Sampo-UI Meets Ontop: Publishing Relational Digital Humanities Data on the Semantic Web

Joren Six[1,*], Jahid Chetti[1], Annastiina Ahola[2], Frederic Lamsens[1], Heikki Rantala[2] and Eero Hyvönen[2]

[1]Ghent University, Ghent Centre for Digital Humanities (GhentCDH), https://www.ghentcdh.ugent.be
[2]Aalto University, Department of Computer Science, Semantic Computing Research Group (SeCo), https://seco.cs.aalto.fi

Abstract

The Sampo model and its Sampo-UI framework have become an established approach for publishing Cultural Heritage and Digital Humanities data on the Semantic Web. Sampo-UI enables exploration of data originating from SPARQL endpoints through faceted search across multiple views with seamlessly integrated visualization and data analytic tools. Here, we present two contributions to the Sampo-UI ecosystem: (i) a separation of Sampo-UI core from application specific configuration; decoupling core and application simplifies updates and facilitates reuse of efforts, and (ii) the use of Ontop — a virtual knowledge graph system — to expose relational databases through a SPARQL endpoint and Sampo-UI. The use of relational databases and reuse of tooling and expertise around relational databases lowers the operational burden of publishing small-to-medium-sized datasets typical of digital humanities projects. These contributions come from outside the original Sampo-UI team but were developed in collaboration with it, and are being prepared for upstream integration. We illustrate the approach with two projects at Ghent University: one on Roman coin finds in Belgium (NuMAD) and one on historic Belgian company data (BelHisFirm). Both portals run the same Sampo-UI core with a different configuration, and operate on existing institutional relational database infrastructure without a dedicated graph database.

Keywords

Linked data, SPARQL, Sampo-UI, Semantic Web, Virtual Knowledge Graph, Ontop-VKG

1. Introduction

Sampo-UI[1] [1, 2] is an open-source framework for creating web applications or "semantic portals" on top of SPARQL [3] endpoints following the principles of the Sampo Model [4]. The framework is built around *application perspectives* which determine how to filter – using facets – and view RDF [2] data. Views on the data can take the form of maps, timelines, networks, and tables for semantic exploration. In Finland, Sampo-UI has been used in more than twenty "Sampo portals" (see [5, 6, 7]). In recent years, it has also been used in multiple international projects [8, 9, 10, 11].

Operating Sampo-UI portals over the long term remains challenging in a project-funded research landscape, where infrastructure must survive well beyond active development. We argue that two architectural choices in the current Sampo-UI ecosystem—the entanglement of core framework and application-specific configuration, and the reliance on SPARQL-native graph databases—raise operational costs and skill requirements in ways that are increasingly untenable. This paper presents two contributions to the Sampo-UI system. The first contribution is an architectural change to split core and application specific code. The second is an operational change: presenting data originating from relational databases via a virtual knowledge graph. This avoids the need to run graph databases and the specialist knowledge needed to run these systems efficiently. To evaluate the feasibility of the changes,

Semantics'26: 22nd International Conference on Semantic Systems, September 15–17, 2026, Ghent, Belgium
*Corresponding author.
joren.six@ugent.be (J. Six); jahid.chetti@ugent.be (J. Chetti); annastiina.ahola@aalto.fi (A. Ahola); frederic.lamsens@ugent.be (F. Lamsens); heikki.rantala@aalto.fi (H. Rantala); eero.hyvonen@aalto.fi (E. Hyvönen)
https://ghentcdh.ugent.be/ (J. Six)
0000-0001-7671-1907 (J. Six); 0009-0008-6369-4712 (A. Ahola); 0000-0002-1527-5723 (F. Lamsens); 0000-0002-4716-6564 (H. Rantala); 0000-0003-1695-5840 (E. Hyvönen)

[1]Sampo-UI project homepage: https://seco.cs.aalto.fi/tools/sampo-ui/
[2]RDF of W3C: https://www.w3.org/RDF/

two new Sampo portals are presented: BelHisFirm, a project around Belgian company data and NuMAD, a project on historic coin finds.

2. Sampo-UI contributions

In close collaboration with the original Sampo-UI development team, we at Ghent University contributed to Sampo-UI [3] with the aim to increase the sustainability of the project in two ways. One deals with software architecture and the other is on the operational side.

1. Splitting Sampo-UI core and application-specific configuration The previous Sampo-UI design made it difficult to decouple application-specific configuration or components from the core Sampo-UI system. Until recently, starting a new Sampo-UI project meant forking the GitHub repository [4] and editing the fork to fit project-specific requirements. This makes maintenance difficult, since updating needs to be done in parallel versions.

The previous workflow also posed a hurdle to contribute reusable components, or other work that could be shared across Sampo projects. Thanks to a serious refactor effort, a strict separation between application and core system has been established. Now, a new project means using the Sampo-UI core system and configuring application specifics. A new plug-in system lets developers register custom application-specific view components (maps, timelines). Versioned releases of Sampo-UI core in the form of OCI container now also makes deployment and upgrades straightforward.

2. Exposing relational databases via a Virtual Knowledge Graph Current Sampo-UI projects are backed by SPARQL-native graph/RDF databases such as QLever [12], Jena [13], Virtuoso, Apache AGE, or similar systems. The expertise on how to run, optimize, back up, restore, migrate, and upgrade these systems is not always readily available and requires experience. The software and support ecosystem around graph databases is still relatively limited and of variable quality. For example, a centrally hosted Virtuoso instance at Ghent University lost university-wide support due to limited demand, and commercial RDF hosts such as TriplyDB are unaffordable for projects without long-term funding.

The operational costs for running relational databases are much lower. Low-cost, supported ways to run — and keep running — relational databases are readily available. Within universities and other contexts, there is typically more expertise and knowledge on relational databases. The software ecosystem around relational databases is richer; it simply has a longer history. Relational databases are, however, not directly compatible with Sampo-UI. A mapping layer is needed.

The Ontop project [14, 15] exposes data in a relational database through a SPARQL endpoint The relational database is mapped to a virtual knowledge graph. This makes a curated subset of the data typically locked in a database accessible in a standardized way and it is an ideal fit to combine with Sampo-UI. The Ontop setup preserves tooling around relational databases in place while having the advantages of data published on the semantic web.

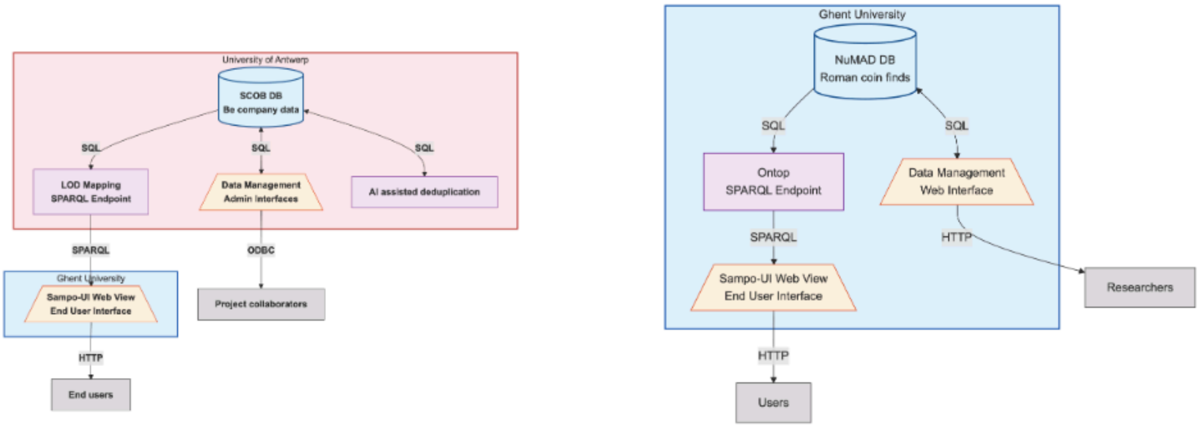
Digital humanities datasets are often of human-sized scale: several thousands of records are typical. For workloads with much larger datasets, without legacy tooling around relational databases or strict latency requirements, investing in a native graph database remains warranted.

3. BelHisFirm, two centuries of Belgian corporate history

BelHisFirm[11] is a large-scale, inter-university infrastructure project that converts two centuries of Belgian corporate records — all published in the Annexes of the Belgian Official Gazette since 1873 — into structured, firm-level linked data integrated with the already established database. The project combines high-resolution digitization, OCR and multi-level record linkage to reconstruct corporate trajectories across decades.

The database – an Oracle system hosted at Antwerp University – goes back many years and has custom tooling and ingrained workflows and expertise around it (see Figure 1). A new service around

[3] The GhentCDH Sampo-UI fork to be mainlined: https://github.com/GhentCDH/Sampo-ui
[4] Sampo-UI GitHub repository: https://github.com/SemanticComputing/sampo-ui



(a) (b)

Figure 1: Architectures of the two case studies. Both systems have tooling and workflows around a relational database. (a) BelHisFirm: Belgian historic company data exposed via an Ontop SPARQL endpoint and a Sampo-UI front-end, with additional tooling to manage and deduplicate data. (b) NuMAD: Roman coin finds database exposed via an Ontop SPARQL endpoint and a Sampo-UI front-end. The data is managed through a generic data management environment operating on the relational DB.

the relational database maps and publishes data according to the EURHISFIRM RDF schema using the Ontop Virtual Knowledge Graph system. The mapping step allows to configure which parts of the data becomes public via a SPARQL endpoint.

Sampo-UI is used to make the company data explorable [5]. With the Ontop system running at Antwerp University – close to the database – and Sampo-UI hosted at Ghent University, a caching layer helps to keep the Sampo-UI interface snappy. By using the Sampo-UI core with only application specific configuration, development time is limited to a minimum and involves mainly configuration. [6]

The architecture allows to keep existing tooling, workflows and operational knowledge around the database; it avoids the need to run a parallel graph database infrastructure. Additionally, it provides a data publication endpoint based on an established schema (EURHISFIRM) and - with Sampo-UI - the data is accessible in a user-friendly way without much custom development.

4. NuMAD, a Belgian coin find database

NuMAD is a project shared between the Belgian national library (KBR) and Ghent University around Roman coin finds in Belgium and their archaeological context. Over many decades, information on coin finds was collected in various databases, notebooks and archives. These are now concentrated in a single relational database. A web environment allows project contributors to add, update and maintain coin find data in a modern web interface. The public Sampo-UI platform [7] allows users to explore and map the finds, see Figure 2 for a faceted exploration of coin finds together with the mints that produced them, overlaid on a Roman-roads map layer[16].

To publish the data in a reusable form, the data is mapped to the Nomisma ontology[17] using the Ontop Virtual Knowledge Graph. The mapping step allows to configure which parts of the data becomes public. For example, the open endpoint does not include precise find locations: in Belgium, exact coordinates of coin finds are not made public to prevent enthusiasts from unintentionally disturbing archaeological sites. Again, by using the Sampo-UI core with only application specific configuration, development time is limited and involves mainly configuration[8]. This project has the additional advantage that it could build upon a similar platform for coin data: coinsampo[18].

[5]BelHisFirm portal: https://belhisfirm.be
[6]BelHisFirm source code: https://github.com/GhentCDH/BelHISFirm-Frontend
[7]NuMAD platform: https://numad.ugent.be
[8]NuMAD source code: https://github.com/GhentCDH/NuMAD

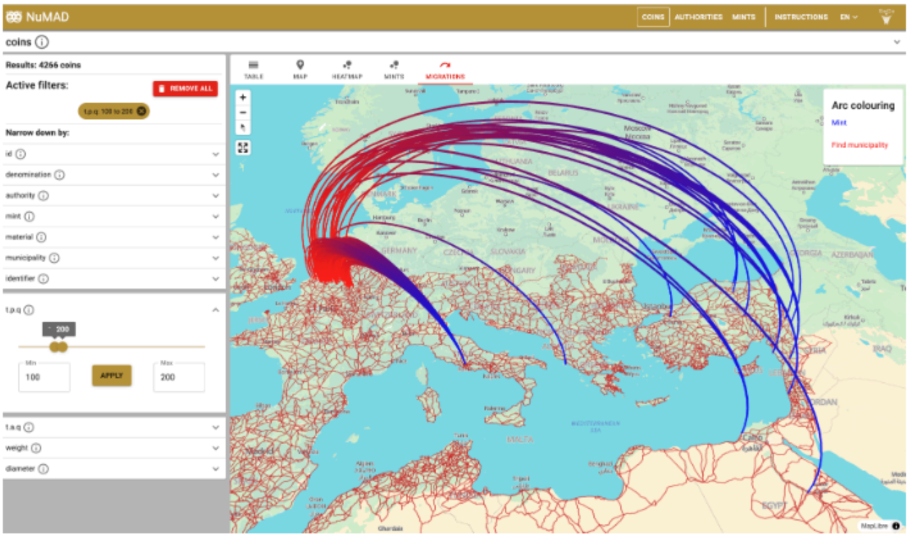


Figure 2: Links between mints and a filtered set of coin finds in Belgium. To situate finds, the map features a layer with Roman roads[16].

Similar to the previous project, the architecture (see Figure 1) facilitates the (re)use of generic tooling to manage data in relational databases. Additionally, it provides a data publication endpoint based on an established schema (Nomisma) and with Sampo-UI the data is accessible in a user-friendly way with only limited custom development efforts.

5. Conclusion

We presented two contributions to Sampo-UI aimed at lowering the operational cost of running semantic portals in project-funded digital humanities settings: a decoupling of core framework and application configuration, and an Ontop-based path from existing relational databases to a Sampo-UI portal. The BelHisFirm and NuMAD portals show that both contributions are practical: each runs the same Sampo-UI core with only configuration-level changes, on top of pre-existing institutional Oracle and PostgreSQL infrastructure. Both contributions are being prepared for upstream integration with the original Sampo-UI team. We expect this combination to lower the barrier for digital humanities projects that already maintain relational data and lack the resources to operate a dedicated graph database.

Acknowledgments The Ghent University acknowledges the CLARIAH-VL+ FWO funding (2025–2028; Project I001525N), the BelHisFirm (2024–2028; Project I015424N), and the funding by BOF UGent (Project NuMAD) by Koenraad Verboven to fund work at GhentCDH. Thanks to the Ghent University IT-services for computational resources such as container orchestrators, relational databases and virtual machines. Support of the Finnish Digital Humanities 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.

The authors used Claude Opus 4.7 in order to: check spelling, verify grammar, and suggest synonyms to improve text flow. The authors have reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] E. Ikkala, J. Tuominen, J. Raunamaa, T. Aalto, T. Ainiala, H. Uusitalo, E. Hyvönen, NameSampo: A linked open data infrastructure and workbench for toponomastic research, in: Proceedings of the 2nd ACM SIGSPATIAL Workshop on Geospatial Humanities, GeoHumanities'18, ACM, New York, NY, USA, 2018, pp. 2:1–2:9. doi:10.1145/3282933.3282936.

[2] H. Rantala, A. Ahola, E. Ikkala, E. Hyvönen, How to create easily a data analytic semantic portal on top of a SPARQL endpoint: introducing the configurable Sampo-UI framework, in: Proceedings of 8th International Workshop on the Visualization and Interaction for Ontologies and Linked Data co-located with the 22nd International Semantic Web Conference (ISWC 2023) in Athens, Greece, CEUR Workshop Proceedings, Vol. 3508, 2023. URL: https://ceur-ws.org/Vol-3508/paper3.pdf.

[3] S. Harris, A. Seaborne, SPARQL 1.1 Query Language, W3C Recommendation, World Wide Web Consortium (W3C), 2013. https://www.w3.org/TR/sparql11-query/ [accessed 15.10.2025].

[4] E. Hyvönen, Digital Humanities on the Semantic Web: Sampo Model and Portal Series, Semantic Web – Interoperability, Usability, Applicability 14 (2023) 729–744. doi:10.3233/SW-223034.

[5] E. Hyvönen, L. Sinikallio, P. Leskinen, S. Drobac, R. Leal, M. L. Mela, J. Tuominen, H. Poikkimäki, H. Rantala, Publishing and Using Parliamentary Linked Data on the Semantic Web: ParliamentSampo System for Parliament of Finland, Semantic Web 16 (2025). doi:10.3233/SW-243683.

[6] E. Hyvönen, P. Leskinen, H. Poikkimäki, H. Rantala, R. Leal, J. Tuominen, S. Drobac, O. Koho, I. Pikkanen, H.-L. Paloposki, Searching, Exploring, and Analyzing Historical Letters and the Underlying Networks: LetterSampo Finland – Finnish 19th-Century Letters on the Semantic Web, DH in the Nordic and Baltic Countries Publications 7 (2026). doi:10.5617/dhnbpub.12932.

[7] A. Ahola, E. Hyvönen, H. Rantala, A. Kauppala, Historical Opera and Music Theatre Performances on the Semantic Web: OperaSampo 1830–1960, in: Knowledge Graphs in the Age of Language Models and Neuro-Symbolic AI. Proceedings of the 20th International Conference on Semantic Systems, Amsterdam, The Netherlands, IOS Press, 2024, pp. 386–402. doi:10.3233/SSW240031.

[8] N. K. Grislis, K. Čerāns, M. Grasmanis, H. Rantala, E. Hyvönen, How to Add a User Interface on Top of an External SPARQL Endpoint: Case Nobel Prize Sampo, in: SEMANTiCS-PDWT 2025, Posters, Demos, Workshops, and Tutorials at SEMANTiCS 2025, volume 4064, CEUR Workshop Proceeding, 2025. URL: https://ceur-ws.org/Vol-4064/PD-paper17.pdf.

[9] F. Pannach, L. Cheng, F. Pianzola, The GOLEM-Knowledge Graph and Search Interface: Perspectives into Narrative and Fiction, in: Proceedings of the Computational Humanities Research Conference 2024, CEUR WS Proceedings, Vol. 3834, 2024. URL: https://ceur-ws.org/Vol-3834/paper80.pdf.

[10] P. Leskinen, E. Hyvönen, F. Cafiero, A. Lionnet, B. Blukacz-Louisfert, P. E. Bourneuf, D. Rodogno1, G. Mallard, A Linked Open Data Service and Semantic Portal to Study the Assembly Minutes and Prosopography of the League of Nations (1920–1946), in: Proceedings of ESWC 2026, Springer-Verlag, 2026. URL: https://seco.cs.aalto.fi/publications/2026/leskinen-et-al-lon-2026.pdf, in press.

[11] Vercruysse, Bas and Ducatteeuw, Vincent and Poukens, Johan and Birkholz, Julie M. and Verbruggen, Christophe, BELHISFIRM : Belgian company records as linked data, in: eabh Annual Meeting 2026, Abstracts, The European Association for Banking and Financial History (eabh), 2026. URL: https://bankinghistory.org/events/datafication-interoperability-in-historical-archives/.

[12] H. Bast, B. Buchhold, Qlever: A ery engine for efficient sparql+ text search (2017).

[13] B. McBride, Jena: A semantic web toolkit, IEEE Internet computing 6 (2002) 55–59.

[14] T. Bagosi, D. Calvanese, J. Hardi, S. Komla-Ebri, D. Lanti, M. Rezk, M. Rodríguez-Muro, M. Slusnys, G. Xiao, The Ontop framework for ontology based data access, in: D. Zhao, J. Du, H. Wang, P. Wang, D. Ji, J. Z. Pan (Eds.), The Semantic Web and Web Science, Springer, 2014, pp. 67–77.

[15] R. E. Wandji, D. Calvanese, Realizing bidirectional virtual knowledge graphs using ontop, in: H. Takeda, Y. Tzitzikas, G. Flouris, S. He, D. Plexousakis, S. Ferré, R. Yu, V. Efthymiou, E. Ilkou, E. Jiménez-Ruiz, X. Wang (Eds.), Knowledge Graphs, Springer-Verlag, 2026, pp. 220–235.

[16] P. De Soto, A. Pažout, T. Brughmans, P. B. Vahlstrup, Á. Auir, T. Bongers, J. E. B. Christoffersen, M. Crépy, M. H. Johansen, J. Lewis, et al., Itiner-e: A high-resolution dataset of roads of the roman empire, Scientific data 12 (2025) 1731.

[17] E. Gruber, S. Heath, A. Meadows, D. Pett, K. Tolle, D. Wigg-Wolf, Semantic web technologies applied to numismatic collections, CAA2012: Archaeology in the Digital Era II., Amsterdam University Press, Amsterdam (2013) 264–274.

[18] H. Rantala, E. Oksanen, F. Ehrnsten, E. Hyvönen, Publishing numismatic public finds on the semantic web for digital humanities research–coinsampo linked open data service and semantic portal, in: CEUR Workshop Proceedings, volume 3724, CEUR, 2024.