="FI" tyyppiKoodi="Kansanedustaja">
   <HenkiloNro>126</HenkiloNro>
   <EtunimetNimi>Elsi Maria</EtunimetNimi>
   <SukuNimi>Hetemäki–Olander</SukuNimi>
   <LajitteluNimi>hetemäki–olander elsi</LajitteluNimi>
   <KutsumaNimi>Elsi</KutsumaNimi>
   <MatrikkeliNimi>Hetemäki–Olander(e. Rinne, e. Hetemäki), Elsi Maria</MatrikkeliNimi>
   <Ammatti>Master of Arts, Councellor of Parliament</Ammatti>
   <SyntymaPvm>1927</SyntymaPvm>
   <SyntymaPaikka>Oulainen</SyntymaPaikka>
   ...
   < Vaalipiirit >
      < EdellisetVaalipiirit >
         < VaaliPiiri >
            <Nimi>Electoral District of Uusimaa</Nimi>
            <AlkuPvm>23.03.1970</AlkuPvm>
            <LoppuPvm>21.03.1991</LoppuPvm>
            <Tunnus>uus01</Tunnus>
         </ VaaliPiiri >
      </ EdellisetVaalipiirit >
   </ Vaalipiirit >
   ...
   <Edustajatoimet>
      <Edustajatoimi>
         <AlkuPvm>23.03.1970</AlkuPvm>
         <LoppuPvm>21.03.1991</LoppuPvm>
      </Edustajatoimi>
   </Edustajatoimet>
   ...
   <EdustajanJulkaisut>
      <EdustajanJulkaisu>
         <Nimi>Suomen vaikuttajanaisia</Nimi>
         <Vuosi>1977</Vuosi>
         <Tekijat/>
      </EdustajanJulkaisu>
   </EdustajanJulkaisut>
   ...
</Henkilo>
```

Fig. 2. Partial extract from XML data for the politician *Elsi Hetemäki-Olander*

contains biographical basic information, e.g., family name (*SukuNimi*) and given names (*EtunimetNimi*), places (*SyntymaPaikka*) and times (*SyntymaPvm*) of birth and death (if applicable), and vocations (occupations). In addition, there are detailed descriptions of the person's political, professional, and educational career. The text sample has three examples of career events: being a candidate in an electoral district (*Vaalipiiri*), being a Member of the Parliament (*Edustajatoimi*), and being a member in a parliamentary group (*Eduskuntaryhma*). These descriptions contain the label (*Nimi*) and id (*Tunnus*) of the related group and the start (*AlkuPvm*) and end (*LoppuPvm*) timestamps pertaining to the data. The data may also contain information about the publications authored by the person or about him/her. Due to privacy issues the data does not contain family-related information about the spouses and children of the politicians in contrast to many other biographical dictionaries.

3 Data Model for Parliamentary Actors and Events

To represent the biographical information about MPs and other politicians the data model presented in Fig. 3 was developed. The key idea of the model is to represent an actor's life and activities as a sequence of events (*bioc:Event*) in places (*crm:E53_Place*) and in time (*:Timespan*) with the actors (*bioc:Person*) participating in different roles (*bioc:Actor_Role*), such as *:Member*, *:Representative*, etc. The data model follows the Bio CRM [14] ontology, an extension of CIDOC CRM[9] for representing biographical information based on role-centric modeling. Bio CRM makes a distinction among attributes, relations, and events, where entities participate in different roles in a qualified manner. The namespaces used in the model are described in the figure on the left. In this extended model, there are almost 200 different roles in use. The data model has been populated by using a set of domain ontologies, such as places based on YSO places[10], groups and organizations (harvested from the data), and vocations based on the AMMO ontology [15].

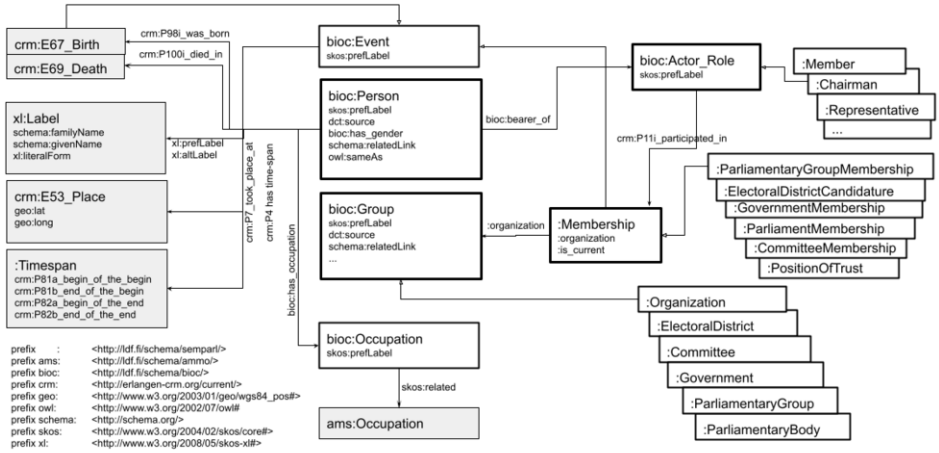

Fig. 3. Schema for the P-KG knowledge graph based on Bio CRM

For example, the XML data about the MP Elsi Hetemäki-Olander in Fig. 2 is translated into the RDF depicted in Fig. 4. Samples of extracted roles and events related to her life by the property *bioc:bearer_of* are listed in Fig. 5. As an example, *event:e2044* defines her the role of being a representative relating to *event:e2043*, being a MP during the time March 23 1970 to March 21 1991.

[9] https://cidoc-crm.org
[10] https://finto.fi/yso-paikat/en/

```
PREFIX bioc:     <http://ldf.fi/schema/bioc/>
PREFIX crm:      <http://erlangen-crm.org/current/>
PREFIX event:    <http://ldf.fi/semparl/event/>
PREFIX label:    <http://ldf.fi/semparl/label/>
PREFIX occupations: <http://ldf.fi/semparl/occupations/>
PREFIX people:   <http://ldf.fi/semparl/people/>
PREFIX rdf:      <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:     <http://www.w3.org/2000/01/rdf-schema#>
PREFIX roles:    <http://ldf.fi/semparl/roles/>
PREFIX schema:   <http://schema.org/>
PREFIX semparl: <http://ldf.fi/schema/semparl/>
PREFIX skos:     <http://www.w3.org/2004/02/skos/core#>
PREFIX times:    <http://ldf.fi/semparl/times/>
PREFIX xl:       <http://www.w3.org/2008/05/skos-xl#>

people:p126 a bioc:Person ;
    crm:P98i_was_born event:b126 ;
    bioc:bearer_of event:e2044, ..., event:e2050 ;
    bioc:has_gender schema:Female ;
    bioc:has_occupation occupations:o32, occupations:o95 ;
    semparl:authored publications:b114 ;
    semparl:id "126" ;
    schema:relatedLink <https://www.eduskunta.fi/FI/kansanedustajat/Sivut/126.aspx> ;
    skos:prefLabel "Hetemäki-Olander, Elsi (1927-)"@fi ;
    xl:altLabel label:l52, label:l53, label:l54 ;
    xl:prefLabel label:l51 .

label:l51 a xl:Label ;
    schema:familyName "Hetemäki-Olander" ;
    schema:givenName "Elsi" ;
    skos:prefLabel "Hetemäki-Olander, Elsi"@fi .

label:l53 a xl:Label ;
    schema:familyName "Rinne" ;
    schema:givenName "Elsi Maria" ;
    skos:prefLabel "Rinne, Elsi Maria"@fi .
```

Fig. 4. Partial extract from RDF data for the politician *Elsi Hetemäki-Olander*

4 Transformation of Parliamentary Actor Data into a KG

The data contains in total 2800 person entries, i.e., instances of the class *bioc:Person*. Out of this, 2605 are MPs from the main data source Parliament Open Data portal. This data was further enriched with data extracts from the web pages of the Finnish Government[11] and Wikidata in order to account for other people mentioned in the data and in the parliamentary speeches dataset S-KG [9] integrated with the P-KG. These ca. 200 additional resources are important people mentioned in the documents, such as Presidents of Finland, Ministers, or Parliamentary Ombudsmen[12] who have never been elected as MPs and therefore are not included in the MP database.

In addition to the people (*bioc:Person*), the groups and organizations (*bioc:Group*) mentioned in the XML data elements where extracted, disambiguated, and linked to the corresponding resources in the ontologies used. These groups contain the related parliamentary bodies and committees, governments,

[11] https://valtioneuvosto.fi

[12] https://www.oikeusasiamies.fi/en/web/guest

```
event:e2044 a roles:r1 ;
    crm:P11i_participated_in event:e2043 ;
    skos:prefLabel "edustaja Hetemäki—Olander"@fi .

event:e2043 a semparl:ParliamentMembership ;
    crm:P4_has_time_span times:t814 ;
    skos:prefLabel "edustajuus 23.03.1970—21.03.1991"@fi .

event:e2050 a roles:r166 ;
    crm:P11i_participated_in event:e2049 ;
    skos:prefLabel "ehdokas Hetemäki—Olander"@fi .

event:e2049 a semparl:ElectoralDistrictCandidature ;
    crm:P4_has_time_span times:t814 ;
    semparl:is_current false ;
    semparl:organization districts:uus01 ;
    skos:prefLabel "ehdokas: Uudenmaan läänin vaalipiiri "@fi .

districts:uus01 a semparl:ElectoralDistrict ;
    skos:prefLabel "Uusimaa constituency"@en,
        "Uudenmaan läänin vaalipiiri "@fi,
        "Nylands läns valkrets "@sv .

publications:b114 a semparl:Publication ;
    crm:P4_has_time_span times:t576 ;
    skos:prefLabel "Suomen vaikuttajanaisia" .
```

Fig. 5. Samples of resourses relating to the politician *Elsi Hetemäki-Olander*

electoral districts, and furthermore also groups out of political fields, such as companies, schools, and colleges. Also references to vocations (*bioc:Occupation*) were identified and linked to the resources of the AMMO ontology of historical occupations.

As a method for knowledge extraction, patterns of regular expressions were applied to the XML data fields, especially when extracting the person name variations and expressions of time. The source data contained all terms in Finnish. In addition, also the corresponding terms in English (1710) and Swedish (5420) were extracted. In the XML only recent data entries had translations in English. Since the main XML data came from a curated database, entities could be extracted with high precision and recall.

Table 1 summarizes the number of instances of the main classes of the data model of Fig. 3, and Table 2 lists the number of different event types extracted.

For validating the transformed P-KG data, the data model and its integrity constraints are presented in a machine-processable format using the ShEx Shape Expressions language[13]. We have made initial validation experiments with the PyShEx[14] validator. Based on the experiments, we have identified errors both in the schema and the data. We plan a full-scale ShEx validation phase integrated in the data conversion and publication process to spot and report errors in the dataset.

[13] https://shex.io
[14] https://github.com/hsolbrig/PyShEx

Table 1. Resources

Resource type	Count
Timespan	9168
Label	6061
Person	2801
Publication	1727
School, College	669
Place	607
Vocation	104
Parliamentary Group	89
Government	76
Committee	54
Organization	54
Electoral District	46
Parliamentary Body	38
Party	32
Ministry	12
Affiliation Group	10

Table 2. Events

Event type	Count
Career Event	14371
Position of Trust	12761
Committee Membership	6344
Municipal Position of Trust	4745
Event of Education	3712
Birth	2801
Electoral District Candidature	2205
Death	2025
Parliamentary Group Membership	1966
Government Membership	1622
Governmental Position of Trust	1621
Affiliation	1331
Parliament Membership	966
Honourable Mention	543
International Position of Trust	364
Membership Suspension	25

5 Prosopographical Data as a Linked Open Data Service

The prosopographical data P-KG presented above and the accompanying data model RDF schema have been published on the Linked Data Finland platform[15] [13] according to the Linked Data publishing principles and other best practices of W3C [1], including, e.g., content negotiation and provision of a SPARQL[16] endpoint.

In our work, the "FAIR guiding principles for scientific data management and stewardship" of publishing Findable, Accessible, Interoperable, and Re-usable data are used[17]. The data can be used via the SPARQL endpoint in two ways. Firstly, the underlying SPARQL endpoint can and is being applied to custom data analyses in Digital Humanities research using tools, such as YASGUI, Google Colab, and Jupyter notebooks. Secondly, a portal called *ParliamentSampo – Finnish Parliament on the Semantic Web* is under development, a new member in the "Sampo series" of semantic portals and LOD services[18]. The portal is targeted to both researchers and the public for studying parliamentary debates, the language used, networks of Finnish politicians, and political culture. ParliamentSampo is based on the Sampo model [16] for sharing collaboratively enriched linked open data, using a shared ontology infrastructure.

[15] https://ldf.fi
[16] https://www.w3.org/TR/sparql11-query/
[17] https://www.go-fair.org/fair-principles/
[18] https://seco.cs.aalto.fi/applications/sampo/

The SPARQL endpoint is hosted on an Apache Jena Fuseki[19] SPARQL server. The LDF platform provides dereferencing of URIs for both human users and machines, and a generic RDF browser for technical users, which opens when a URI is visited directly with a web browser. The URI routing, content negotiation, and caching is implemented using the Varnish Cache web application accelerator[20]. The LDF data service is based on a microservice architecture, using Docker containers[21]. Each individual component (Fuseki with the KG data and Varnish) is run in its own dedicated container, making the deployment of the services easy due to installation of software dependencies in isolated environments, enhancing the portability of the services. The data and the service are currently used internally in the SEMPARL project but will be opened by the CC BY 4.0 license to external users later on.

6 Using the SPARQL Endpoint for Data Analysis

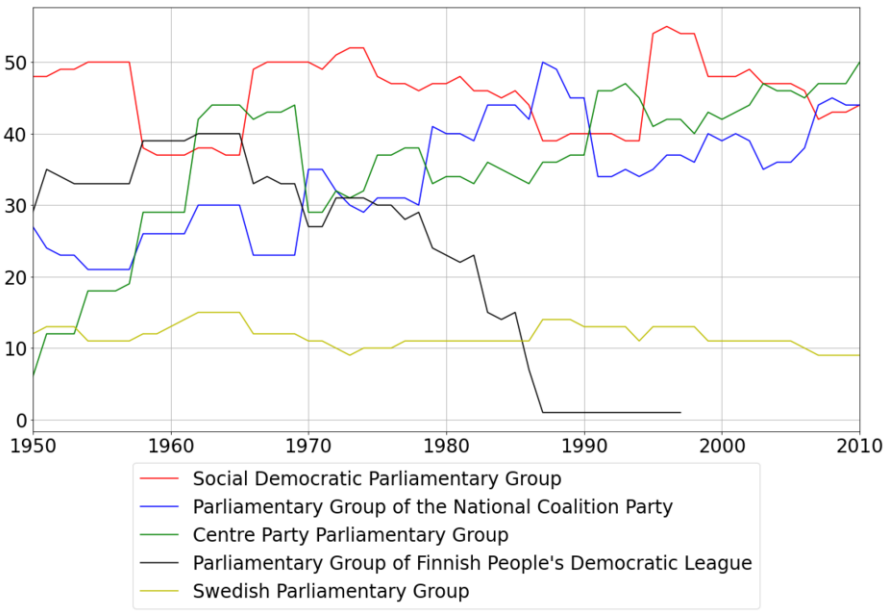

Fig. 6. Number of MPs of five most common Parliamentary Groups on a timeline

This section illustrates how the P-KG data can be used in researching the parliamentary culture in Finland, as suggested in Fig. 1.

[19] https://jena.apache.org/documentation/fuseki2/
[20] https://varnish-cache.org
[21] https://www.docker.com

A typical question in politics is to find out or forecast popularity of parties among the voters. Such data is available for recent times but not for historical times. By using P-KG such questions can be answered starting from 1907. For example, Fig. 6 depicts the number of MPs of the five most common Finnish parties during the years 1950–2010. The curves show how the *Social Democratic, National Coalition*, and *Centre Party* constantly share the top three positions. However, the *Finnish People's Democratic League* had a significant number of representatives from 1950's to the end of 1980's; the party was later replaced by the *Left Allience*. Furthermore, during the entire period of time, the *Swedish Parliamentary Group* has had an almost constant number of MPs.

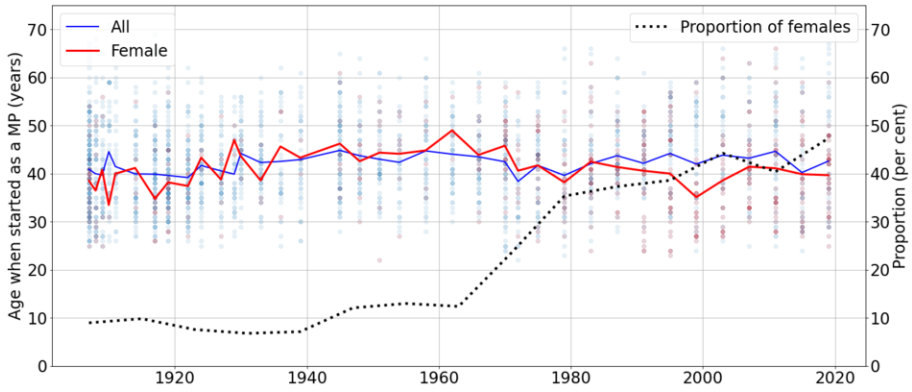

Fig. 7. Timeline with average ages of new MPs

Fig. 7 depicts on a timeline the ages of people when they were selected as MPs for the first time. The blue curve shows the average age for all MPs, and the red curve for female MPs. The values are calculated in time windows of four years. The black dotted line shows the relative proportion of female MPs in percents. It can been observed that between the years 1960 and 1980 this proportion constantly grows approximately from 10% to 40%. Generally, the average age of entering the Parliament is 42.1 which remains relatively constant during the entire timeline. However, after the 1980's the new female representatives are a few years younger than the men.

An interesting part of the P-KG is information about the vocations of the people, based on the AMMO ontology that has been aligned with the international HISCO classification[22] [17]. It provides an international comparative classification system of history of work, particularly for occupational titles in the 19th and early 20th centuries. HISCO encodes not only occupations, but also information about prestige, property, and family relations can be included. As an example of a data analysis based on vocations, Fig. 8 depicts a correlation matrix between

[22] https://iisg.amsterdam/en/data/data-websites/history-of-work

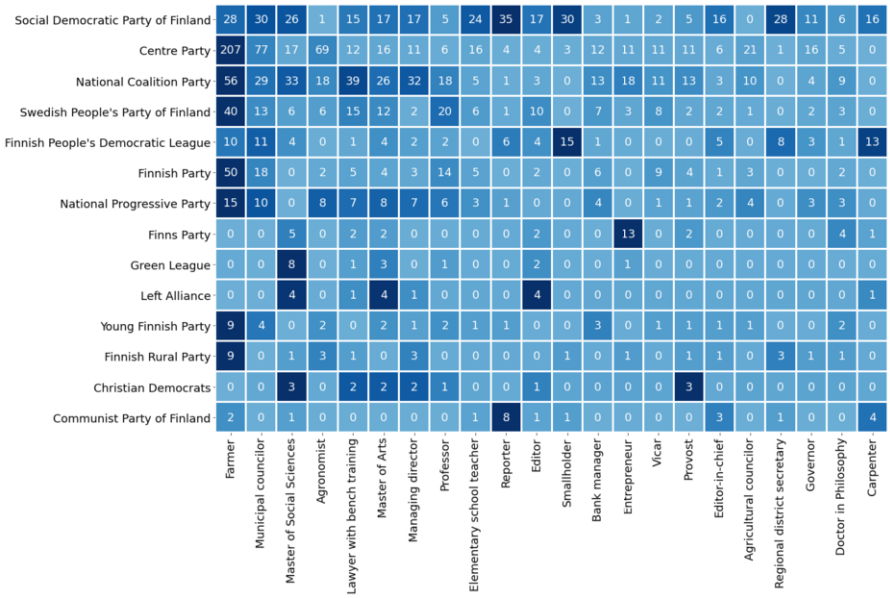

Fig. 8. Correlations between parties and vocations

the parties and vocations of the MPs. In the figure, the rows correspond to the ten parties with most MPs and columns to their vocations. The figure shows the vocations during the entire time period from 1907 to 2021. Finland was a before the Second World War a rural country, which explains why the vocation Farmer is on the first rank. From the results it can be noticed that, e.g., *Smallholder* and *Carpenter* have been common vocations for MPs of *the Finnish People's Democratic League* or *Provost* and *Master of Social Sciences* are common among the MPs of *the Christian Democrats*.

As a final example of data analysis, Fig. 9 depicts a correlation matrix between the parties and committees of PoF. In this figure, each row corresponds to a party and each column to a committee. The darker the cell background color is, the more members of that party haves been in the corresponding committee. Generally, the largest committee *the Grand Committee* has had a large amount of members from most of the parties. It can be noticed that, e.g., *the Finns Party* has had more members in *the Legal Affairs Committee* and *the Swedish People's Party of Finland* in *the Finance Committee*. The data model facilitates to easily analyze similar correlations between, e.g., parties, vocations, or genders.

These data analyses and visualizations were created easily by using a SPARQL query and then analysing its result with Python scripting and libraries on Google Colab Jupyter documents. According to our experiences in these and several other examples, the underlying data model and the populated data seems useful, semantically rich, and complete enough for studying political culture in versatile

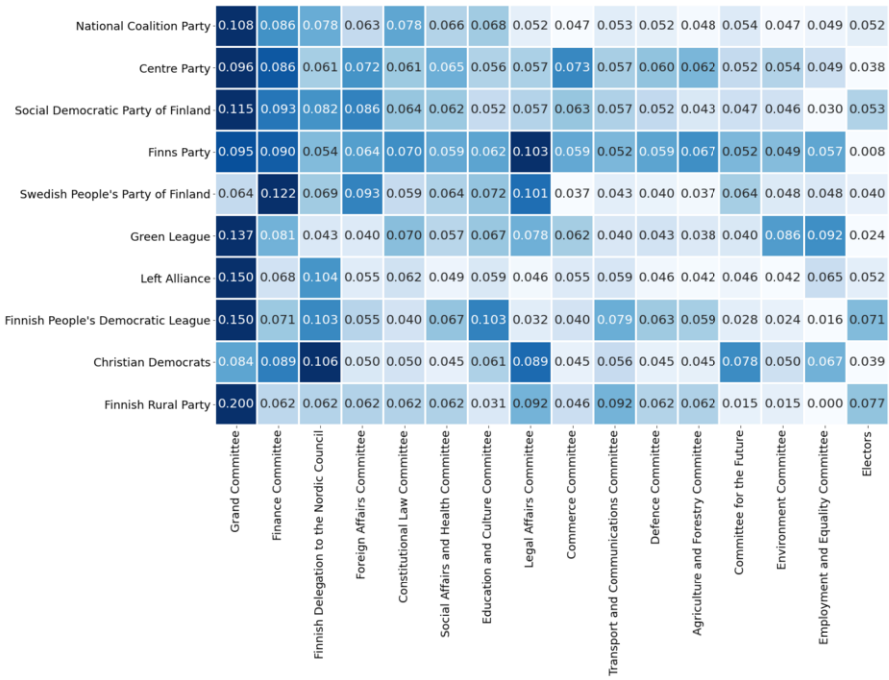

Fig. 9. Correlations between parties and committees of PoF

ways. In order to get feedback from external users, too, the data was used in the Helsinki Digital Humanities Hackathon in May 2021 for research purposes.

Of course, the data is limited to what is openly available from PoF and to additional data and links aggregated into the P-KG from related data sources during the data transformation into RDF. When using a dataset such as P-KG, where much of the content has been created or transformed automatically, new kind of data literacy [18] is needed when interpreting the results. Tools based on distant reading [19] are good for finding and exploring efficiently interesting patters of information in the data but for the final interpretation and error analysis close reading is needed, too.

7 Discussion

Related Work Many national projects have transformed parliamentary data[23], such as plenary session debates [9], into structured formats and enriched the data with biographical metadata, including, e.g., the Canadian Lipad project [3] and the Norwegian Talk of Norway [4]. Linked data has also been used in some

[23] See the CLARIN page www.clarin.eu/resource-families/parliamentary-corpora for a list of various national parliamentary corpora projects.

works, such as the LinkedEP about the European Parliament linked data 1999–2017 [6], the Latvian LinkedSAEIMA project [5], and the Italian Parliament[24]. Speech data can be used for analysing the language and topics of speeches (cf. e.g. [20,21,22]) and also the activities of the parliament and networks of its members. For example, speeches of male and female MPs or other groups, such as political parties, can be analyzed and compared [23].

The P-KG is in nature a biographical dictionary even if focused on parliamentary data and events. The idea of analysing such proposographical data quantitatively, as was illustrated in section 6, have been already made for some national dictionaries of biography, such as for the British ODNB [24] and the Irish Ainm [25]. As is [26], our goal is to combine quantitative approach and distant reading methods with the qualitative approach, often based on close reading, typical to biographical research.

Contributions This paper introduced the first Linked Data model and publication of the Finnish parliament actor data, covering the whole history of PoF since 1907. In comparison to related works, the underlying data model is arguably unique in employing the semantically rich event-based ontology model presented for harmonizing data about the politicians and their lives, extending CIDOC CRM to representing prosopographical data. Our experience on developing biographical Sampo systems [8] suggests that an event-based approach is needed for integrating biographical data of different kinds instead of using only traditional document-centric models, such as Dublin Core. Furthermore, the actor data is enriched and interlinked with several additional external data sources, and is based on a national level ontology infrastructure [10] for even more extensive interlinking. The first experiments presented in using the data service for Digital Humanities research suggest that the model is fit for its purpose and can be used effectively in SPARQL queries for visualizations and parliamentary data analyses, and for creating the large Finnish parliamentary debate dataset [9] and the larger *FinnParla* LOD cloud.

Future Research Digital humanities studies are underway in the Semantic Parliament project project using the P-KG interlinked with its sister dataset S-KG about the Finnish parliamentary debates. The P-KG will also be used as part of the semantic portal *ParliamentSampo – Finnish Parliament on the Semantic Web* that is being developed based on the Sampo model [16] and the Sampo-UI framework [27].

Acknowledgements Thanks to Ari Apilo, Sari Wilenius, and Päivikki Karhula at the Parliament of Finland for co-operation, and the Semantic Parliament project team for discussions. Our work was funded by the Academy of Finland as part of the Semantic Parliament project, the EU project InTaVia: In/Tangible European Heritage[25], and is related to the COST action NexusLinguarum[26] on linguistic data science. CSC – IT Center for Science, Finland, provided computational resources for the work.

[24] http://data.camera.it/data/en/datasets/
[25] https://intavia.eu
[26] https://nexuslinguarum.eu

References

1. Heath T, Bizer C. Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan & Claypool; 2011. Available from: http://linkeddatabook.com/editions/1.0/.
2. Gutierrez C, Sequeda JF. Knowledge Graphs. Commun ACM. 2021 Feb;64(3):96—104. Available from: https://doi.org/10.1145/3418294.
3. Beelen K, Thijm TA, Cochrane C, Halvemaan K, Hirst G, Kimmins M, et al. Digitization of the Canadian Parliamentary Debates. Canadian Journal of Political Science. 2017;50(3):849-64.
4. Lapponi E, Søyland MG, Velldal E, Oepen S. The Talk of Norway: a richly annotated corpus of the Norwegian parliament, 1998–2016. Language Resources and Evaluation. 2018 Sep;52(3):873-93.
5. Bojārs U, Darģis R, Lavrinovičs U, Paikens P. LinkedSaeima: A Linked Open Dataset of Latvia's Parliamentary Debates. In: Acosta M, Cudré-Mauroux P, Maleshkova M, Pellegrini T, Sack H, Sure-Vetter Y, editors. Semantic Systems. The Power of AI and Knowledge Graphs. Cham: Springer; 2019. p. 50-6.
6. van Aggelen A, Hollink L, Kemman M, Kleppe M, Beunders H. The debates of the European Parliament as Linked Open Data. Semantic Web. 2017;8(2):271-81.
7. Hyvönen E, Tamper M, Oksanen A, Ikkala E, Sarsa S, Tuominen J, et al. LawSampo: A Semantic Portal on a Linked Open Data Service for Finnish Legislation and Case Law. In: The Semantic Web: ESWC 2020 Satellite Events. Revised Selected Papers. Springer; 2019. p. 110-4.
8. Hyvönen E, Leskinen P, Tamper M, Rantala H, Ikkala E, Tuominen J, et al. BiographySampo - Publishing and Enriching Biographies on the Semantic Web for Digital Humanities Research. In: The Semantic Web. ESWC 2019. Springer; 2019. p. 574-89.
9. Sinikallio L, Drobac S, Tamper M, Leal R, Koho M, Tuominen J, et al. Plenary Debates of the Parliament of Finland as Linked Open Data and in Parla-CLARIN Markup. In: Proceedings, Language, Data and Knowledge (LDK 2021); 2021. Forth-coming.
10. Hyvönen E, Viljanen K, Tuominen J, Seppälä K. Building a National Semantic Web Ontology and Ontology Service Infrastructure—The FinnONTO Approach. In: Proceedings of the 5th European Semantic Web Conference (ESWC 2008). Springer; 2008. p. 95-109.
11. McCarty W. Humanities Computing. Palgrave, London; 2005.
12. Gardiner E, Musto RG. The Digital Humanities: A Primer for Students and Scholars. New York, NY, USA: Cambridge University Press; 2015. https://doi.org/10.1017/CBO9781139003865.
13. Hyvönen E, Tuominen J, Alonen M, Mäkelä E. Linked Data Finland: A 7-star Model and Platform for Publishing and Re-using Linked Datasets. In: Presutti V, Blomqvist E, Troncy R, Sack H, Papadakis I, Tordai A, editors. The Semantic Web: ESWC 2014 Satellite Events. ESWC 2014. Springer-Verlag; 2014. p. 226-30. Available from: https://doi.org/10.1007/978-3-319-11955-7_24.
14. Tuominen J, Hyvönen E, Leskinen P. Bio CRM: A Data Model for Representing Biographical Data for Prosopographical Research. In: Proceedings of the Second Conference on Biographical Data in a Digital World 2017 (BD2017). vol. 2119. CEUR Workshop Proceedings; 2018. p. 59-66. Available from: http://ceur-ws.org/Vol-2119/paper10.pdf.

15. Koho M, Gasbarra L, Tuominen J, Rantala H, Jokipii I, Hyvönen E. AMMO Ontology of Finnish Historical Occupations. In: Proceedings of the First International Workshop on Open Data and Ontologies for Cultural Heritage (ODOCH'19). vol. 2375. CEUR Workshop Proceedings; 2019. p. 91-6. Available from: http://ceur-ws.org/Vol-2375/.
16. Hyvönen E. "Sampo" Model and Semantic Portals for Digital Humanities on the Semantic Web. In: DHN 2020 Digital Humanities in the Nordic Countries. Proc. of the Digital Humanities in the Nordic Countries 5th Conference. CEUR Workshop Proceedings, vol. 2612; 2020. p. 373-8. Available from: http://ceur-ws.org/Vol-2612/poster1.pdf.
17. Van Leeuwen MHD, Maas I, Miles A. HISCO: Historical international standard classification of occupations. Leuven University Press; 2002.
18. Koltay T. Data literacy for researchers and data librarians. Journal of Librarianship and Information Science. 2011;49(1):1-14.
19. Moretti F. Distant reading. Verso Books; 2013.
20. Greene D, Cross JP. Exploring the Political Agenda of the European Parliament Using a Dynamic Topic Modeling Approach. Political Analysis. 2017;25(1):77-94.
21. Won M, Martins B, Raimundo F. Automatic extraction of relevant keyphrases for the study of issue competition. EasyChair; 2019. EasyChair Preprint no. 875.
22. Kettunen K, La Mela M. Digging Deeper into the Finnish Parliamentary Protocols – Using a Lexical Semantic Tagger for Studying Meaning Change of Everyman's Rights (allemansrätten). In: DHN 2020 Digital Humanities in the Nordic Countries. Proc. of the Digital Humanities in the Nordic Countries 5th Conference. CEUR Workshop Proceedings, vol. 2612; 2020. p. 63-80. Available from: http://ceur-ws.org/Vol-2612/paper5.pdf.
23. Blaxill L, Beelen K. A Feminized Language of Democracy? The Representation of Women at Westminster since 1945. Twentieth Century British History. 2016;27(3):412-49.
24. Warren C. Historiography's Two Voices: Data Infrastructure and History at Scale in the Oxford Dictionary of National Biography (ODNB). Journal of Cultural Analytics. 2018.
25. Bhreathnach Ú, Burke C, Mag Fhinn J, Ó Cleircín G, Ó Raghallaigh B. A quantitative analysis of biographical data from Ainm, the Irish-language Biographical Database. In: Proceedings of the Third Conference on Biographical Data in a Digital World (BD 2019). CEUR-WS; 2019. Forth-coming.
26. Jänicke S, Franzini G, Cheema MF, Scheuermann G. Visual text analysis in digital humanities. Computer Graphics Forum. 2017;36(6):226-50.
27. Ikkala E, Hyvönen E, Rantala H, Koho M. Sampo-UI: A Full Stack JavaScript Framework for Developing Semantic Portal User Interfaces. Semantic Web – Interoperability, Usability, Applicability. 2021. Accepted.