PM-Sampo: Semantic Portal for Heritage Object Provenance Research

Sarah Binta Alam Shoilee[1][0000−0001−9458−8105], Annastiina Ahola[2][0009−0008−6369−4712], Heikki Rantala[2][0000−0002−4716−6564], Eero Hyvönen[2,3][0000−0003−1695−5840], Victor de Boer[1][0000−0001−9079−039X], Jacco van Ossenbruggen[1][0000−0002−7748−4715], and Susan Legene[1][0000−0002−2826−9541]

[1] Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, the Netherlands
{s.b.a.shoilee,v.de.boer,jacco.van.ossenbruggen,s.legene}@vu.nl
[2] Aalto University, Semantic Computing Research Group (SeCo), Konemiehentie 2, 02150 Espoo, Finland {annastiina.ahola, heikki.rantala, eero.hyvonen}@aalto.fi
[3] University of Helsinki, Helsinki Centre for Digital Humanities (HELDIG), Unioninkatu 40, 00170 Helsinki, Finland

Abstract. Provenance research in cultural heritage is a complex domain requiring the integration of diverse datasets, intricate historical narratives, and evolving research methodologies. Traditional manual approaches to tracking object ownership, acquisition, and transfer are often time-intensive and fragmented. To bridge this gap, we present a Linked Data-driven approach using PM-Sampo, a semantic portal for provenance research, that effectively meets complex domain requirements through harmonising provenance metadata, data quality assessment, and visualisation. This poster will showcase how domain-specific provenance challenges can be addressed through existing semantic web technologies, paving the way for knowledge discovery and knowledge inference.

Keywords: Semantic Web · Knowledge Discovery · Provenance Research · Cultural Heritage.

1 Motivation and Background

The increasing digitisation of museum collections has opened new possibilities for provenance research. However, provenance records are often fragmented and textual, making analysis challenging. Transforming such records into Linked Open Data (LOD) enables a large-scale analysis of provenance networks, trends, and institutional collaboration [5,6]. Furthermore, structured provenance data supports greater transparency, accountability, and inclusivity in museum research, aligning with broader efforts in restitution and decolonisation.

Our previous work [7] highlights three key requirements for effective heritage object provenance representation: (1) machine-readable, event-centric models for accessibility, (2) structured metadata, including object details, actor networks,

acquisition routes, and historical context, and (3) multiple perspective provenance representation for transparency. Although the paper shows that existing semantic web technologies and domain ontologies can meet these requirements for interoperable knowledge representation, communicating structured provenance information effectively to support provenance research and new knowledge discovery remains a challenge.

This work presents PM-Sampo, a lightweight and easy-to-implement system designed for provenance research, that enables interactive visualisation and dynamic exploration of entity relationships. It demonstrates how existing semantic web technologies support scalable, data-driven provenance exploration and can be easily adopted by cultural heritage institutions. We showcase real-world case studies that highlight how semantic interoperability, visualisation, and knowledge discovery techniques have the potential to improve provenance research.

2 PM-Sampo

PM-Sampo builds on the Sampo model [2], a Linked Data approach for cultural heritage analysis, where data services are available via an open SPARQL endpoint. The Sampo-UI framework [3,4] complements this model by providing a "stadardized" UI development framework for faceted search and visualisation. PM-Sampo extends existing Sampo portals by modifying configurations declaratively to align with the specific requirements and data models of the new application [4]. It supports CIDOC-CRM ontologies [1] and Linked Art standards, ensuring compatibility with provenance data models [7]. To optimise query performance for semantic portal, a data conversion pipeline was built to simplify ontology paths, accessible at `https://github.com/Shoilee/PM-SampoDataManager`, with data published through SPARQL endpoint at `http://ldf.fi/pm-sampo/sparql`.

The primary objective of PM-Sampo is to enable users to dynamically filter, explore, and analyse dynamic relationships among provenance-related entities. Beyond merely providing structured provenance data to ensure interoperability and reusability, the system is designed to facilitate knowledge discovery and comparative analysis. On the PM-Sampo homepage, users can select *target entities* through four distinct tabs: Objects, Provenance Events, Historical Events, and Actors. Once a target entity is chosen and landed on the result page of all instances of target entities, users can further filter instances of the selected entity class(es) based on attributes or related entity instances (*facet entities*) using faceted semantic search.

To provide insights of these filtered dataset, several data-analytic tools have been integrated namely faceted search tables, summarisation pie charts, production places maps, provenance tabs, and event timelines. The *related tab* infers new relationships beyond direct links, introducing actor-to-actor and actor-to-historical event connections via SPARQL queries, enhancing context-aware provenance analysis. The mapping between competency questions from prior work [7] and PM-Sampo's visual solutions is detailed in Table 2. The PM-Sampo

Entities	Competency Questions	Visual Interface	Knowledge Inference
place ⇒ object	Which objects were collected from a given place of origin?	Map visualisation of place and objects.	Identify geographical patterns in object collections.
time ⇒ object	Which objects were collected during a given year?	Timeline with objects and collection dates.	Discover temporal trends in object collections.
actor ⇒ object	Which objects were collected by person A?	Faceted search result table.	Trace the collection history of a specific individual.
event ⇒ object	Which objects were collected during a given historical event?	Event-object network graph.	Understand how historical events shaped object collections.
place, time ⇒ object	What objects are collected from a specific place of origin and produced in a specific time period?	Object timeline with geographical overlays.	Analyse patterns between place, time, and object production.
actor, time ⇒ object	Which objects are collected by person A during a given time?	Timeline and object-person connections.	Explore an individual's collection activity over time.
event ⇒ time	What are the common acquisition times attributed to a historical event?	Bar chart showing acquisition-time and object number correlations filtered by event.	Identify common acquisition time related to historical events.
event ⇒ place	What are the common places of origin for objects attributed to a given historical event?	Heatmap showing object-place of origin correlations filtered by event.	Understand geographical patterns tied to historical events.
event ⇒ actor	Which constituents are related to objects attributed to a historical event?	Instance tab of person with roles in event.	Discover relationships between actor and events in provenance.
object ⇒ actor	Which persons were involved in the provenance of this object?	Object instance provenance tab.	Identify key individuals in the object's provenance.
actor ⇒ event	Which historical events is this person attributed to and with which role?	Event tab in Person instance page with roles.	Trace a person's involvement in historical events.
actor ⇒ actor	Which persons are related through object acquisition?	Person-to-person relationship graph.	Identify interactions between individuals through object collection.
place ⇒ actor	What are the places this person has collected objects from?	Map of object collection locations by person.	Identify geographical patterns in a person's acquisitions.
place, time ⇒ objects	Is there any trend in objects collected from a place over time?	Animated map with objects and time overlays.	Discover trends in object collection related to time and location.

Table 2. Mapping of provenance research requirements to visual interfaces and possible knowledge discovery. This table outlines domain interest related to objects, places, actor, time, and events, along with adapted visualisation interface by PM-Sampo that facilitate intuitive exploration and knowledge discovery in provenance studies.

demonstrator is available at https://pmsampo.demo.seco.cs.aalto.fi/en/, with its source code published at ms demonstrator is available at https://pmsampo.demo.seco.cs.aalto.fi/en/, with its source code published at https://github.com/Shoilee/PM-Sampo/releases/tag/v1.0.

3 Conclusion

This paper demonstrates that provenance research can be conducted effectively using the Linked Data principles. By mapping domain requirements to visualisation techniques, we show that semantic web technologies are capable of addressing complex provenance research challenges. Future work includes integrating data from multiple sources, enabling entity linking support among data sources, and improving knowledge inference through pattern discovery.

Overall, PM-Sampo is designed for easy adoption by cultural heritage institutions, such as Getty Provenance Research and the Rijksmuseum, supporting a more comprehensive and transparent approach to provenance research. The demonstration will showcase real-world case studies, illustrating how semantic interoperability, visualisation tools, and knowledge discovery techniques enhance provenance research. Our approach underscores the potential of lightweight, domain-adaptable Semantic Web solutions in addressing the functional needs of provenance research and beyond.

References

1. Doerr, M.: The cidoc conceptual reference module: an ontological approach to semantic interoperability of metadata. AI magazine **24**(3), 75–75 (2003)
2. Hyvönen, E.: Digital humanities on the Semantic Web: Sampo model and portal series. Semantic Web journal **14**(4), 729–744 (2022). https://doi.org/10.3233/SW-223034
3. Ikkala, E., Hyvönen, E., Rantala, H., Koho, M.: Sampo-UI: A full stack JavaScript framework for developing semantic portal user interfaces. Semantic Web journal **13**(1), 69–84 (2022)
4. 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: 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), https://ceur-ws.org/Vol-3508/paper3.pdf
5. Rother, L., Koss, M., Mariani, F.: Taking care of history: Toward a politics of provenance linked open data in museums. Perspectives on data (2022)
6. Rother, L., Mariani, F., Koss, M.: Hidden value: Provenance as a source for economic and social history. Jahrbuch für Wirtschaftsgeschichte/Economic History Yearbook **64**(1), 111–142 (2023)
7. Shoilee, S.B.A., de Boer, V., van Ossenbruggen, J.: Polyvocal knowledge modelling for ethnographic heritage object provenance. In: Knowledge Graphs: Semantics, Machine Learning, and Languages, vol. 56, pp. 127–143. IOS Press, Leipzig, Germany (2023)