Citizen Science Archaeological Finds on the Semantic Web: The FindSampo Framework

Eero Hyvönen, Heikki Rantala, Esko Ikkala, Mikko Koho, Jouni Tuominen, Babatunde Anafi,
Aalto University and University of Helsinki (HELDIG), http://seco.cs.aalto.fi
Suzie Thomas, Anna Wessman, Eljas Oksanen
University of Helsinki (Dept. of Cultures)
VIlle Rohiola, Jutta Kuitunen, Minna Ryyppö
Finnish Heritage Agency (FHA)

Abstract: FindSampo fosters collecting, sharing, publishing, and studying archaeological finds discovered by the public. The framework includes a mobile find reporting system, a semantic portal for researchers, the public, and collection managers to use, and a Linked Open Data service for creating custom data analyses and for application developers.

Keywords: Citizen Science; Metal Detecting; Archaeology; Linked Open Data; Semantic Web; Finland

Introduction

FindSampo develops a prototype framework system for supporting mobile finds data reporting on the field, and for studying archaeological artefacts discovered and reported by the public. It is unique in responding to the archaeological conditions in Finland, and in providing solutions to its users' needs (Wessman et al. 2019). However, the framework and its implementation are open source and can be re-used. FindSampo sits within a broader context of digitizing Finnish heritage (Hyvönen 2020), of the European Public Finds Recording Network[1], and the ARIADNEPlus[2] infrastructure project.

Challenges

Metal-detectorists' finds can contribute to archaeological research. However, in Finland it has been laborious to access data regarding new metal-detected finds, especially from a researcher perspective, and there is a backlog in the cataloguing process at FHA preventing up-to-date research. Hence, a user-friendly, easy-to-use tool for reporting, viewing, browsing, and researching metal-detected finds to get high quality metadata in a timely manner is needed. We adopted a 'citizen science' approach, conducting surveys, interviews, and focus groups for future users to express their preferences.

FindSampo Framework in Action

FindSampo's framework consists of three major components:

[1] https://www.helsinki.fi/en/networks/european-public-finds-recording-network
[2] https://ariadne-infrastructure.eu

- *FindSampo Reporter*: a mobile app demonstrator to minimise the reporting work in the field, foster collaborative work with FHA, and create a Linked Open Data repository to be used in applications.
- *FindSampo Portal*: a forthcoming research prototype of a semantic portal for studying the finds online.
- *FindSampo Data Service*: based on the Linked Data (LD) approach, standards, and best practices of W3C (Heath and Bizer 2011), opens the finds data to application and data analysis developers to use with modest programming skills.

FindSampo is based on the “Sampo model” (Hyvönen 2020) using the "FAIR guiding principles for scientific data management and stewardship"[3]. This model includes three components:
1) A “business model” for harmonizing, aggregating, and publishing heterogeneous, distributed data from different content providers based on a shared ontology infrastructure.
2) An approach to interface design, where the data can be re-used and accessed independently from multiple application perspectives, while the data resides in a single SPARQL endpoint.
3) A two-step model for accessing and analyzing the data, where the focus of interest is first filtered out using faceted semantic search, and then visualized and analyzed by ready-to-use Digital Humanities tools of the portal. Implementing user interfaces based on this model is supported by the open source Sampo-UI framework[4].

In FindSampo, archaeological finds can be searched using the faceted search paradigm (Tunkelang 2015), allowing to narrow the result set by making orthogonal category value selections, such as object type, material, time period, and place, based on underlying ontologies. Once a result set of interest has been found, ready-to-use data analytic tools and visualizations can be applied to it with additional contextual information. For example, it is possible to visualize on maps at the same time as seeing protected archaeological sites. If the question is about an individual find, its “home page” can be studied further.

[3] https://www.go-fair.org/fair-principles/
[4] https://seco.cs.aalto.fi/tools/sampo-ui/

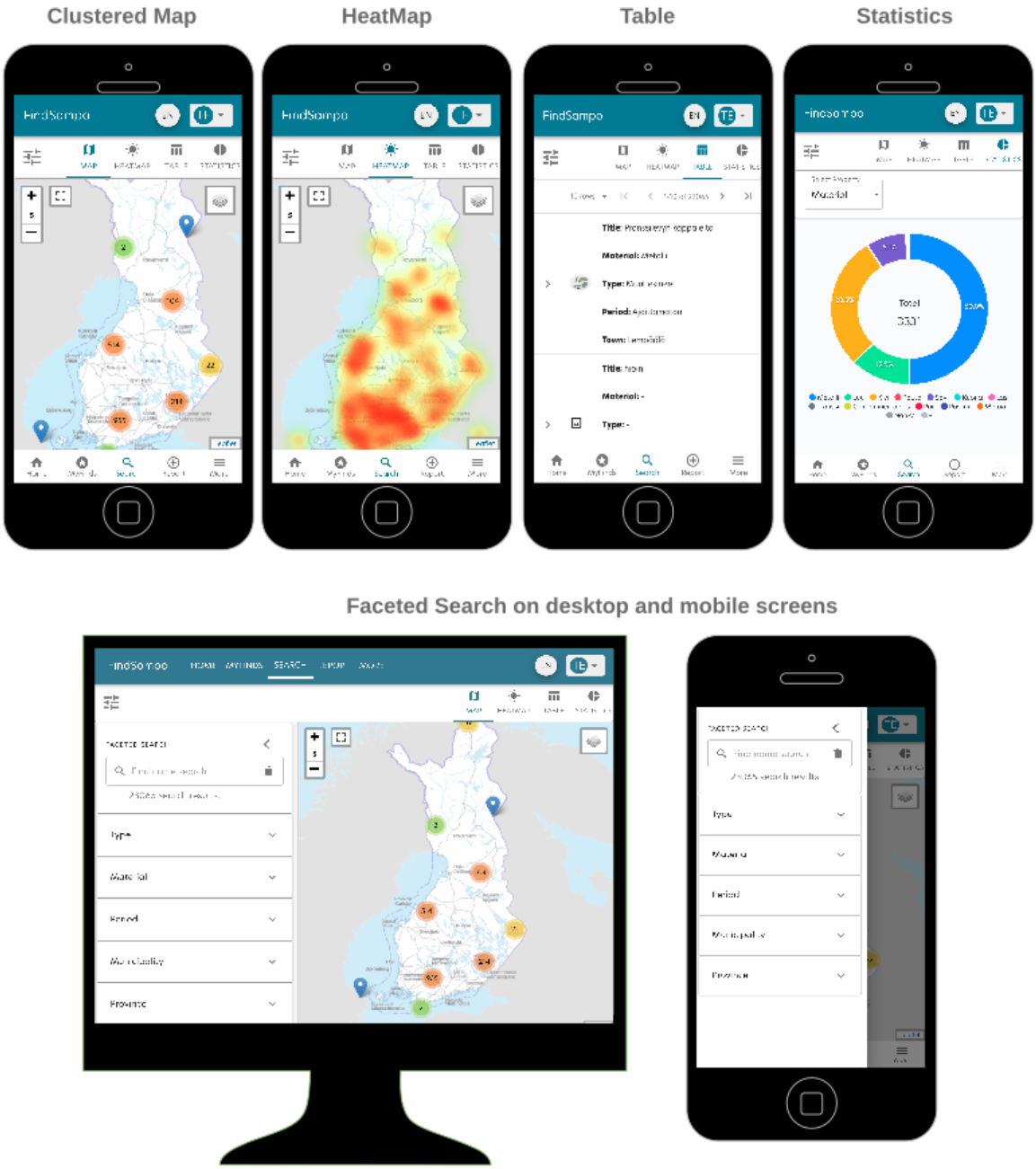

Fig 1. Different views in FindSampo Reporter. The Clustered Map view (left) provides an aggregated view of filtered finds on the map; the HeatMap view visualizes the filtered finds distribution in colours; the Table view lists the finds in a traditional way; the Statistics view illustrates statistical distributions of the finds along different facet dimensions, here based on the selected finds' material.

Figure 1 illustrates different views for visualising archaeological data with faceted search in FindSampo Reporter (Hassanzadeh et al 2020). In FindSampo Portal similar and more advanced data-analytic tools will be available. For example, the finds can also be projected on timelines (Figure 2) (Anafi et al 2020) and analysed statistically using a larger variety of seamlessly integrated data analytic and visualization tools.

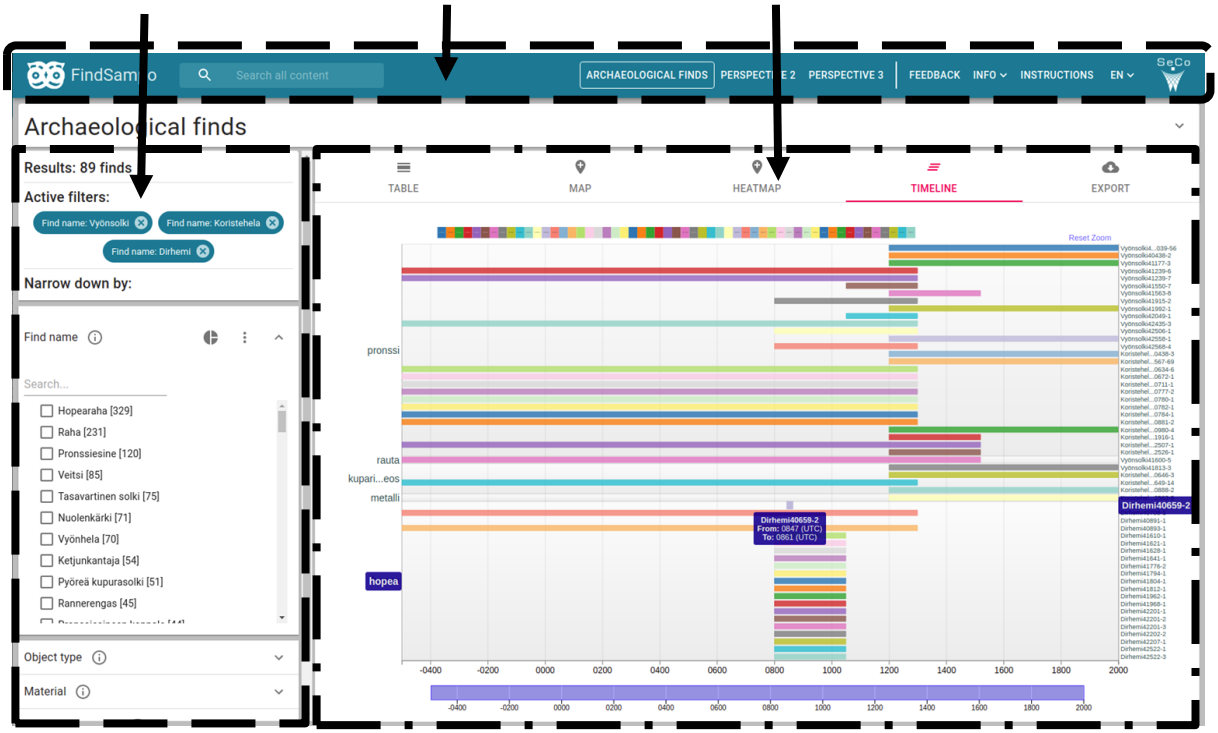

Fig. 2. The novel timeline visualization in the FindSampo portal's user interface (after Anafi et al 2020). The activated filters are shown on top of the facets in the faceted search section and the result of the filters is displayed on the activated tab in the result area. The timeline visualization shows the distribution of the filtered finds over time. The finds are grouped by material type, providing the user with a new kind of perspective on the material distribution of the finds in time.

The FindSampo Data Service with its SPARQL[5] endpoint and data download facility can be used for custom-made analyses. Different software tools can be employed for this. Fig. 3 presents a matrix showing probabilities that two types of items of the same era, here Iron Age, are found in the same area, here a municipality.

[5] SPARQL: https://www.w3.org/TR/sparql11-query

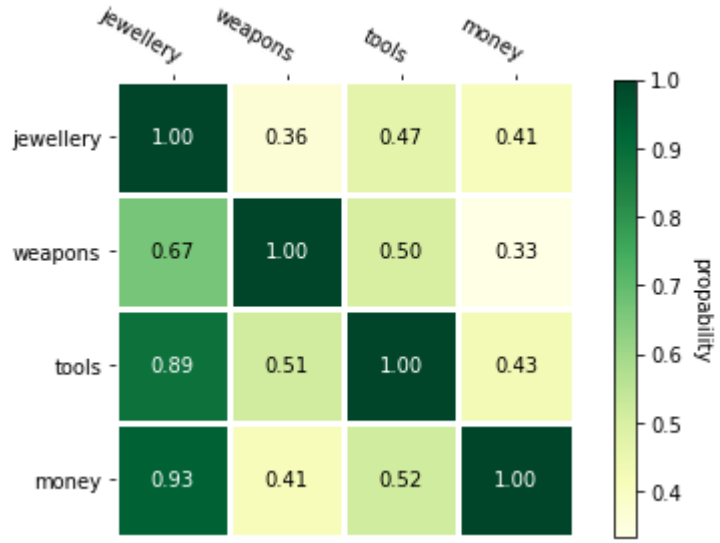

Fig 3. Analysis and visualization of co-occurring Iron Age object types found in the same municipality, made using Python Matplotlib library and a Google Colab notebook. If coins (money) are found, then the probability for jewellery is 093, but finding jewellery indicates coins with less probability 0.41. Probability for co-occurrence of weapon and coin finds seems low.

Infrastructure: Data and Ontologies

The data used in FindSampo contains initially approximately 3000 finds transformed into LD and published on the LDF.fi platform. This data model used describes the finds in terms of values taken from a set of (hierarchical) ontologies, such as object types, materials, and time periods. The ontologies harmonize heterogeneous data from different data sources, and are used to enrich the data by data linking to external data sources and by reasoning based on the Semantic Web logical standards[6]. The shared ontology infrastructure includes a new object type ontology of archaeological finds interlinked with the MAO/TAO ontology for Museum Domain and Applied Arts[7] and the AAT Thesaurus of the Getty Research Centre[8], and a time period ontology interlinked with the PeriodO[9] ontology, as recommended for international semantic interoperability in the ARIADNEplus project.

Future visions

The surge of new metal-detected find records in Finland since the 2010s is rewriting our understanding of material culture and associated fields in social, cultural and economic history. To actualising these developments, FindSampo framework will offer novel, groundbreaking qualitative and quantitative research tools to advance digital humanities and citizen science research. Further, a new Marie Skłodowska-Curie project[10], which began in September 2020, will deploy FindSampo and other FHA archaeological data to produce new analysis of large-scale and long-term development of Finnish archaeological landscapes. To

[6] https://www.w3.org/standards/semanticweb/
[7] https://finto.fi/maotao/fi/
[8] https://www.getty.edu/research/tools/vocabularies/aat/
[9] https://perio.do/en/
[10] https://cordis.europa.eu/project/id/896044

test the generalizability of the FindSampo framework, we plan to apply it to the large Portable Antiquities Scheme database managed by the British Museum[11]. These initiatives push forward to a deeper understanding of the agency of the public as creators of new knowledge about the past.

Acknowledgements
This paper is an output of the research project "SuALT—The Finnish Archaeological Finds Recording Linked Open Database (2017–2021)", funded by the Academy of Finland, decision numbers 310854, 310859, and 310860. Thanks to AriadnePlus and the Marie Skłodowska-Curie project *DeepFIN* grant agreement No. 896044) for (additional funding. CSC – IT Center for Science, Finland has provided computational resources for the work.

References

ANAFI, B., M. KOHO & E. HYVÖNEN. 2020. Temporal Visualization and Data Analysis of Archaeological Finds FindSampo. Poster, *Conference on Cultural Heritage and New Technologies (CHNT 25)*, Museum Stadt Archäologie Wien, Nov, 2020.

HASSANZADEH, P., E. HYVÖNEN, E. IKKALA, J. TUOMINEN, S. THOMAS, A. WESSMAN & V. ROHIOLA. 2020. FindSampo Platform for Reporting and Studying Archaeological Finds Using Citizen Science. *Proceedings of the 3rd Workshop on Humanities in the Semantic Web (WHiSe), Heraklion, Greece, June 2, 2020*: 33-40. CEUR Workshop Proceedings.

HEATH, T. & C. BIZER. 2011. (2011) *Linked Data: Evolving the Web into a Global Data Space* (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan & Claypool. doi:10.2200/S00334ED1V01Y201102WBE001

HYVÖNEN, E., J. TUOMINEN, M. ALONEN & E. MÄKELÄ. 2014. Linked Data Finland: A 7-star Model and Platform for Publishing and Re-using Linked Datasets, in V. Presutti, E. Blomqvist, R. Troncy, H. Sack, I. Papadakis, & A. Tordai (ed.) *The Semantic Web: ESWC 2014 Satellite Events. ESWC 2014*. 226-230, Cham: Springer.

HYVÖNEN, E. 2020. "Sampo" Model and Semantic Portals forDigital Humanities on the Semantic Web, in S. Reinsone, I. Skadiņa, A. Baklāne & J. Daugavietis (ed.) *DHN 2020: Digital Humanities in the Nordic Countries. Proceedings of the Digital Humanities in the Nordic Countries 5th Conference, Riga, Latvia, October 21-23, 2020*: 373-378. CEUR Workshop Proceedings.

TUNKELANG, D. 2009. Faceted Search. Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan & Claypool. doi:10.2200/S00190ED1V01Y200904ICR005

WESSMAN, A., S. THOMAS, V. ROHIOLA, M. KOHO, E. IKKALA, J. TUOMINEN, E. HYVÖNEN, J. KUITUNEN, H. PARVIAINEN, & M. NIUKKANEN. 2019. Citizen Science in Archaeology: Developing a Collaborative Web Service for Archaeological Finds in Finland, in J. Jameson & S. Musteaţă (ed.) *Transforming Heritage Practice in the 21st Century. One World Archaeology*, 337-352. Cham: Springer.

[11] https://www.britishmuseum.org/our-work/national/treasure-and-portable-antiquities-scheme