U.S. Congress Prosopographer
–A Tool for Prosopographical Research of Legislators

Goki Miyakita[1], Petri Leskinen[2], and Eero Hyvönen[2,3]

[1] Research Institute for Digital Media and Content (DMC), Keio University, Japan
[2] Semantic Computing Research Group (SeCo), Aalto University, Finland
[3] HELDIG – Helsinki Centre for Digital Humanities, University of Helsinki, Finland
http://seco.cs.aalto.fi, http://heldig.fi

1 Prosopographical Method

Person registries and biographies are widely used to document and describe life stories of historical people, with the aim of getting a better understanding of their personality, actions, and motivations in history. In *biography* [4] the focus is on individual protagonists, while in *prosopography* [5] life histories of groups of people are studied in order to find out some kind of commonness or average in them. The prosopographical research method [5, p. 47] consists of two steps. First, a target group of people is selected that share desired characteristics for solving the research question at hand. Second, the target group is analyzed and compared with other groups to solve the research question.

This paper shows how the prosopographical method can be used in practice in Digital Humanities by presenting a tool and application based on the Linked Data (LD) paradigm [1]. It is shown how faceted search and data visualization tools can be integrated with a SPARQL endpoint allowing the end user to 1) filter out target groups of people, and 2) then to study them. A key novelty of this paper is the idea to support comparing analyses and visualizations based on different target subgroups.

As a use case, a database about the United States Congress Legislators[4,5] is used. We pulled and linked two different datasets: 1) a dataset of the members of the United States Congress and 2) a dataset based on ICPSR ID[6] accompanying Congress numbers[7], as a basis. It contains biographical records of 11 987 persons who served in the U.S. Congresses from the 1st (1789) to the 115th (2018) one. We converted and extracted the data (in CSV and YAML format[8]) into RDF, and developed a SPARQL compliant data service and an online application named *U.S. Congress Prosopographer*[9] to complement both quantitative and qualitative inquiry in American political history.

---

[4] https://github.com/unitedstates/congress-legislators

[5] http://k7moa.com

[6] The Inter-university Consortium for Political and Social Research (ICPSR) ID number

[7] https://www.senate.gov/reference/Years_to_Congress.htm

[8] http://yaml.org/spec/1.2/spec.html

[9] https://semanticcomputing.github.io/congress-legislators

2   Data Model and Linked Data Service

The target RDF data model for representing biographical records is based on the Norssi biography model [3]. The ontology model representing people and their biographical information is based on the schema.org vocabulary[10]. And the data model of schema.org is extended by additional properties and classes in the domain specific namespace[11].

All basic biographical data (family name, gender, given name, etc.) are modeled using the schema.org namespace, and all the data relating to his or her career are in the domain namespace. The resources can be linked to external databases or services, such as DBpedia, Wikidata, Wikipedia, and Twitter, for more information.

The data is available as a Linked Open Data service at the Linked Data Finland platform[12] in an open SPARQL endpoint[13] with resolvable URIs, using the W3C Linked Data publishing principles and best practices [1]. For example, the URI http://ldf.fi/congress/p10079 refers to Harry Truman (1882–1972), and can be used for retrieving the related RDF data or for Linked Data browsing depending on the need and HTTP protocol header data used. The data in the service contains altogether ca 830 000 triples, 790 000 in the people graph and 40 000 in the place graph.

3   Supporting Prosopographical Research
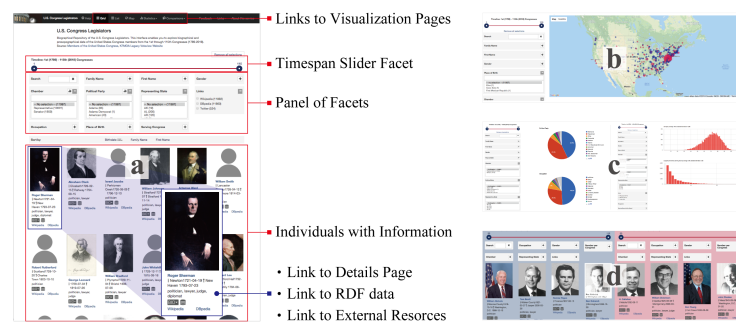
Fig. 1. Overview of the *U.S. Congress Prosopographer* (Main Four Tools: a,b,c,d)

As for prosopographical research, *U.S. Congress Prosopographer* provides a macroscopic viewpoint in historical time scale following with a microscopic viewpoint of individual Congress members (Fig. 1). The interface was implemented by extending SPARQL Faceter [2], a tool for creating faceted search interfaces on top of a SPARQL

---

[10]http://schema.org/docs/schemas.html

[11]http://ldf.fi/congress/

[12]http://ldf.fi

[13]http://ldf.fi/congress/sparql

endpoint. AngularJS[14] framework was used to organize linked data together with a timespan slider[15] that is included as a canonical facet. Other filtering facets include personal attributes, political characteristics, and links to external resources. As shown in Fig. 1, the interface contains the following four tools for prosopographical research:

a) Multifaceted Search Views The end user is able to customize the facets and explore the patterns, variations, and uniqueness in the Congress data. Increased accessibility is supported with intuitive interactive elements, such as simple click and drag.

b) Map visualization To understand the intellectual mobility of Congress members through the places of their birth and death, Angular Google Maps[16] is used in the visualization to locate, map, and explain historical trends in geographical space.

c) Two Views of Statistical Visualizations To examine the data through structured charts, and to provide glanceable overviews of the temporal features, this page generate statistics in Google Chart diagrams[17] based on the extracted filtering results.

d) Comparing Visualizations These visualizations allow the user to examine the similarities and differences between the Democratic and Republican parties. In these visualizations, all functions and visualizations used in a), b), and c) are implemented to identify and compare the properties of the two different target groups.

Besides this, every view enables investigation of individual legislator attributes in depth for biographical investigation.
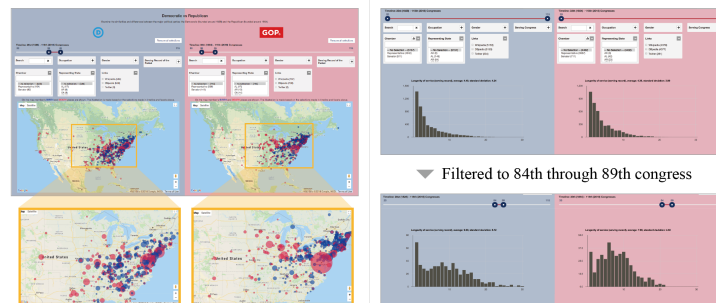
Fig. 2. LEFT: Mapped Birth and Death Places / RIGHT: Longevity of Service in Graph

Use Case Examples The comparison visualizations can be used in different research studies. For example, it can be shown that during the Reconstruction era from the 38th through the 45th Congresses (1863–64 to 1877–78) there is a large difference in the locations of birth and death of the legislators (cf. Fig. 2 LEFT). Most legislators were born and died in the eastern side. However, the distribution reveals a further clear tendency during this period: while the Democrats have a longitudinal spreading,

---

[14] http://angularjs.org

[15] https://github.com/angular-slider/angularjs-slider

[16] http://angular-ui.github.io/angular-google-maps/

[17] https://developers.google.com/chart/

Republicans remain in the Northeastern megalopolises. Another example is from the 84th through the 89th Congresses (1955–56 to 1965–66) when the federal government aimed to revitalize cities though funding urban renewal programs.[18] During this period of time, the poor were displaced and suffered from the series of policies. However, comparing with the overall trend in longevity of service of legislators which continuously decreases (cf. Fig. 2 RIGHT upper part), there is a wide variation in the longevity during 1955–66 (cf. Fig. 2 RIGHT bottom part). This indicates that incumbent re-election rates were extremely high in both parties during this time, despite the fact that the social situation was very unstable. Hence, through revealing such correlated continuities and changes, these examples demonstrate how historical patterns correspond to biographical information and further intertwine with politics, economics, and historical knowledge.

4   Related Work and Discussion

U.S. Congress Prosopographer explores different ways to support prosopographical research. Its different types of visualization for target groups establish context and maintain orientation while revealing details also about the individuals. This combination of macro- and microscopic viewpoints offers both qualitative and quantitative understanding of the biographical and prosopographical aspects of the Congress legislators.

The idea of combining faceted search and visualizations has been applied, e.g., in ePistlarium[19]. However, this tool deals with epistolary data and is not based on Linked Data. The research of this paper stands unique in providing a comprehensive coverage of U.S. Congressional biography from its beginning until today through the dynamic integration of querying and visualizing Linked Data under one single system.

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), http://linkeddatabook.com/editions/1.0/
2. Koho, M., Heino, E., Hyvönen, E.: SPARQL Faceter—Client-side Faceted Search Based on SPARQL. In: Troncy, R., Verborgh, R., Nixon, L., Kurz, T., Schlegel, K., Vander Sande, M. (eds.) Joint Proc. of the 4th International Workshop on Linked Media and the 3rd Developers Hackshop. CEUR Workshop Proceedings, Vol-1615 (2016), http://ceur-ws.org/Vol-1615/semdevPaper5.pdf
3. Leskinen, P., Tuominen, J., Heino, E., Hyvönen, E.: An ontology and data infrastructure for publishing and using biographical linked data. In: Proceedings of the Workshop on Humanities in the Semantic Web (WHiSe II). pp. 15–26. CEUR Workshop Proceedings, Vol-2014 (2017)
4. Roberts, B.: Biographical Research. Understanding social research, Open University Press (2002), https://books.google.fi/books?id=04ScQgAACAAJ
5. Verboven, K., Carlier, M., Dumolyn, J.: A short manual to the art of prosopography. In: Prosopography Approaches and Applications. A Handbook, pp. 35–70. University of Ghent (2007), http://hdl.handle.net/1854/LU-376535

---

[18]Widely known as "The Urban Renewal Projects"

[19]http://ckcc.huygens.knaw.nl/epistolarium/