

Linked Open Data Approach to Study the Assembly Minutes of International Organizations and Their Underlying Prosopography in the Real World Context

Eero Hyvönen^{2,1}, Petri Leskinen^{1,2}, Alexandre Lionnet^{3,1}, Blandine Blukacz-Louisfert⁴, Pierre-Etienne Bourneuf¹, Davide Rodogno¹, Grégoire Mallard¹, and Florian Cafiero^{3,1}

¹ Geneva Graduate Institute, Center for Digital Humanities and Multilateralism, Geneva, Switzerland

² Aalto University, Department of Computer Science, Semantic Computing Research Group (SeCo), Finland

³ École nationale des chartes, Paris, France

⁴ UN Library & Archives Geneva, Geneva, Switzerland

Abstract. The notion of democracy and the functions of the state of the law in modern liberal countries are based on legislation and agreements that are discussed and decided in democratically elected national parliaments. In an analogous way, various international organizations have been established to discuss and agree upon international matters, such as state of the law principles, war and peace, climate change, global health issues, and food and agriculture. Examples of such international organizations include Parliament of EU, the United Nations (UN), Inter-Parliamentary Union (IPU), World Health Organization (WHO), Red Cross, and Food and Agriculture Organization (FAO) to name a few. This paper presents a Linked Open Data (LOD) approach for publishing and studying assembly minutes data of international organizations, based on lessons learned in developing analogous systems for parliamentary discussion minutes. As a case study, minutes of the Assembly of the League of Nations (LoN) (1920–1946) are considered. A new system LEAGUE OF NATIONS SAMPO is presented, re-using the ParliamentSampo framework for the speeches of Parliament of Finland (1907–). LEAGUE OF NATIONS SAMPO is based 27 000 pages of minutes of LoN assembly meetings, a prosopographical knowledge graph about some 3100 people mentioned in the minutes, and on contextualizing data about the real world.

Keywords: knowledge graphs · digital humanities · information retrieval · data analysis

1 Studying national and cross-national organizations

To make decision making in parliaments and international organizations organizations transparent, minutes of discussions among the decision makers are

documented and made open to the public and researchers as books and as PDF and other documents online. This paper argues that opening the documents for people to read is not enough. Even if the documents are openly available, they are typically not Findable, Accessible, Interoperable, and Re-usable according to the FAIR principles. As a remedy, the minutes should be published not only in human-readable form but also as semantic data for computers to interpret and use, and including contextual data about the decision makers and organizations involved, as well as about the underlying real world. This would facilitate development of easy to use intelligent applications for public engagement and scholarly Digital Humanities (DH) research.

From a DH point of view, minutes data differ from organization to organization. In national parliaments, the minutes typically record the speeches of parliamentarians word by word. For example, the Parliament of Finland corpus [5] contains textual transcripts of over million speeches since 1907. On the other hand, in the assembly minutes of the League of Nations (1920–1946) [6], the focus is more on documenting the decisions. However, in all cases the minutes documents are essentially texts making references to various topical matters, people, organizations, and places in time. Our research hypotheses therefore is that a similar kind of technical solution can be applied to different kind of minutes cases for FAIRifying the data and for building applications on top of the data.

2 Related work

Parliaments have created speech corpora and datasets of both historical and contemporary parliaments. Parliamentary data has been used in many fields of research, such as linguistics, political science, legal studies, media studies, economics, and history, including long-term studies [3]. Semantic web technologies have been applied for linking and enriching parliamentary data [9]. In the same way international organizations have produced vast corpora of assembly minutes, plenary debates, committee reports, and voting records for accountability and global governance [2]. The vision in our work is to turn such dispersed materials into a sustainable LOD ecosystem [1, 5] that allows users to trace who spoke, on what issues, in which capacities, and how agendas evolved over time, while remaining explicit about uncertainties, biases, and curatorial choices. Our ultimate goal is to create and align with each other minutes data from several related international organizations, with a focus on Geneva-based ones.

3 Case study: LEAGUE OF NATIONS SAMPO

To test and demonstrate these ideas, we present the case of the League of Nations (LoN) based on data from the LONTAD project, [8, 10]. These materials are accessible online for humans to read, but not as data for research and application development.

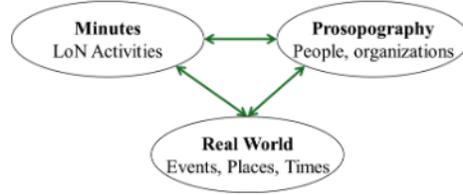


Fig. 1. Interlinked components of the LEAGUE OF NATIONS SAMPO knowledge graph.

Table 1. LONTAD minutes were enriched by person data from seven external sources. Column People tells the number of shared entries with the 3100 LONTAD people.

#	Data Service	People	Domain
1	Lonsea ^a	2653	Database about the history of the League of Nations
2	Wikidata	1682	Knowledge graph and data service underlying Wikipedias
3	VIAF.org	1440	Virtual International Authority File system combining other authority services of national libraries etc.
4	Wikipedia	1296	English or French Wikipedia
5	GND ^b	1181	Integrated authority file system of the the German National Library
6	Metagrid.ch ^c	385	Linking service designed to connect metadata and resources of research projects throughout the humanities and social sciences
7	DODIS ^d	369	Diplomatische Dokumente der Schweiz

^a Lonsea: <http://www.lonsea.de/>

^b GND: <https://www.dnb.de/EN/>

^c Metagrid API service: <https://metagrid.ch/docs/enhancer/>

^d Dodis: <https://www.dodis.ch/>

We transformed 27 000 pages of LoN Assembly minutes related to some 3100 mentioned representatives and other people into a LOD service and semantic portal LEAGUE OF NATIONS SAMPO. Proposographical data about the actors and the real world (cf. Fig. 1) were extracted from the related data services of Table 1.

The LEAGUE OF NATIONS SAMPO SPARQL endpoint can be used directly by scripting in DH research and for application development. To demonstrate this, LEAGUE OF NATIONS SAMPO portal was built by using the Sampo model ⁵ [4] and Sampo-UI framework [7]⁶ and the domain specific ParliamentSampo framework [5].

⁵ Sampo portal series: <https://seco.cs.aalto.fi/applications/sampo/>

⁶ Sampo-UI home: <https://seco.cs.aalto.fi/tools/sampo-ui/>; Github: <https://github.com/SemanticComputing/sampo-ui>

The idea of the “Sampo framework” is to take an existing Sampo in a domain of interest, with its ready to use user interface (UI) model and knowledge graph as a starting point to be modified. This approach allows for extremely rapid software development, if the framework used is mostly fit for the new purpose. See [6] for more details about the LEAGUE OF NATIONS SAMPO LOD service and portal application available openly on the Web.

Acknowledgments. We thank the LONTAD archives project, UN Library & Archives Geneva, Lonsea project, Dodis, and Metagrid for fruitful collaborations. Support of the Finnish DH research infrastructure initiative FIN-CLARIAH/DARIAH-FI funded by the Research Council of Finland and NextConnectionEU is acknowledged. Computational resources of CSC – IT Center for Science have been used.

Bibliography

- [1] Cafiero, F.: Datafying diplomacy: How to enable the computational analysis and support of international negotiations. *Journal of Computational Science* **71**, 102056 (2023). <https://doi.org/10.1016/j.jocs.2023.102056>
- [2] Cafiero, F., Cointet, J.P., Mallard, G.: Digital accountability can re-legitimate multilateralism (2025), working paper
- [3] Guldi, J.: Parliament’s debates about infrastructure: An exercise in using dynamic topic models to synthesize historical change. *Technology and Culture* **60**(1), 1–33 (2019). <https://doi.org/10.1353/tech.2019.0000>
- [4] Hyvönen, E.: Digital humanities on the semantic web: Sampo model and portal series. *Semantic Web* **14**(4), 729–744 (2023). <https://doi.org/10.3233/SW-223034>
- [5] Hyvönen, E., Sinikallio, L., Leskinen, P., Drobac, S., Leal, R., La Mela, M., Tuominen, J., Poikkimäki, H., Rantala, H.: Publishing and using parliamentary linked data on the semantic web: ParliamentSampo System for Parliament of Finland. *Semantic Web* **16**(1) (2025). <https://doi.org/10.3233/SW-243683>
- [6] Leskinen, P., Hyvönen, E., Cafiero, F., Lionnet, A., Blukacz-Louisfert4, B., Bourneuf, P.E., Rodogno1, D., Mallard, G.: A linked open data service and semantic portal to study the assembly minutes and prosopography of the league of nations (1920–1946 (December 2025), <https://seco.cs.aalto.fi/publications/2026/leskinen-et-al-lon-2026.pdf>
- [7] Rantala, H., Ahola, A., Ikkala, E., Hyvönen, E.: How to create easily a data analytic semantic portal on top of a SPARQL endpoint: introducing the configurable Sampo-UI framework. In: *VOILA! 2023 Visualization and Interaction for Ontologies, Linked Data and Knowledge Graphs 2023*. vol. 3508. *CEUR Workshop Proceedings* (2023), <https://ceur-ws.org/Vol-3508/paper3.pdf>
- [8] UN Library & Archives Geneva: LONTAD: Total digital access to the league of nations archives. <https://libraryresources.unog.ch/lontad> (2022), digitization project (2017–2022) providing comprehensive online access to the League of Nations archives
- [9] Van Aggelen, A., Hollink, L., Kemman, M., Kleppe, M., Beunders, H.: The debates of the European Parliament as Linked Open Data. *Semantic Web* **8**(2), 271–281 (2017). <https://doi.org/10.1007/s42001-019-00060-w>
- [10] Wells, C.M.: Total digital access to the league of nations archives: Digitization, digitalization, and analog concerns. In: *Archiving Conference*. vol. 16, pp. 12–16. *Society for Imaging Science and Technology* (2019)