ArtSampo – Finnish Art on the Semantic Web

Annastiina Ahola[1][0009−0008−6369−4712], Heikki Rantala[1][0000−0002−4716−6564], and Eero Hyvönen[1,2][0000−0003−1695−5840]

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

Abstract. This paper presents first results of ArtSampo, a collaborative Finnish Linked Open Data (LOD) infrastructure for publishing fine art collections on the Semantic Web and for facilitating Digital Humanities (DH) research. The infrastructure consists of a Knowledge Graph (KG) whose initial version was compiled from the metadata of the three art museums of the Finnish National Gallery. A semantic ArtSampo portal was built on top of the KG for searching, browsing, and analyzing the underlying data. The Finnish ontology infrastructure and international datasets are used for harmonizing and enriching the data.

Keywords: digital humanities · fine art · cultural heritage · portals

1 Introduction

Art is an important part of cultural identity. Contrary to the fears that web services showcasing information on museum collections would cause a decrease in museum visitors, these kind of services have instead become an important tool for marketing and distributing information for museums, and the number of visitors has been growing, e.g., in Finland. These collections, however, have been divided between both public museums of various types as well as private collections. This makes it hard for the visitor to get a general picture of national collections and access a particular art object at a suitable time and place.

Though not an comprehensive replacement to physically going to view an art object, a web service containing information on the collections could alleviate these problems and enrich the physical visit experience by providing additional insightful information that can be accessed without the limitations based on time and place. Another use case for web services would be for the curators of these art collections to be able to get a good overview on what their collections consist of. Various ontologies and vocabularies already exist for describing art, such as the Getty Research Institute's[3] Art & Architecture Thesaurus (AAT), Universal List of Artist Names (ULAN) and Getty Iconography Authority, as well as the ICONCLASS [1] iconography classification system and Chinese Iconography Thesaurus CIT[4] for describing the contents of art works. Linked Data (LD) on

[3] www.getty.edu/research/tools/vocabularies/
[4] https://chineseiconography.org/

art has also been published as part of projects like Europeana[5] as well as the Linked Art [2] project involving various high-profile international museums.

ArtSampo is a LOD service and semantic portal for Finnish art collections. It facilitates an easy way of searching, browsing, and analyzing fine art data for both the general public as well as researchers. The idea is to combine collection data from different museums into one KG and utilize Semantic Web technologies to enhance the users' experience and to provide means for studying art. In this work, collection data is transformed into RDF and made available as a LOD service for DH research. A user interface (UI) utilizing faceted search [6] and offering integrated data-analytic tools is also built on top of the data to make it accessible and explorable without SPARQL knowledge. ArtSampo is one of the Sampo systems[6] [3] for publishing cultural heritage data on the Semantic Web.

2 ArtSampo Knowledge Graph and User Interface

The current version of the ArtSampo KG is based on the metadata on the collections of Finnish National Gallery provided as open data[7]. These collections include over 58,000 art objects and 6,400 artists and span multiple different museums (Ateneum Art Museum, Museum of Contemporary Art Kiasma and Sinebrychoff Art Museum). This data was transformed from the original JSON files to Turtle-serialized RDF data. The transformation was done using simple, custom Python scripts with the rdflib[8] library that read the original data and created the necessary triples based on the JSON data. As the data for the different museums came all from the same source, the same script could be used for all the data without need for alignment between the data of different collections. Basic data type casting for relevant values (e.g., dimensions as decimals) and language tagging labeling was done during the process, but the values itself were not touched.

The initial version of the KG consists of approximately 1,100,000 triples. The triples in the KG are stored on a Apache Jena Fuseki[9] SPARQL server on the Linked Data Finland platform[10] and is accessible from a restricted, read-only SPARQL endpoint. The KG uses three classes for modeling the data: *Art Object*, *Person* and *Multimedia*. All types of art objects are instances of the *Art Object* class. People are modeled with the *Person* class and linked to art objects via the art object instance. Images of art objects are modeled through the *Multimedia* class and linked to the art objects similarly to the artists. All other metadata related to the aforementioned classes is modeled through properties getting literal values. This data model structure is not intended to be the final structure, but rather just an initial version to get a better idea of what can be done with the

[5] https://www.europeana.eu/
[6] https://seco.cs.aalto.fi/applications/sampo/
[7] https://www.kansallisgalleria.fi/en/api-sovelluskehittajille
[8] https://rdflib.readthedocs.io/en/stable/
[9] https://jena.apache.org/documentation/fuseki2/
[10] https://ldf.fi/

data, and the final data model design and usage of existing vocabularies will be iterated upon as further data sources and their interoperability with the original data set is explored in more detail.

The user interface (UI) for the ArtSampo KG is built with the Sampo-UI framework [5] to facilitate an easy way of browsing, exploring, and searching the data as well as to provide integrated data-analytic tools for analyzing the data without the need for technical expertise. The UI queries the data from the restricted SPARQL endpoint. The landing page of the portal lists the *application perspectives* available in the portal. The portal is split into two different perspectives: *art objects* and *persons*: By choosing the *art objects* perspective the user is presented the data as rows of art objects, while the *persons* perspective lists all the people related to various art objects in the data.

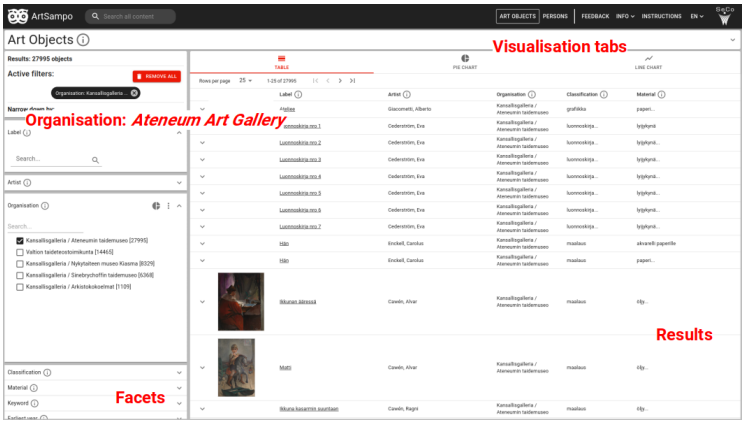

Fig. 1. The faceted search view of the portal with *Ateneum Art Museum* selected in the *Organisation* facet

The user can filter the search results in the faceted search view by using the provided facets to select the wanted values for various properties. For example, in Fig. 1 the use has selected *Ateneum Art Museum* as the wanted organisation in the *Organization* facet on the left, and all results are thus from the collections of the Ateneum Art Museum. The search results are listed as a table on the right and any of them can be selected for close reading and browsing for further information. With the provided data-analytic tools available as tabs the user can also easily visualize the result set as a whole. For example, Fig. 2 visualizes the most common art object materials for the selected subset of data (i.e., Ateneum Art Museum's collections) as a pie chart, while Fig. 3 visualizes a temporal aspect of the data, the years the creation processes of works have been started, on a timeline.

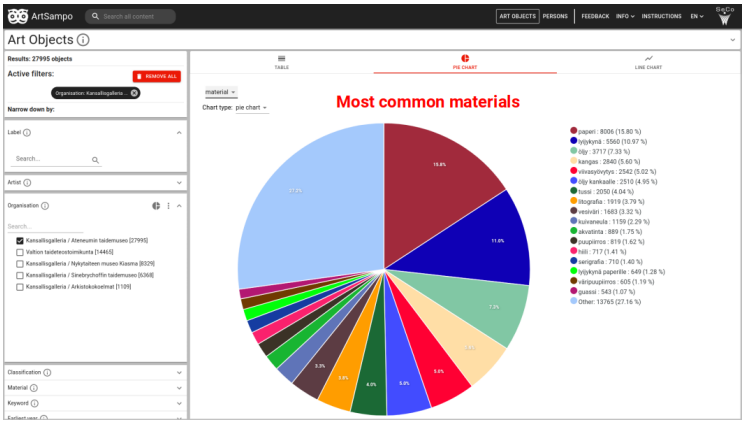

Fig. 2. A pie chart visualization showcasing the most common materials for works belonging to the Ateneum Art Museum's collections

3 Discussion & Future Work

The ArtSampo adds Finnish art and art history to the growing number of Finnish Cultural Heritage (CH) "Sampo" KGs and portals. Our plan is to expand the current data included in the KG by integrating openly available data from other Finnish museums and galleries as well as from international collections featuring Finnish art. With the new UI for faceted searching, browsing, and analyzing the data, the collection contents are more accessible to the common art museum visitor, DH researcher, and museum curator in comparison to publishing the data dump and using the current legacy search system online[11].

An integral part of the ArtSampo project is to connect it to other data sources, such as other Sampos, Wikidata and ULAN, to contextually enrich the existing data. For example, biographies about 1,900 of the 6,400 artists can be found in other Sampo KGs, such as BiographySampo [4]. In CoCoSampo[12] [7] , there is a letter correspondence KG of more than 10,000 letters of Finnish artists from the archives of the Finnish Art Society and the Finnish National Gallery. Outside links to, e.g., Wikidata can be extracted using person instances from Sampo KGs where these links are already available.

Another aspect of the project is to look into automatic content description by analyzing the available images and generating iconographic descriptions of those to supplement the possibly already-existing keywords that have been used for those particular art objects. The goal is to release the finalized KG from the project as LOD for use in research on the Linked Data Finland[13] platform.

[11] https://www.kansallisgalleria.fi/en/search
[12] https://seco.cs.aalto.fi/projects/coco/
[13] https://www.ldf.fi/

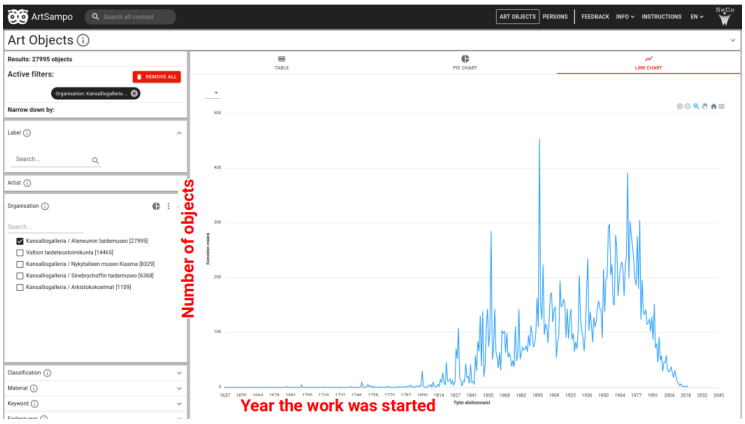

Fig. 3. A line chart visualization showcasing the distribution of years when the works belonging to Ateneum Art Museum's collections have been started by the artists

Acknowledgements This work has received partial funding from the European Union – NextGenerationEU instrument and is funded by the Research Council of Finland under grant number P3C3I6. Computing resources provided by the CSC – IT Center for Science were used in our work.

References

1. Couprie, L.D.: Iconclass: an iconographic classification system. Art libraries journal **8**(2), 32–49 (1983)
2. Delmas-Glass, E., Sanderson, R.: Fostering a community of pharos scholars through the adoption of open standards. Art Libraries Journal **45**(1), 19–23 (2020)
3. Hyvönen, E.: Digital humanities on the semantic web: Sampo model and portal series. Semantic Web **14**(4), 729–744 (2023)
4. Hyvönen, E., Leskinen, P., Tamper, M., Rantala, H., Ikkala, E., Tuominen, J., Keravuori, K.: BiographySampo – Publishing and enriching biographies on the Semantic Web for digital humanities research. In: Proceedings of the 16th Extended Semantic Web Conference (ESWC 2019). pp. 574–589. Springer (2019)
5. 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). https://doi.org/10.3233/SW-210428
6. Tunkelang, D.: Faceted search. Morgan & Claypool Publishers, CA, USA (2009)
7. Tuominen, J., Koho, M., Pikkanen, I., Drobac, S., Enqvist, J., Hyvönen, E., La Mela, M., Leskinen, P., Paloposki, H.L., Rantala, H.: Constellations of correspondence: a linked data service and portal for studying large and small networks of epistolary exchange in the grand duchy of finland. In: CEUR Workshop Proceedings. vol. 3232, pp. 415–423. RWTH Aachen University (2022)