Aalto University publication series
Doctoral Theses 225/2025

Publishing, exploring, and analyzing cultural heritage linked data

Heikki Rantala

A doctoral thesis completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Science at a public examination held at the lecture hall TU2 of the school on 21.11.2025 at 12 noon.

Aalto University
School of Science
Department of Computer Science

Supervising professor
Professor Eero Hyvönen, Aalto University & University of Helsinki, Finland

Thesis advisors
Dr Jouni Tuominen, University of Helsinki & Aalto University, Finland
Dr Eljas Oksanen, Title of Docent, University of Helsinki, Finland & University of Reading, UK

Preliminary examiners
Professor Harald Sack, FIZ Karlsruhe – Leibniz Institute for Information Infrastructure & Karlsruhe Institute of Technology, Germany
Professor Isto Huvila, Uppsala University, Sweden

Opponent
Professor Isto Huvila, Uppsala University, Sweden

Aalto University publication series
Doctoral Theses 225/2025

ISBN 978-952-64-2823-9 (paperback)
ISBN 978-952-64-2822-2 (pdf)
ISSN 1799-4934 (print)
ISSN 1799-4942 (pdf)

http://urn.fi/URN:ISBN:978-952-64-2822-2

PunaMusta Oy
2025

Aalto University

ABSTRACT

Author Heikki Rantala

Name of the doctoral thesis Publishing, exploring, and analyzing cultural heritage linked data

Article-based thesis

Number of pages 73

Keywords linked data, cultural heritage, digital humanities, archaeology, war history

Memory organizations produce various kinds of Cultural Heritage (CH) data that is of great importance for research and society in general, but there are challenges in making the data available to the general public and humanities researchers. The data often exists in data silos, sometimes even just as tabular files on computers. This thesis aims to address these issues by answering questions concerning how CH data can be published in interoperable form for semantic applications, what kinds of functions such applications can have, and how the creation of such CH web applications can be supported.

The hypotheses selected in this thesis are that Linked Data (LD) can offer a practical way of publishing CH data and creating semantic applications; that faceted search, combined with visualizations for data analysis, offers both researchers and the general public a useful way to explore data; that CH web applications can be created on top of SPARQL endpoints; and that their creation can be supported with a software framework created for that purpose.

The research questions and hypotheses were tested by creating new knowledge graphs, web applications, and methods to open CH as LD and explore it using faceted search. The example cases used in this thesis to test the research questions and hypotheses include death records from the Finnish Civil War, records of archaeological finds made by the public, and biographies of prominent individuals from Finland and various other European countries. The existing legacy CH datasets were converted to LD and enriched in cooperation with experts from heritage institutions.

The work in this thesis demonstrates that LD can be a practical way to open CH data, that faceted search applications can be built on top of SPARQL endpoints, and that faceted search combined with visualizations is a useful tool for opening CH data to be utilized by both researchers and the general public. The creation of CH applications requires high-quality data, but such applications can also support improving data quality. The research in this thesis also extends the concept of faceted search by introducing a novel method for applying faceted search to relations between entities such as persons and places in addition to the entities themselves.

Aalto-yliopisto

TIIVISTELMÄ

Tekijä Heikki Rantala

Väitöskirjan nimi Kulttuuriperintöön liittyvän linkitetyn datan julkaiseminen, tutkiminen ja analysoiminen

Artikkeliväitöskirja

Sivumäärä 73

Avainsanat linkitetty data, kulttuuriperintö, digitaaliset ihmistieteet, arkeologia, sotahistoria

Muistiorganisaatiot tuottavat erilaista kulttuuriperintöön (KP) liittyvää dataa, jolla on suuri merkitys tutkimukselle ja yhteiskunnalle. Datan avaamisessa yleisöä ja tutkijoita varten on kuitenkin esteitä. Data sijaitsee usein "datasiiloissa", esimerkiksi taulukkotiedostoina tutkijoiden tietokoneilla. Tämä väitöskirja pyrkii parantamaan tilannetta vastaamalla kysymyksiin siitä, miten KP-dataa voidaan julkaista yhteentoimivassa muodossa semanttisille sovelluksille, millaisia ominaisuuksia sellaisilla sovelluksilla voi olla ja miten sellaisten sovellusten luomista voi helpottaa.

Tähän työhön on valittu seuraavat hypoteesit: linkitetty data (LD) voi tarjota käytännöllisen tavan julkaista KP-dataa ja luoda semanttisia sovelluksia; fasettihaku, yhdistettynä data-analyysejä mahdollistaviin visualisaatioihin, tarjoaa tutkijoille ja suurelle yleisölle hyödyllisen tavan tutkia dataa; web-sovelluksia voidaan luoda SPARQL-päätepisteiden varaan ja sovellusten luomista voi helpottaa hyödyntämällä valmista sovelluskehystä.

Tutkimuskysymyksiä ja hypoteeseja testattiin luomalla uusia tietämysgraafeja, web-sovelluksia ja metodeja KP-datan avaamiseen ja datan tutkimiseen. Esimerkkitapauksina tässä työssä käytettiin Suomen sisällissodan uhrien tilastoja, arkeologisia kansalaislöytöjä sekä merkittävien henkilöiden elämäkertoja. Esimerkkitapauksissa olemassa olevat tietokannat muunnettiin linkitetyksi dataksi ja niitä rikastettiin yhteistyössä KP-asiantuntijoiden kanssa.

Tämä väitöskirja osoittaa, että LD voi olla käytännöllinen tapa avata kulttuuriperintödataa, että fasettihaku on hyödyllinen työkalu kulttuuriperintödatan avaamisessa ja että sovelluksia voidaan luoda SPARQL-päätepisteiden varaan. KP-sovellusten luominen edellyttää laadukasta dataa, mutta tällaiset sovellukset voivat myös auttaa datan laadun parantamisessa. Tutkimus laajentaa lisäksi fasettihaun käsitettä tietämysgraafien solmujen, kuten henkilöiden ja paikkojen, välisten yhteyksien hakemiseen.

Preface

The research presented in this thesis was conducted at the Semantic Computing Research Group (SeCo) at the Department of Computer Science, Aalto University, in collaboration with the Helsinki Centre for Digital Humanities at the University of Helsinki (HELDIG). The work has benefited from a large body of previous work and the existing infrastructure created by the research group.

I would like to thank my supervisor, professor Eero Hyvönen. I'm also grateful to my advisors Jouni Tuominen and Eljas Oksanen. I also thank Professor Petri Vuorimaa, who almost, but not quite, ended up as my supervisor.

My co-writers have obviously been indispensable for writing this thesis. They include Annastiina Ahola, Frida Ehrnsten, Esko Ikkala, Ilkka Jokipii, Mikko Koho, Ville Rohiola, Anna Wessman, Petri Leskinen, and Lilli Peura. I would also like to thank all the other co-workers I have had during my years in the Semantic Computing Research Group. I would especially like to mention Senka Drobac, Erkki Heino, Joonas Kesäniemi, Rafael Leal, Matti La Mela, Henna Poikkimäki, Minna Tamper, Sarah Shoilee, Sami Sarsa, Telma Peura, Babatunde Anafi, Patrik Boman, Sisko Pajari, Muhammad Faiz Wahjoe, Pejam Hassanzadeh, Kasper Apajalahti, Hien Cao, Arttu Oksanen, Ossi Koho, Lia Gasbarra, and Laura Sinikallio.

The research for this thesis was conducted in multiple projects and involved numerous partners. I would like to thank all the collaborators, particularly in the Finnish Heritage Agency, The National Museum of Finland, National Archives of Finland, and the Finnish Literature Society.

Semantic Web Publications – Texts as Data Services (Severi) project was funded by Business Finland. I particularly thank Aki Hietanen and Kirsi Keravuori, who served on the steering committee, for collaboration on this and other projects.

WarVictimSampo was funded by the Ministry of Education and Culture and was a collaboration with the National Archives of Finland. I would like to thank Päivi Happonen, Vili Haukkovaara, Jarmo Nieminen, and Markku Mäenpää who served on the steering committee of the project and provided useful discussions.

The SuALT project was funded by the Research Council of Finland under decision numbers 310854, 310859, and 310860. I would like to particularly thank Suzie Thomas, who led the project.

For the work in the DigiNUMA project, I was supported by a grant from the Jenny and Antti Wihuri Foundation. The project was also supported by Marie Skłodowska-Curie Actions of the European Commission, Aalto University, and University Helsinki, Helsinki Institute for Social Sciences and Humanities (HSSH). I would like to thank Michael Lewis and David Wigg-Wolf, who advised the project.

InTaVia project was funded by the European Commission within the topic DT-TRANSFORMATIONS-12-2018-2020 "Curation of digital assets and advanced digitisation" (project ID: 101004825). Out of the many people all over Europe I worked with in this project, I would like to particularly thank Victor de Boer, Eva Mayr, Matthias Schlögl, and Florian Windhager.

I have also worked on multiple other projects. FIN-CLARIAH research infrastructure program is funded by the Research Council of Finland and the European Union – NextGenerationEU instrument under grant number 346323. SEMPARL project involving parliamentary speeches was funded by the Research Council of Finland as part of the DIGIHUM 2020–2022 program. Constellations of Correspondence project dealing with 19th century correspondence was funded by the Research Council of Finland. ARCH-ON project on archaeological ontologies has been funded by KU Leuven, Aalto University, and University of Helsinki. WarMemorySampo project was funded by Tammenlehvän Perinneliitto ry. OperaSampo was a collaboration between the University of the Arts Helsinki and Aalto University. I also worked on cross-border legislation with the Nordic-Baltic project Achieving the World's Smoothest Cross-Border Mobility and Daily Life Through Digitalisation which was funded by the Nordic Council of Ministers and the Cross-border Digital Services program.

I also received grants to support my thesis work from the Finnish Cultural Foundation and Etelä-Pohjalainen Osakunta. CSC – IT Center for Science provided computational resources for the research.

I would also like to thank Pieterjan Deckers, Kristiina Mannermaa, Mika Lavento, Anne Kauppala, Kārlis Čerāns, Uldis Bojars, Ilona Pikkanen, Johanna Enqvist, and Hanna-Leena Paloposki for collaboration and interesting discussions. I am also grateful to Doctor Vesa Muhonen, as well as to my other friends and teachers in pursuits outside academia.

Finally, I would like to thank my family who supported me in every way and Kaisa who was always optimistic and promised that I would be ready soon.

Helsinki, November 3, 2025,

Heikki Rantala

Contents

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I Eero Hyvönen and Heikki Rantala. Knowledge-Based Relational Search in Cultural Heritage Linked Data. *Digital Scholarship in the Humanities*, Volume 36, Issue Supplement 2, Pages 155–164, Oxford University Press, online `https://doi.org/10.1093/llc/fqab042`, October 2021.

II Heikki Rantala, Ilkka Jokipii, Esko Ikkala and Eero Hyvönen. WarVictimSampo 1914–1922: a National War Memorial on the Semantic Web for Digital Humanities Research and Applications. *ACM Journal on Computing and Cultural Heritage (JOCCH)*, Volume 15, issue 1, pages 1–18, Association for Computing Machinery, online `https://doi.org/10.1145/3477606`, January 2022.

III Esko Ikkala, Eero Hyvönen, Heikki Rantala and Mikko Koho. Sampo-UI: A Full Stack JavaScript Framework for Developing Semantic Portal User Interfaces. *Semantic Web – Interoperability, Usability, Applicability*, Volume 13, Issue 1, Pages 69–84, IOS Press, `https://doi.org/10.3233/SW-210428`, January 2022.

IV Heikki Rantala, Esko Ikkala, Ville Rohiola, Mikko Koho, Jouni Tuominen, Eljas Oksanen, Anna Wessman and Eero Hyvönen. FindSampo: A Linked Data Based Portal and Data Service for Analyzing and Disseminating Archaeological Object Finds. In *The Semantic Web: 19th International Conference, ESWC 2022, May 29–June 2, 2022, Proceedings, Hersonissos, Greece*, Paul Groth, Maria-Esther Vidal, Fabian Suchanek, Pedro Szekley, Pavan Kapanipathi, Catia Pesquita, Hala Skaf-Molli, Minna Tamper (editors), Lecture Notes in Computer Science, volume 13261, pages 478–494, Springer, Cham, `https://doi.org/10.1007/978-3-031-06981-9_28`, May 2022.

V Eljas Oksanen, Frida Ehrnsten, Heikki Rantala and Eero Hyvönen. Semantic Solutions for Democratising Archaeological and Numismatic Data Analysis. *ACM Journal of Computing and Cultural Heritage*, Volume 16, issue 4, Pages 1–18, Association for Computing Machinery, `https://doi.org/10.1145/3625302`, January 2024.

VI Heikki Rantala, Annastiina Ahola, Esko Ikkala and Eero 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 the 8th International Workshop on the Visualization and Interaction for Ontologies (VOILA! 2023), Linked Data and Knowledge Graphs, Athens, Greece*, CEUR Workshop Proceedings, Vol-3508, `https://ceur-ws.org/Vol-3508/paper3.pdf`, November 2023.

VII Heikki Rantala, Eljas Oksanen, Frida Ehrnsten and Eero Hyvönen. Publishing Numismatic Public Finds on the Semantic Web for Digital Humanities Research – CoinSampo Linked Open Data Service and Semantic Portal. In *Proceedings of the First International Workshop of Semantic Digital Humanities (SemDH 2024, Hersonissos, Greece)*, CEUR Workshop Proceedings, Vol-3724, `https://ceur-ws.org/Vol-3724/paper3.pdf`, May 2024.

VIII Heikki Rantala, Petri Leskinen, Lilli Peura and Eero Hyvönen. Representing and Searching Associations in Cultural Heritage Knowledge Graphs Using Faceted Search. In *Proceedings of the 20th International Conference on Semantic Systems, Amsterdam, The Netherlands, Studies on the Semantic Web, Volume 60: Knowledge Graphs in the Age of Language Models and Neuro-Symbolic AI*, pages 420 – 435, IOS Press, `https://doi.org/10.3233/SSW240033`, September 2024.

Author's Contribution

Publication I: "Knowledge-Based Relational Search in Cultural Heritage Linked Data"

Heikki Rantala contributed to the writing of the article as a co-author and wrote about 40% of the article. Rantala did all the technical work for this paper. The idea for the article was proposed by Eero Hyvönen.

Publication II: "WarVictimSampo 1914–1922: a National War Memorial on the Semantic Web for Digital Humanities Research and Applications"

Heikki Rantala is the lead author of the publication and wrote about 50% of the article. Ilkka Jokipii is a historian and provided subject matter expertise for the project. He was responsible for curation of the data and wrote about 10% of the article, including chapter 1.1. Esko Ikkala wrote around 20% of the article, including chapter 4.2 on the technical implementation of the user interface. Rantala was responsible for technical aspects for data conversion and designing data models. The user interface was created by Rantala and Ikkala. Eero Hyvönen wrote about 20% of the article including chapters 2, 1.3 and 3.6.

Publication III: "Sampo-UI: A Full Stack JavaScript Framework for Developing Semantic Portal User Interfaces"

Heikki Rantala contibuted to the writing of the article as a co-author, and participated to the design of the tool. He wrote about 15% of the article. Rantala wrote chapter five about a use case. Rantala participated in the planning and testing of the framework. Esko Ikkala is the main author and wrote about 65 percent of the article. Eero Hyvönen wrote about 15 percent of the article. Mikko

Koho wrote about 5 percent of the article. Esko Ikkala did most of the design and technical work in implementing the framework.

Publication IV: "FindSampo: A Linked Data Based Portal and Data Service for Analyzing and Disseminating Archaeological Object Finds"

Heikki Rantala is the lead author of the publication and wrote about 50% of the article. Rantala was mainly responsible for the technical implementation of data conversion and creation of the user interface. Esko Ikkala wrote about 10% of the article and is responsible for about half of the technical work of creating the web application. Ville Rohiola is responsible for citizen finds in the Finnish Heritage Agency and provided subject matter expertise in the project. Rohiola worked on the vocabularies and data. He wrote less than 5% of the article. Eljas Oksanen is a historian and archaeologist and provided subject matter expertise on metal detecting. He wrote around 10% of the article, including parts of the introduction and related work chapters. Mikko Koho and Jouni Tuominen worked with the data and wrote less than five percent of the article. Anna Wessman is an archaeologist and provided subject matter expertise for the publication. She wrote less than 5% of the article. Eero Hyvönen wrote around 10% of the article including chapter 1.3 dealing with the Sampo Model and parts of the related work chapter 1.2.

Publication V: "Semantic Solutions for Democratising Archaeological and Numismatic Data Analysis"

Heikki Rantala contibuted to the writing of the article as a co-author and wrote about 25% of the article with focus on 3. Methods and 5. Data Service and Semantic Portal. Eljas Oksanen conceptualized the article and wrote about 45%, with a focus on 1. Introduction, 2. Related Work, 4. Research Data, 7. Discussion and 8. Conclusion. Frida Ehrnsten wrote about 25% of the article with a focus on chapter 6. about CoinSampo Use Cases. Eero Hyvönen wrote about 5% of the article and edited the final version. However, all authors contributed to all chapters. Rantala did all the technical work for this paper.

Publication VI: "How to Create Easily a Data Analytic Semantic Portal on Top of a SPARQL Endpoint: Introducing the Configurable Sampo-UI Framework"

Heikki Rantala is the lead author of the publication and wrote about 55% of the article. Annastiina Ahola wrote about 25% of the article. Eero Hyvönen wrote about 20% of the article. Esko Ikkala wrote less than 5% of the article. Rantala wrote most of Chapters 1, 2, and 5. Chapter 4 was written equally by Rantala and Ahola. Hyvönen wrote most of Chapter 3. Esko Ikkala is the main developer of the Sampo-UI framework.

Publication VII: "Publishing Numismatic Public Finds on the Semantic Web for Digital Humanities Research – CoinSampo Linked Open Data Service and Semantic Portal"

Heikki Rantala is the lead author of the publication and wrote about 60% of the article. Eljas Oksanen wrote about 20% of the article. Frida Ehrnsten wrote about 15% of the article. Eero Hyvönen wrote about 5% of the article. Rantala wrote chapters 3 and 4 regarding data and web application. Oksanen led on chapter 2 on Related Work, and Oksanen and Ehrnsten co-wrote chapter 5 on the use case. The introduction, related work, and conclusion chapters were written in co-operation with all the authors. Rantala did the technical work for the paper. Ehrnsten created the original data.

Publication VIII: "Representing and Searching Associations in Cultural Heritage Knowledge Graphs Using Faceted Search"

Heikki Rantala is the lead author of the publication and wrote about 50% of the article. Petri Leskinen wrote about 20% of the article. Lilli Peura wrote less than 5% of the article. Eero Hyvönen wrote about 30% of the article. Rantala and Hyvönen wrote the chapters 1 and 2 together. Rantala wrote chapter 3 and most of the chapter 4, except for the Wikipedia case study which was written by Leskinen. Leskinen did all the technical work related to the Wikipedia case. In other example cases, Rantala created the user interfaces and designed the data models. Peura did most of the work of creating SPARQL queries as instructed by Rantala.

List of Figures

List of Tables

Abbreviations

AAT Art & Architecture Thesaurus

CH Cultural Heritage

CIDOC International Committee for Documentation

CIDOC CRM CIDOC Conceptual Reference Model

CSV Comma-Separated Values

DC Dublin Core

DH Digital Humanities

EDM Europeana Data Model

FHA Finnish Heritage Agency

KG Knowledge Graph

LD Linked Data

LLM Large Language Model

LOD Linked Open Data

OWL Web Ontology Language

PAS Portable Antiquites Scheme

PDF Portable Document Format

RDF Resource Description Framework

SKOS Simple Knowledge Organization System

SOCH Swedish Open Cultural Heritage

SPARQL SPARQL Protocol and RDF Query Language

UI User Interface

ULAN Union List of Artist Names

URI Uniform Resource Identifier

XML Extensible Markup Language

1. Introduction

1.1 Background and Research Environment

Cultural Heritage (CH) [35, 59] refers to the inherited objects, traditions, and environment of a culture. *Memory organizations*, such as museums, libraries, and archives, manage large amounts of heterogeneous and interlinked CH (meta)data. Such data is important for understanding our societies, which makes opening the data to researchers and the general public in the best possible form important. However, CH data often exists in separate “data silos”. This refers to isolated data repositories that use incompatible data models and vocabularies and cannot be easily accessed. Sometimes, they are just tabular files on the computers of researchers or museum curators; in other cases, there may be a web application with limited functions using closed endpoints. The information in the data is also often expressed using uncontrolled terms that are sometimes not consistent even in the context of that dataset.

This situation is not ideal for humanities research. Opening the data in a way that is easy to access and that is interoperable with other CH data makes comparing the data easier. It also allows for easy reuse of resources, so that, for example, coordinates or a hierarchy of places do not need to be redone for each research project. To enable CH data to be used to its full potential, the data should be findable, accessible, interoperable, and reusable for people and computers, as stated by the FAIR principles [98]. The approach selected and tested in this thesis to address these issues is to publish the data as *Linked Data* (LD) [21, 14] using *Resource Description Framework (RDF)* [30] *knowledge graphs* (KG).

LD can also be referred to as *Linked Open Data* (LOD) when focusing on the openness of the data. The concept of openness can be seen to encompass multiple aspects, including technical openness, legal openness, and accessibility [88]. In this thesis, openness is seen as something that can always be improved in some way with, for example, better documentation or better search functions.

This thesis presents multiple cases of opening CH data of memory organizations using LD. The data used in the case studies presented in this thesis is mainly from national Finnish heritage organizations: the National Archives of Finland, the Finnish Heritage Agency (FHA), and the National Museum of Finland. The datasets range from records of victims of the Finnish Civil War to records of archaeological metal detecting finds. The development of these datasets has been done in cooperation with leading experts from memory organizations.

Often in CH a distinction is made between data and metadata. Metadata is “data about data”, and plays an important role in the description and administration of CH data [40]. Although metadata is an important concept for CH research, in this thesis the distinction between data and metadata is generally not made, since the distinction is not very important for the purposes of this thesis, and metadata is always also data[1]. The examples in this thesis mainly deal with structured data about, for example, archaeological objects. If the distinction were made, this kind of data could generally be considered metadata.

Opening LD KGs using *triplestores* allows technically inclined researchers and application developers to query the data directly using the SPARQL [54] query language. However, most researchers and especially the general public often need user interfaces such as web applications that open the data in a more easy-to-use manner. Web applications can be very important in making it possible for end users, both the general public and researchers, to find the information they are interested in [65]. When applications use the semantics of a domain, they can be called *semantic applications* [10].

A major part of this thesis deals with how such applications can be designed, what kind of features they can have, and how the creation of such applications can be supported. Web applications presented in this thesis use SPARQL to query the data. The approach for designing CH web applications selected and tested in this thesis is based on the *faceted search* paradigm [149]. The work gives multiple examples of how faceted search can be applied to explore and disseminate CH data. The thesis also presents a new way to approach faceted search by searching relations.

An example of how such an application can look is shown in Figure 1.1 showing the FindSampo[2] web application, where a user browses swords found in Finland using the search functions of the application. The facets on the left allow users to make selections and the hit counts of the facets are automatically updated after each selection. The user has selected swords (“miekka”) from the “object type” facet. Because the facet uses a vocabulary with a hierarchy, this also selects all the concepts that are defined to be below the sword in the hierarchy, such as sword hilts (“miekan kahva”). The user could also further refine the search by, for example, selecting the medieval period from the period facet. The user can

[1] See for example discussion by Tim Berners-Lee on metadata and Web in `https://www.w3.org/DesignIssues/Metadata.html` and `https://www.w3.org/DesignIssues/Semantic.html`.

[2] The FindSampo web application can be tested online at: `https://findsampo.fi/`

browse the result set for each individual find, or they can visualize the entire result set with various charts or maps. However, the hit counts of facets can be interesting in themselves and reveal various insights. For example, in the example the hit count for swords is 35 while the hit count for spears ("keihäs") is 36, meaning that a roughly similar amount of finds in the data are swords and spears. The application queries an LD graph with SPARQL and is created using the Sampo-UI framework presented in this thesis.

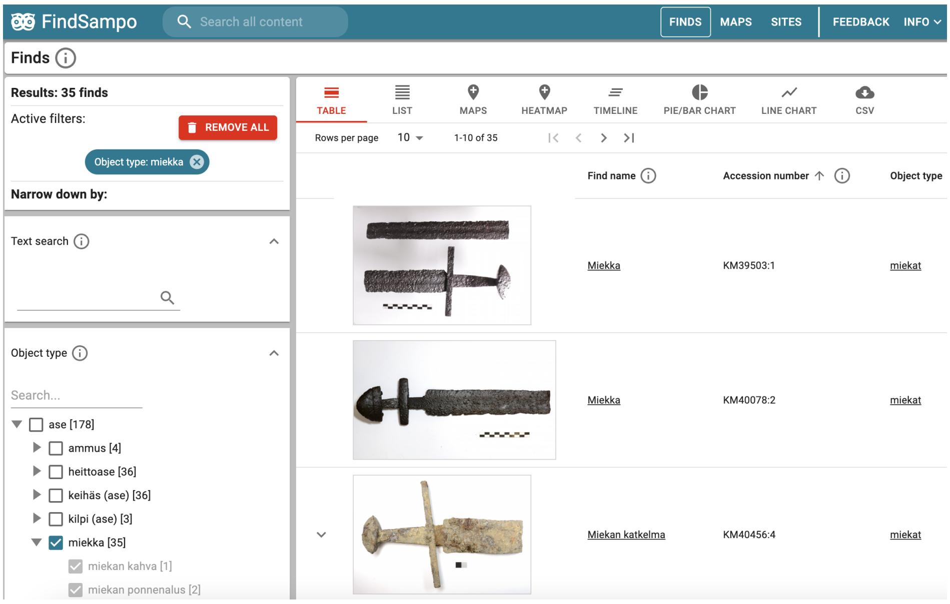

Figure 1.1. An example of opening CH LOD with web application: browsing the swords in FindSampo KG.

An example of a service that a CH web application can offer to end users is *relational search* [65]. Relational search or *association search* [136, 28] deals with finding potentially interesting connections between entities in data. In the CH field, this might mean, for example, finding connections formed by genealogical or teacher-student relations [77], or finding out how people are related to different places based on biographical life events. This thesis presents a novel approach to how such a search can be implemented with CH data.

The individual papers of this thesis have been created as part of multiple research projects. War Victims of Finland 1914–1922 database [159] collects information about Finnish people killed in wars between 1914 and 1922. This is a nationally important database that was originally compiled in a government funded project and opened with a web interface in 2002. It is maintained by the National Archives of Finland. Publication II describes work done to technically update the database into a knowledge graph based on ontologies and LD. The new KG was also improved by adding new information, which resulted in the creation of the WarVictimSampo 1914–1922 (referred to as WarVictimSampo) LOD service and portal. This project[3] was funded by the Finnish government

[3] `https://seco.cs.aalto.fi/projects/sotasurmat-1914-1922/en/`.

and was carried out in cooperation between the Finnish National Archives and Aalto University.

Publication IV deals with the opening of the data on archaeological finds made by the public in Finland. The work was carried out as part of the Finnish Archaeological Finds Recording Linked Open Database (SuALT)[4] [148, 158, 78] project. This was a multidisciplinary project conducted in cooperation with Aalto University, the University of Helsinki, and the Finnish Heritage Agency to develop solutions to respond to the need caused by the increased number of archaeological objects found by the public, usually by hobby metal detecting.

Publication V and Publication VII were written as a part of DigiNUMA: Digital Solutions for European Numismatic Heritage[5] project. The project was carried out in cooperation with Aalto University, the University of Helsinki, and the National Museum of Finland to address the challenges in creating, publishing, and analyzing information on numismatic finds reported by citizens.

Publication I was created as part of Semantic Web Publications – Texts as Data Services (Severi)[6] project and is related to BiographySampo [73, 74, 76] data service and portal, that publish biographical data of the Finnish Literature Society as Linked Data. Publication VIII was created as part of the In/Tangible European Heritage – Visual Analysis, Curation, and Communication (InTaVia)[7] project and is based on the work and reports created during the project. InTaVia was a project to harmonize tangible and intangible cultural heritage with participants from multiple European countries.

Publication III and Publication VI deal with the development of web interfaces for LOD. The created framework is called Sampo-UI[8]. The Sampo-UI framework has been and continues to be developed in the course of multiple projects aimed at disseminating LOD with applications to search, browse, and analyze data.

In addition to these publications, the author has contributed to several other related articles. They include work dealing with Finnish war victims of the Second World War [97], publishing interviews of Finnish war veterans [96], Finnish historical occupations [41], publishing biographical data [73, 75], 19th century letter data [37, 72], Finnish Opera performances [2], art history [3], harmonizing cross-border legislation [70], intangible cultural heritage [69], effects of human decision on archaeological data [104], and parliamentary speeches [63]. The FindSampo project is discussed further in [78] and the DigiNUMA project in [117]. In addition, a number of reports, abstracts, posters, and shorter conference papers have been written as part of the process of creating the articles included in this thesis. The work in this thesis builds on and extends the LOD infrastructure for CH [67] and the Sampo series of systems [66]. The work

[4] https://blogs.helsinki.fi/sualt-project/.
[5] https://seco.cs.aalto.fi/projects/diginuma/.
[6] https://seco.cs.aalto.fi/projects/severi/.
[7] https://intavia.eu/
[8] https://seco.cs.aalto.fi/tools/sampo-ui/

continues as part of the national FIN-CLARIAH[9] research infrastructure for Social Sciences and Humanities in Finland, which combines the Pan-European CLARIN[10] and DARIAH[11] infrastructures in the Finnish CH context.

1.2 Objectives and Scope

The aim of this thesis is to find ways to make CH data more accessible and usable to the general public and researchers. This includes improving the quality of the data itself by making it more findable, accessible, interoperable, and reusable, but also creating web applications that utilize semantics that can make searching and analyzing the data faster and require less technical knowledge. This thesis concentrates on three aspects of the process of bringing CH data to the end users: modeling and enriching CH data, recognizing potential features that applications based on CH data can have, and the process of creating semantic applications for the CH domain. Each of these aspects is represented by one of the following research questions.

RQ1. How can CH data be published in a way that is interoperable and usable in semantic applications aimed at researchers and the general public?

RQ2. How can semantic applications be used to disseminate and analyze CH data in a way that does not require technical expertise from users?

RQ3. How can the process of developing semantic applications employing faceted search be supported?

RQ1 examines the process of making CH data accessible for research purposes and for the development of semantic applications. This includes modeling, linking, and enriching the data. The hypothesis used in this thesis to answer this research question is that LD can be used to open CH data and can serve as a basis for creating semantic applications.

RQ2 deals with how semantic web applications can be used to open CH data. The hypothesis used in this thesis to answer this research question is that faceted search combined with visualizations for data analysis can be a useful way to explore the data for both researchers and the general public.

RQ3 focuses on the technical aspects of developing web applications based on LD and faceted search. Specifically, the focus is on how to support the development process so that it requires less effort and expertise from developers. The hypothesis used in this thesis for this research question is that such applications can be built on top of SPARQL endpoints, and that development time and

[9] https://seco.cs.aalto.fi/projects/fin-clariah/
[10] https://www.clarin.eu/
[11] https://www.dariah.fi/

required technical expertise can be reduced by using a software framework that is a working application that can be used as a starting point for making a new application with new data.

Table 1.1 shows how the publications are mainly related to the research questions.

Publication	RQ1	RQ2	RQ3
Publication I	x	x	
Publication II	x	x	
Publication III			x
Publication IV	x	x	
Publication V	x	x	
Publication VI			x
Publication VII	x	x	
Publication VIII	x	x	

Table 1.1. The relationship between the publications and research questions

1.3 Research Methods and Dissertation Structure

The methodology of this thesis is inspired by the paradigm of *design science* [57, 125]. Design science is a research methodology where research is approached by attempting to solve important problems by building and then evaluating artifacts. The created artifacts can in practice be software systems, models, methods, vocabularies, or other useful artifacts. The design science process is iterative. It consists of defining a problem, developing an artifact to help solve the problem, and then evaluating the artifact. Finally, the results should be communicated to the appropriate audience [125]. The artifacts created as part of this thesis process include a variety of methods, knowledge graphs, ontologies, frameworks, and web services.

The work in this thesis can also been seen to represent action research [11, 6]. Action research is a class of research approaches that has its origin in social and medical sciences and that has also gained importance in information systems research. In action research researcher often works in collaboration with practitioners of certain field with complex problems relevant to the practitioners. The goal is to gain shared understanding of the issues faced and to improve the issues faced by practitioners as part of a collaborative, continuous, and often iterative process. A great part of the work presented in this thesis has been done in collaboration with cultural heritage practitioners from heritage institutions, such as the National Archives of Finland and the Finnish Heritage Agency, and humanities researchers, to help with their existing issues and needs.

Design science and action research have been described as similar [86] and dissimilar [81] from each other. This thesis considers them as distinct but complementary approaches and sources of inspiration for conducting research. This thesis can be considered to fall under the field of Digital Humanities (DH) [145, 106], as the articles of the thesis offer computational solutions to problems in humanities. DH can be seen as either a discipline of it's own, or a cross-disciplinary approach that applies computational methods to humanities related topics.

This thesis is structured as follows. In Chapter 2, the theoretical foundations related to the research are presented. In Chapter 3, the results of the publications are reviewed and summarized. Chapter 4 discusses the significance of the results and the reliability and validity of the research, and provides experiences and lessons learned with suggestions for the directions of further research.

2. Theoretical Foundation

2.1 Publishing Cultural Heritage Data

CH can be tangible, intangible, or natural [59]. The research in this thesis deals with both tangible CH in the form of archaeological object finds and intangible CH in the form of death records and other biographical information. The key data producers in the CH sector are memory organizations, but citizen volunteers can sometimes contribute to the creation of data. An important challenge of CH data is that it is often separated into individual data silos [122]. There is also a shortage of interfaces, especially for intangible CH [162]. Making data better available for research and dissemination to the general public is a central theme in this thesis.

The datasets opened as part of the projects related to this thesis were created in part based on reports from nonprofessional volunteers. The new war victims and corrections to errors in WarVictimSampo are mainly based on information collected by a nonprofessional individual who collected the data as his own personal project. Similarly, the data in FindSampo and CoinSampo are based on reports from hobby metal detectorists. This kind of activity can be described as *citizen science* [52]. Although there are varying definitions for the term, for the purposes of this thesis, it can be broadly seen as an activity where volunteer members of the general public engage in an activity aimed at improving public knowledge. Examples of this kind of activity relevant to this thesis include a project [135] initiated by the Finnish Broadcasting Company YLE, where members of the public could record and share locations relevant to conflicts in Finland in the 20th century. Many of the map visualizations used in web applications created for the research in this thesis are created using data from the OpenStreetMap[1] [51] which is a community project where geographic data is created and shared openly by volunteers. In addition, many of the programming libraries, tools, and systems used in work related to this thesis are community-generated.

[1] `https://www.openstreetmap.org/`

Three of the publications in this thesis are related to opening data of archaeological object finds made by the public. These kinds of finds are mainly reported by hobby metal detectorists. Hobby metal detecting [157, 156, 118] is a leisure activity that has grown considerably in popularity during the past several decades in multiple European countries, including Finland. The legal status of metal detecting varies from country to country. In Finland[2], metal detection is generally legal outside certain protected areas, such as known ancient monuments, with the permission of the landowner. Legally, all finds older than one hundred years need to be reported to the FHA. Irresponsible metal detecting can lead to the theft of valuable artifacts or the destruction of archaeological sites, but responsible metal detecting can lead to the recording of new finds and the identification of new archeological sites that may otherwise be at risk of destruction. Multiple recording schemes have been developed in Europe for archaeological finds made by the public. One of the earliest and largest of these is the Portable Antiquities Scheme (PAS) [103] started in 1997 which now covers all of England and Wales and includes more than one million individual object finds.

2.1.1 FAIR Principles and Linked Data

The FAIR[3] Principles [161, 85] are a set of guidelines that are meant to make scholarly assets more usable to machines and therefore also more useful to humans. There are four foundational FAIR principles that are supplemented with 15 guiding principles. These principles refer to data (or any digital object), metadata, and infrastructure. They are intended as high-level concepts that are not meant to be tied to any specific technical implementation. The foundational principles of FAIR are as follows.

- Findability: Digital resources should be easy to find for humans and computers. The data should be assigned persistent globally unique identifiers and described with rich metadata.

- Accessibility: protocols for retrieving digital resources should be made clear. Data should be accessible through standardized communication protocols.

- Interoperability: Data should be integrated with other data and interoperable with tools and workflows. The vocabularies used to describe the data should also comply with the FAIR principles.

- Reusability: Data should be well described, so that it can be easily reused

[2] The FHA guidlines for metal detectorists are available at https://stmuseovirastoprod.blob.core.windows.net/museovirasto/Palvelut_ja_ohjeet/Antiquities_and_metal_detectors_guide_2020_final.pdf.
[3] https://www.go-fair.org/fair-principles/

by both humans and computers. Data should be released with a clear and accessible license.

The FAIR principles deal with questions of interoperability and reuse of research data and have been embraced by a wide range of stakeholders in various fields of science [25]. Using FAIR principles to improve data have also been discussed in the humanities, for example, for historical [12] data and archaeological [100] data.

Linked Data [21, 14] refers to a set of practices for publishing data on the Web in an interoperable way. RDF [30] is a standard for representing information as KGs using *Uniform Resource Identifiers (URIs)*, which often take the format of Uniform Resource Locators (URLs). *Semantic Web*[4] [15, 16, 58] is a concept that includes elements such as structured machine-understandable presentation of information and an interlinked network of linked data graphs. The original vision for the Semantic Web also included elements such as autonomous intelligent agents, which have not been realized. The Semantic Web has received less attention than many other approaches to artificial intelligence, but has potential with increasing interest in issues such as explainability and data provenance [128].

Tim Berners-Lee has proposed [14] four guidelines for publishing data on the Web: **1)** Use URIs that are unique persistent identifiers to name things, **2)** Use HTTP URIs so that people can access the names, **3)** Use standards like RDF and SPARQL Query Language to provide descriptions of URIs, **4)** Include links to other URIs to help finding new information

Berners-Lee has also proposed the so-called 5-star model[5] for opening data. The idea is that data can be rated with more stars when it is opened in a more open and machine-actionable way:

⋆ Make data available on the Web under an open license.
⋆⋆ Make the data available in structured format.
⋆⋆⋆ Make the data available in an open non-proprietary open format.
⋆⋆⋆⋆ Use URIs to denote things in the data.
⋆⋆⋆⋆⋆ Link the data to other data for context.

To address questions regarding data quality, Hyvönen et al. have proposed additions to the 5-star model, first a 7-star model [64] and later extending to the 8-star model [79]. In this concept a sixth star is awarded for additionally providing a schema for the data, a seventh star is given for providing validation that the data are formatted correctly according to the schema, and finally, the eight star is awarded for giving explanation of how the data corresponds to the facts and of the possible factual errors.

[4] `https://www.w3.org/standards/semanticweb/`

[5] Information and resources about the 5-star model are available at `https://5stardata.info/en/`.

LDF.fi[6] [64] is a service based on the principles of the star model for data publishing. It makes publishing new RDF data sets easier by automating publication of schemas and other documentation, and providing automatically generated human-readable landing pages for URIs. The KGs related to the work in this thesis are published using the LDF.fi system.

The LD principles and the FAIR principles have a clear overlap, but also differences [55]. A big philosophical difference is that the LD principles are more focused on the openness of the data, while the FAIR principles do not explicitly require the data to be fully open.

The CH sector goals of sharing resources and advancement of human knowledge are naturally compatible with LD principles, and there has been interest in using LD in the CH field. Although there have been numerous projects, the adoption of LD in the CH field is still in the early stages. Potential barriers to adopting LD in the CH field include the lack of financial and human resources and the lack of technical infrastructure [32].

2.1.2 Cultural Heritage Ontologies

An *ontology* is a specification of a conceptualization for a certain domain, providing semantic vocabulary for representing the domain [47]. Ontologies are also often considered to be shared between multiple parties and to be defined in a machine-readable manner [48]. The term *data model* can sometimes be used for the specification of classes and relations for purposes of a specific application [141]. In the field of library and information sciences, the term *Knowledge Organization Systems* (KOS) [163, 18] is often used to refer to all kinds of schemes aimed at organizing information. Ontologies can be seen as a special form of KOS that represents complex relations between objects. Ontologies can be described using, for example, the RDF Schema[7] or the Web Ontology Language (OWL)[8] which can facilitate complex logical reasoning.

Ontologies that focus on categorizing objects, instead of describing them in more complicated ways, and provide hierarchical structures for the concepts, can be called *lightweight ontologies* [44]. A commonly used vocabulary that can be used to express lightweight ontologies is *Simple Knowledge Organization System (SKOS)*[9] [112, 8]. For example, Finnish Finto ontologies use the SKOS vocabulary. Such hierarchies can be very useful in helping with information retrieval tasks; however, problems may arise if a user understands the hierarchy in a different way than the creator of the hierarchy [149]. Sometimes it can be beneficial to create separate *annotation ontologies* that are used to describe the data, and *facet ontologies* that are created to be more intuitive to end users [144]. When vocabularies are just a collection of controlled terms that are used in data,

[6] `https://www.ldf.fi/`

[7] `https://www.w3.org/TR/rdf-schema/`

[8] `https://www.w3.org/OWL/`

[9] `https://www.w3.org/2009/08/skos-reference/skos.html`

they can be called *controlled vocabularies*.

CH domain can be characterized by its interdisciplinary nature and the need to integrate information from a wide variety of sources [35]. Ontologies can be a useful tool for addressing these integration issues. Ontologies have also been used in CH applications, for example, to provide semantic recommendations with explanations, to construct facet hierarchies, and for association discovery [59]. Different types of ontologies can include general concept ontologies, actor ontologies, place ontologies, time ontologies, event ontologies, and domain terminologies. A challenge for the CH field is that because it covers such a wide range of human life, it can require a wide variety of different ontologies.

Numerous ontologies have been developed for CH data. The CIDOC Conceptual Reference Model (CRM)[10] [34] is perhaps the most widely used ontology in the CH field. It is a high-level event-based ontology created to allow integration of CH data. Using events as the basis of the ontology can be beneficial in harmonizing heterogeneous CH data. A potential drawback is that there can be added complexity when information needs to be expressed through events. CIDOC CRM is used in a variety of CH applications, for example in [121] to integrate content from the Ephesus Museum and the British Museum, in [22] to integrate metadata of medieval manuscripts from various institutes, and for war history in [113, 95]. CIDOC CRM has also been used in the archaeology domain. It has, for example, been used as a basis for creating the ARIADNE reference model for integrating European arcaheological data [36], and in integrating excavation data in England [20]. CIDOC CRM and Bio CRM [150], an extension of CIDOC CRM for modeling biographical data, have been the basis for data models in BiographySampo and InTaVia KGs that have been used as sources in parts of this thesis.

Europeana[11] [84, 24, 27] is a major project that aggregates data from thousands of CH institutions. It is funded by the European Union to promote the digitization of European culture. The Europeana project developed the Europeana Data Model [83, 139] (EDM) to help harmonize the data provided to it from diverse sources. EDM has been aligned to the CIDOC CRM [27]. This will make conversion from CIDOC CRM to Europeana, or vice versa, relatively easy. However, some European institutions use EDM instead of CIDOC CRM because it can be simpler to implement and makes it easier to add the data of an institution to Europeana [134].

Dublin Core[12] is a widely used metadata schema in the CH domain [87, 4]. In contrast to the event-based model used in CIDOC CRM, Dublin Core uses an object-based or document-based approach, where it defines a set of common properties for describing documents or other objects.

Finland has an existing Semantic Web infrastructure, significant parts of which were originally developed in the FinnONTO project between 2003 and

[10] `http://cidoc-crm.org`

[11] `https://www.europeana.eu/`

[12] `https://www.dublincore.org/specifications/dublin-core/`

2012, where the CH domain has played a significant role from the beginning [80, 152]. Many Finnish lightweight ontologies and thesauri are hosted in the Finto[13] [143] service maintained by the National Library of Finland. For CH domain central Finnish lightweight ontology is the MAO/TAO - Ontology for Museum Domain and Applied Arts[14]. It is designed to help index the material that is related to the museum domain and is maintained by the Finnish Heritage Agency. Finnish place ontology YSO places[15], which is maintained by the National Library of Finland, is also relevant to Finnish CH data. MAO/TAO and YSO places have been used in multiple projects related to this thesis.

Internationally an important linked data resource is Wikidata [39] which is based on Wikipedia. Wikidata has been used in many ways to enrich the data in projects related to this thesis. The vocabularies of the Getty Research Institute[16] [53, 7] have also been important for this thesis. For international harmonization of archaeological objects, Getty Art & Architecture Thesaurus (AAT)[17] is an important vocabulary. In the FindSampo project Finnish object terms were aligned with AAT and through that to FISH[18] Archaeological Sciences Thesaurus used by the British Museum. Union List of Artist Names (ULAN) includes rich relations between artists, such as teacher-student and patron relations. These relations have been used in a case study in this thesis.

The international Nomisma.org[19] [56, 46] ontology provides classes, properties, and detailed vocabularies for numismatic objects. However, those are incomplete and tend to focus on the classical era at the moment. Although Nomisma.org is currently mainly adopted by the classical era numistatic community, the project is working on adding concepts from the medieval era and beyond. In DigiNUMA project, the vocabularies related to Finnish numismatic finds were aligned to the Nomisma concepts.

Integrating heterogeneous CH data can be challenging. For example, the concepts in CIDOC CRM are abstract and complex in such a way that even experts can find it difficult to interpret them consistently [116]. Even applying the less complicated Dublin Core vocabulary was found to have challenges in a survey of metadata professionals [124].

2.1.3 Sampo Model

The Sampo Model [66] is an informal collection of principles for publishing LOD and semantic portals. The six principles of the Sampo Model as outlined by Eero Hyvönen [66] in 2023 are:

[13] https://www.kansalliskirjasto.fi/en/services/finto
[14] https://finto.fi/maotao/en/
[15] https://finto.fi/yso-paikat/en/
[16] https://www.getty.edu/research/tools/vocabularies/
[17] https://www.getty.edu/research/tools/vocabularies/aat/
[18] https://heritage-standards.org.uk/fish-vocabularies/#archaeological-objects-thesaurus.
[19] https://nomisma.org/

- P1 Support collaborative data creation and publishing.
- P2 Use a shared open ontology infrastructure.
- P3 Make clear distinction between the LOD service and the user interface.
- P4 Provide multiple perspectives to the same data.
- P5 Standardize portal usage by a simple filter-analyze two-step cycle.
- P6 Support data analysis and knowledge discovery in addition to data exploration.

P1, P2, and P3 concern publishing data, while P4, P5, and P6 concern the construction of web applications. Principles P1 and P2 are related to general Semantic Web standards. P3 is about the separation of data and the web application. The data is published on a SPARQL endpoint, and web applications built on top of the data access the data only through SPARQL queries. This imposes special requirements for the design of web applications. P4 is about application design and viewing the data from different points of view. For example, a user could be offered separate perspectives to search the persons and places in the data, which can offer new points of view to the same data. P5 is about designing the user interface around the two-step cycle for research: First, the target group that the user is interested in is selected, perhaps using faceted search based on ontologies, and then the target group can be visualized and examined in different ways. P6 is about offering better ways to analyze the data in user interfaces.

The Sampo model has evolved on the basis of practical experiences gained from developing the Sampo series of linked data services and semantic portals since 2002. Currently, there are over 20[20] different LOD services and portals in the Sampo series, most of which are related to CH. The earliest services were Museum Finland[21] [62] (2002, collections of Finnish museums), CultureSampo[22] [61] (2008, Finnish culture), and BookSampo[23] (2011-, fiction literature). Other examples include WarSampo[24] [71] (2015-2019, Second World War in Finland) and BiographySampo[25] [73] (2019, biographies). After 2019 the

[20] A list of Sampo services is available at: `https://seco.cs.aalto.fi/applications/sampo/`.
[21] MuseumFinland portal is available online at `https://www.museosuomi.fi/` as of 27. December 2024, but not actively maintained.
[22] CultureSampo portal is available online at `https://www.kulttuurisampo.fi/` as of 27. December 2024, but not actively maintained.
[23] BookSampo portal is available online at `https://www.kirjasampo.fi/` [107] as of 27. December 2024, and is actively maintained by the Finnish public libraries.
[24] WarSampo portal is available online at `https://www.sotasampo.fi/`.
[25] BiographySampo portal is available online at `https://biografiasampo.fi/`.

web applications for Sampo series have generally been done based on the Sampo-UI framework. These include NameSampo[26] [82] (2019, Finnish place names), Mapping Manuscript Migrations (MMM)[27] [60] (2019, medieval manuscripts), and ParliamentSampo[28] [63] (2022, parliamentary speeches).

2.2 Cultural Heritage Applications

Doer [35] identifies four main functions of the CH information systems: collection management, conservation, research, and presentation. Although there exist highly specialized systems for each of these areas, there is some overlap. The applications presented in this thesis aim to be usable for presentation, research, and even collection management.

2.2.1 Usability and Users

Usability is an important consideration when designing information systems, such as web applications. One classic way to define usability is through five usability attributes: *learnability*, *efficiency*, *memorability*, *errors*, and *satisfaction* [114]. Learnability refers to how easy it is to learn to use a system. Efficiency as a usability attribute refers to how productively a user can use a system once it's use has been learned. Memorability refers to how much effort is required to re-learn to use a system after a period of not using it, and the concept of errors refers to how well users can avoid errors when using a system. Finally, satisfaction refers to the satisfaction that users feel when using a system.

Although user satisfaction is important, especially when targeting the general public, this has not been the main focus of this thesis. The work in this thesis is mainly concerned with efficiency and learnability.

Potential LD users can be divided into two main categories: lay users, who do not have a deep technical understanding of LD, and technically skilled users who do have a deep understanding of LD [31]. An important subcategory of lay users are domain experts. Humanities researchers can be an example of such users. These people have domain expertise, for example, in archaeology, but often only superficial understanding of LD.

2.2.2 Faceted Search

Existing LD interfaces use multiple different interaction paradigms that have their own strengths [13]. The paradigm on which this thesis mainly focuses is called faceted search [149, 120]. Faceted search is based on exploring the data with different structured attributes of the searched resource. These attributes

[26] NameSampo portal is available online at `https://nimisampo.fi/en/`.
[27] MMM portal is available online at `https://mappingmanuscriptmigrations.org`.
[28] ParliamentSampo portal is available online at `https://parlamenttisampo.fi/`.

can be organized into different facets. For example, archaeological objects might be searched with facets based on material, object type, and period. The selections on the facets can be made in any order. The hit counts for each facet selection are also shown and updated after each new selection is made.

Faceted search is a powerful tool in part because it allows for open-ended search where users do not necessarily know exactly what they are searching beforehand, but issues such as a large number of different facets can cause usability problems [115]. Faceted search can also be computationally demanding, as the hit counts need to be updated after each selection is made. An important factor influencing performance is the number of entities searched.

Faceted search is particularly suitable for *exploratory search* [110, 99]. In exploratory search, the user is not only interested in looking up a specific file or person. Instead, exploratory search is about learning and investigating data. The incremental nature of faceted search facilitates exploration and the hit counts of facets can help in discovering patterns in the data.

Faceted search is commonly used in LD applications, especially when they are aimed at non-technical users [1]. Possible ways of visualizing data include lists, tables, maps, time-lines, and graphs. Combining faceted search and visualizations with CH LOD have been an important theme in the Sampo series of portals. Four example portals are discussed in [102]. These examples show how faceted search and visualizations can be implemented on top of SPARQL endpoints. One relevant example for this thesis is the WarSampo portal, and especially it's war victims perspective. It can be compared to WarVictimSampo, as they both deal with Finnish war victims. WarSampo provides nine different application persceptives to the data. These include perspectives for events, persons, war victims, and prisoners. Many of these, such as the war victims and prisoners perspectives, have essentially their own datasets and data models, allowing search and visualization of a specific dataset. WarSampo also harmonizes the data using a model based on CIDOC CRM. This event-based harmonized data is used on the events and persons application perspectives. The war victims perspective offers a faceted search and various ways to visualize the result set.

2.2.3 Relational Search

Relational search (also known as association search) [136, 28] is a concept that deals with finding potentially interesting connections between entities in a KG. In relational search, the task usually is to find interesting semantic associations between two resources. The concept of relational search has been applied, for example, to find potential terrorists [136] and for medical research [153]. The approaches can differ in various ways, including the KG that is searched, the formulation of queries, the methods for finding connections, and the representation of the results [28].

Generally, the main challenge that these systems have is how to select and rank the interesting connecting paths and how to weed out the uninteresting connections. An early work dealing with ranking associations is SemRank [5]. Later work includes, for example, WiSP [147] which finds and ranks several paths with a between two resources in Wikidata KG based on relevance. Various methods have been proposed for ranking connections that can generally be divided into two main categories: data-centric methods, where the ranking is based on properties of the graph such as size, frequency, centrality, informativeness, and specificity of a connection, and user-centric methods, where connections are ranked based on some criteria defined by a user [28, 19].

Explaining relations has been explored in multiple previous studies. In RelFinder [101, 105] the user selects resources, and the result is a graph visualization showing how the queried resources are related to each other. Similarly in [127] a tool for explaining paths between given resources in a KG called RECAP is presented, where the explanations are visualized as a graph showing the labels of resources. A recent example of a relational search application is WoolNet[29] [49]. In [154] human-readable sentences that explain the relations in KG are selected from Wikipedia texts and ranked.

Faceted search has also been applied for searching relations. Some applications presented in literature, e.g., RelFinder [101, 105] and Explass [29], allow filtering relations between entities with facets. However, the user has to first select two entities, and faceted search can only be used to filter relations between the preselected entities.

CultureSampo [61, 109, 77] includes an application in which connections between two selected artists can be searched. This is an example of applying the concept to the CH field. The system is based on data extracted from the Getty ULAN KG.

2.3 Tools for Creating Linked Data Applications

Tools that aim to simplify the creation of semantic portals or web applications have been part of the Sampo model from the very early stages. OntoViews [108] was an early attempt to create a tool for creating web portals for linked data with faceted search functionality. The MuseumFinland [62] and CultureSampo portals were created using it.

SPARQL Faceter[30] [93] is a JavaScript component for building user interfaces with faceted search on top of SPARQL endpoints. It is implemented using AngularJS[31], and is designed so that the application can be implemented with a minimal back-end server. SPARQL Faceter is used, for example, in the WarSampo portal. The Web application for Publication I was implemented using

[29] A demo is available online at: `https://woolnet.dcc.uchile.cl/`. The link is tested to work on November 5th 2024.

[30] Source code for the SPARQL Faceter tool is available at GitHub: `https://github.com/SemanticComputing/angular-semantic-faceted-search`.

[31] `https://angularjs.org/`

SPARQL Faceter. SPARQL Faceter is implemented so that the faceted search is directed to instances of a certain class as the basic search space, and each facet represents a certain property path from the instance of the class to facet values. This approach to faceted search search has also been adopted in Sampo-UI.

A drawback of the SPARQL based approach is that implementing faceted search using SPARQL queries can be computationally demanding and, therefore, slow for users [66]. SPARQL queries are not the only possible way of implementing search in KGs. This is why in many cases various server side solutions such as ElasticSearch[32], and Solr[33] are used to implement the search. For example, the ARIADNE project has used ElasticSearch in its archaeology portal [111].

The SPARQL-based approach for implementing faceted search is in accordance with the principle of the Sampo model, which expects that web applications built on top of the data access the data only through SPARQL queries. This simplifies the architecture of the applications. Designing tools with this principle also makes it possible to use them, at least in principle, with any SPARQL triplestore.

[32] https://www.elastic.co/elasticsearch

[33] https://solr.apache.org/guide/6_6/faceting.html

3. Results

This chapter presents answers to the research questions of the thesis. The results as a whole and the lessons learned are further reflected in Chapter 4.

3.1 Publishing Cultural Heritage Data (RQ1)

The research question 1 concerns the modeling of CH data for publication. The results and contributions of this thesis include new methods and multiple datasets from Finnish heritage institutions that were converted to LD and enriched in cooperation with heritage agencies and humanities researchers in the course of multiple projects. While CH is a diverse field, the following section focuses on war history and archaeological finds, which are the main focus of this thesis.

3.1.1 State of the Art

The state of the art in publishing CH data, and research data in general, is to publish the data so that it conforms to the FAIR principles. Publishing the data in RDF format as LOD is a technical way to address this. State-of-the-art systems often use LD and ontologies to describe data. However, the complexity of ontologies is often limited because complicated constructs can be difficult for human users to master, and the concepts used in CH data can often be vague or uncertain, causing logical reasoning to be complicated [59]. One data model commonly used when publishing LOD is the event-based CIDOC CRM model, which can be useful for interoperability between heterogeneous datasets. Another example is the document-based Dublin Core model, which offers a direct and simple table-like model for representing CH metadata.

CIDOC CRM is used in many state-of-the-art systems. It is usually extended in various ways to fit the needs of the project. The ARIADNEplus project [9] uses a specialized model called the ARIADNE ontology based on CIDOC CRM to harmonize and publish archeological data from around Europe. A model that has elements from CIDOC CRM is also used, for example, in the MEDEA [33] project that publishes archaeological citizen finds from Flanders.

CIDOC CRM aims to provide interoperability with heterogeneous data by providing a generic model that can be used with different kinds of CH data. Another option is to use a specialized ontology. Nomisma.org [56, 46] is an example of a specialized ontology focused on the field of numismatics. It is an international ontology that aims to provide all the necessary classes and properties required to semantically represent numismatic objects. Nomisma.org is also working to provide all the necessary entities, such as mints and rulers, that the field of numismatic requires. Considering the vast history of numismatics, this requires a lot of work.

WarSampo [94] is an example of state of the art in publishing war history data. It publishes data on the Second World War in Finland as LOD. It uses both an event-based model and a table-like model in different contexts. War victims are represented in WarSampo using a simple table-like model, similar to Dublin Core. There is also a CIDOC CRM based model that is generated from the simple RDF. A simple model is used in War Victims KG of WarSampo, while the CIDOC CRM based ontology is used to harmonize data from different KGs in WarSampo. LOD has been used to harmonize data on the events of the First World War in WW1LOD [113], which uses a model based primarily on CIDOC CRM, but uses a simpler model that is not based on events to describe relationships between organizations and people, because the CIDOC CRM model was not considered intuitive for users in those cases. LOD has also been used to publish data about the Spanish Civil War. An example of this is the Ontophoto [133] ontology that was created to disseminate data on historical photos of the Spanish Civil War.

In Sweden, the Swedish Open Cultural Heritage (SOCH) [140] aggregates and publishes CH data from various stakeholders, such as museums, in LD format. The data is mapped to its own common RDF-based model[1] and made available through API. SOCH allows Swedish CH actors to publish their data as LOD, without needing to commit too much of their limited resources. The SOCH data model is focused on entities, or records, and assigning metadata attributes to them. Some information loss can be unavoidable when aggregating data to this kind of common model, as the model needs to remain abstract enough to accommodate different kinds of data.

3.1.2 Improving on the State of the Art

Publication II presents the work of creating the WarVictimSampo KG of Finnish war victims from 1914 to 1922 by converting the original database to LD and updating it with new information. The war victims in the database are from the First World War, the so-called Kindred Nation wars, and most importantly the Finnish Civil War of 1918. Original data was collected from various sources

[1] Documentation about the model that SOCH uses is available in Swedish at https://www.raa.se/hitta-information/k-samsok/att-anvanda-k-samsok/protokoll-och-parametrar/.

between 1998 and 2003 in a large project funded by the Finnish government. The database was an SQL relational database that could only be accessed by the public through a web application. The database had errors and many victims were missing.

In the WarVictimSampo project, the data was updated with new victims, errors were fixed, and the data was converted to RDF with global URI identifiers and partially linked to external resources. More than 1600 new records were added to the database. The new records were added in cooperation with experts from the National Archives. The records are mainly about victims from the "Red" side of the Finnish Civil War and were not available when the original database was collected. Numerous corrections were also made to individual records. This clarifies the large historical picture of the Finnish Civil War and can be important to relatives of the victims. Having the data in RDF format conforms to the FAIR principles and makes the data easier to use in any future research projects.

Due to the nature of the Civil War, many victims did not die in battles but were executed or died from diseases in prison camps. The records kept were of varying quality and were sometimes destroyed on purpose. This means that the data is often uncertain and that different sources can give different, even opposite, information. Modeling the provenance of individual pieces of information is very important in this case. In contrast to the WarSampo war victims data that does not include separate sources for individual pieces of information, WarVictimSampo introduces a model that allows sources to be explicitly expressed. The model is based on individual information resources, which is based on the idea of RDF reification. This allows expressing the source and other information, such as preserving the sometimes erroneous codes used in the original database, for each individual piece of information. WarVictimSampo also includes a separate model with simpler table-like presentation of the most important pieces of information, mainly those used in facets of the portal. This model is similar to Dublin Core or the model used for the war victims perspective of WarSampo, and there each piece of information is connected to a death record with a single triple, which makes them less complicated and faster to use, for example, in a web portal when querying a KG using SPARQL when updating the facet values. However, the more complex model needs to be used when accessing the provenance of information.

Publication IV presents the work of creating a FAIR LOD dataset of cataloged archaeological object finds reported by the public from 2015 to 2020. This means that in contrast to CoinSampo which includes all the reported numismatic object finds, only important objects taken into collections are included in FindSampo. In this project, existing records of objects were converted to LOD with global URI identifiers. The original records were created by experts from the Finnish Heritage Agency, but, for example, the types of object were not recorded using a clear controlled vocabulary. Working together with archaeology experts, a mapping was created from the used terms of object types to the corresponding concepts in the MAO/TAO ontology. A new period ontology was created to represent periods of Finnish history and prehistory with machine-readable dates. Although the MAO/TAO ontology provides a simple hierarchy, it was deemed too difficult to use for many use cases. This is why a new facet ontology was created for the object types. This facet ontology is mapped to object type concepts from MAO/TAO and provides a hierarchy that is more intuitive to use. The object type concepts were also mapped to the Getty AAT thesaurus for international compatibility. The links to the MAO/TAO and AAT vocabularies were made by experts working on tabular files that were then converted to RDF. A special script was created to allow for creating a hierarchy for facet ontology using a tabular file.

To represent classes and properties in the FindSampo KG, the FindSampo Core ontology was created. This is a simple ontology inspired by the Dublin Core where objects are the focus, and most information is connected to the object with properties. This is in contrast to event-based ontology like the CIDOC CRM that provides more compatibility to wider range of data but can be unnecessarily complex if only archaeological object finds are modeled. For example, in this table-like model the production period is expressed with a single triple with the object as the subject and the period as the predicate. An event-based model like CIDOC CRM would require a separate production event and therefore more triples complicating the KG. The property chain from object to period would also be longer. The event-based model would be more general purpose and make it easier to use the data with diverse other data. The object-based approach used here is a compromise that makes the creation of KG easier and also makes the creation of a faceted search application easier. Although internationally data on citizen finds can vary slightly, the core properties tend to be the same. These include object type, material, period, and findspot. This means that the FindSampo Core ontology could be used to harmonize the most important elements of international data that are focused mainly on object finds.

Publication VII presents work of creating a FAIR LOD dataset of all numismatic citizen finds reported by the public in Finland between 2013 and 2023. The original data was collected by the curator of the coin finds in the Finnish National Museum as an excel sheet. Coins are an interesting subtype of citizen finds. They are usually the most common object type, they have many properties that can be represented that most objects do not have, such as a mint and authority (issuer), and they can also usually be dated much more exactly than most objects.

To represent the coin finds, a new data model was created that is based on the FindSampo Core ontology but expanded with properties specific to coin finds. Although Nomisma.org aims to provide classes and properties for numismatic finds, there were some properties that would not fit well. Nomisma.org also relies on representing most of the information as part of a “coin type” instead of the object itself, which can be useful, but it would have been difficult to create the required instances for the coin types, as they largely do not currently exist in Nomisma.org.

During the project, it was found that most of the necessary concepts, such as mints and rulers required to represent coin finds from Finland, did not exist in the Nomisma.org vocabularies. For example, out of the 193 authorities (i.e. rulers) in the CoinSampo only 34 had a corresponding resource in Nomisma.org, and only 146 individual coins had an authority that has a corresponding concept in Nomisma. Only about a percent of all the reported coin finds from Finland could be described using the existing mint and authority resources of Nomisma.org. This is because Nomisma.org is mostly limited to the early era coins that form only a very small part of the Finnish coin finds. Because of this, new vocabularies were created to represent the coin finds from Finland. The initial vocabularies were extracted from the data as tabular files and then cleaned and mapped to the corresponding concepts in Wikidata and the YSO ontology. The tabular files were then used to create RDF vocabularies to represent the data. The vocabularies were then enriched from Wikidata and YSO with relevant information. Translations into English were added to the concepts partially from YSO and Wikidata.

To represent uncertainty that is inherent to many coin finds that often are not in perfect condition, separate concepts that represent uncertain concepts were created. An example of this is a minting place that is either Sweden or Denmark. Using such uncertain concepts made it possible to keep the data model simple, as there was no need to model uncertainty beyond having a separate class for uncertain concepts.

Publication I and Publication VIII present a model for publishing CH data in a way that focuses on representing relations between entities such as persons and places as individual entities themselves. These can be created from existing CH KGs using SPARQL CONSTRUCT shapes that correspond to predetermined types on interesting connections. Natural language explanations can be created for relation entities with simple forms in SPARQL CONSTRUCT queries, where the labels of the entities are inserted in a form that corresponds to the type of the relation, for example “X was the teacher of Y”.

Creating a new KG of interesting relations with human readable explanations made it easier to create web applications for searching and analyzing relations. For example, Figure 3.1 shows the web application[2] presented in Publication VIII where the search is refined with facets to find relations between Austrian and Finnish persons where the Austrian person is the teacher. For example, according to the data, Austrian composer Karl Goldmark taught the Finnish composer Jean Sibelius.

Publication I focuses only on relations between people and places using the BiographySampo KG of biographical data of important Finnish persons as the source data. Publication VIII applies the method on person-to-person relations in InTaVia KG of biographies from multiple European countries, and Getty

[2]Web application for searching relation in InTaVia KG can be tested online at: `https://intaviasampo.demo.seco.cs.aalto.fi/`.

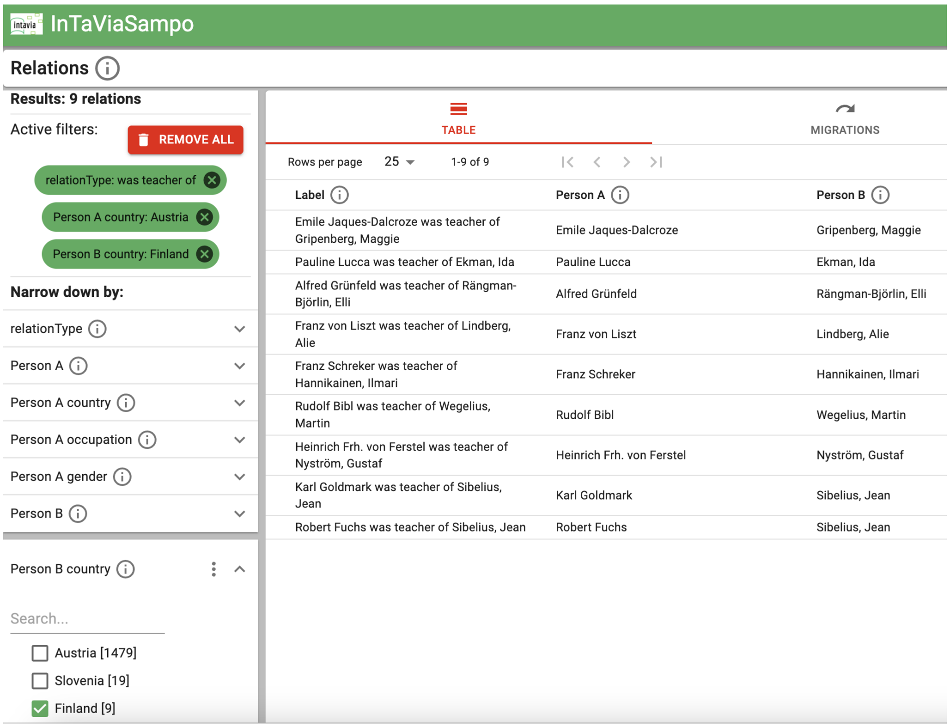

Figure 3.1. An example of searching relations: Connections between Austrian and Finnish persons where an Austrian has been a teacher of a Finnish person.

ULAN art history KG. One drawback of the method is the work required to craft the SPARQL CONSTRUCT queries that need to be done separately for each type of interesting connection, and usually for each new KG. The number of individual relation instances can also grow large for practical applications. It was noted that the number of interesting connections between people and places tended to stay relatively low, while the number of connections between persons could grow very large when dealing with second-degree connections, such as shared teacher relations.

In Publication VIII it is observed that modeling relation as directed connections offers better possibilities for implementing search, as semantics are better preserved. For example, when modeling teacher-student relation as directed relation, the semantics of which person is the teacher and which one is the student is known and this knowledge can be used for searching and visualizing the data. In Figure 3.1 both endpoints, here persons, of the connection have their own separate facets that can be used to filter the relations. The user has used the facets to select "was teacher of" as the Austria as the country of "Person A" and Finland as the country of "Person B". This means that the results are only relations where the person from the Austrian biographical collection is the teacher, and the person from the Finnish collection of biographies is the student. The hits in "Person B country" also show that while there are 9 biographies with Finland as the country of origin, there are 19 persons from the Slovenian collection. A search with roles reversed would show that there are no cases where a person from the Finnish collection would have been a teacher of a person in Austrian collection. Defining search in such a way would not be possible, or at least would be harder, if there were not semantic information of the direction of the connection.

Modeling the relations as undirected connections makes implementing some search functions and visualizations more difficult, as some semantics are lost, but the number of relation entities is smaller, so the computational requirements for memory or faceted search are not as large. This means that if the desired search or visualization functions do not require semantics of relations direction to be preserved, omitting it and using undirected relations can be beneficial.

3.2 Application Features (RQ2)

Web applications can offer a useful access point to CH data for both the general public and researchers. In legacy systems, the web application can be the only access point that a normal user has to the database, and its users often equate the web application used to access a database and the database itself. Even when the data can be accessed directly through, for example, SPARQL queries, the web application can be a vital tool to search and visualize the data for the general public and also researchers. This section presents how Finnish CH web services and the international state of the art have been improved in the case studies of this thesis.

3.2.1 State of the Art

Multiple ways have been implemented in state-of-the-art systems to improve search and browsing [1, 131]. Faceted search has been used for searching and exploring CH KGs, especially in applications aimed at users that are not technically aligned. Semantic auto-completion can also be used to help search. For example, the name of a searched person can be automatically completed, so that users do not necessarily need to fully remember the name of the person they are interested in. Semantic recommendations can be provided to users to link the items they are viewing to other similar items. Some systems also aim to provide users serendipitous results in addition to the expected ones to provide inspiration. State-of-the-art systems also often offer the option of visualizing information about the selected group of entities or an individual entity in different ways, such as charts and maps. Some relevant examples from the fields of war history and archaeology are given below.

PAS finds database[3] of the British Museum can be searched using facets, and the finds can be visualized on a map. PAS is perhaps the earliest and largest example of publishing archaeological citizen finds. In the Netherlands, the Portable Antiquities of the Netherlands (PAN) [26] has an application based

[3] https://finds.org.uk/

on LD[4] that records and documents archaeological finds made by the public. It offers faceted search functions and map visualizations. The MEDEA [33] project from Flanders also has an application[5] based on LD for publishing citizen finds. It offers users the ability to filter results and some basic map visualizations. ARIADNEplus portal[6] [132] opens data from the ARIADNEplus project that aggregates archaeological research data, including object finds, from multiple European countries. The portal offers a faceted search function based on ElasticSearch and a table, heat map, and timeline visualization options. The American Numismatic Society has multiple portals for publishing data on the classical era numismatics based on LD and the Nomisma.org ontology. These include Seleucid Coins Online (SCO)[7] and Online Coins of the Roman Empire (OCRE)[8]. The results can be browsed using filters, but a full faceted search with hit counts is not offered. The results can be viewed as a list and visualized on a map. These services are based on LD, using an XML database, and are built using the Open Source Numishare[9] tool maintained by the American Numismatic Society.

WarSampo casualties perspective allows for faceted search of Finnish casualties of WW2. In addition to the table view of individual results, there is a visualization for age distribution and interactive bar chart visualizations for properties like occupation and gender. Another Finnish resource, in this case based on a relational database, is Sotapolku[10] which was opened in 2016 [92]. It is a military history portal for crowdsourcing the history of Finnish soldiers in WW2. It has simple search functions and offers map-based visualizations.

Internationally, there are multiple war history projects based on both LD and non-LD. For example, 1914–1918-online[11] is a portal for publishing articles about WW1. It is based on Semantic MediaWiki[12], and offers simple faceted search. Morts Pour la France[13] [43] is a database of French war victims in the First World War, which has a web application with a faceted search, but no visualization of results beyond a table view.

[4] `https://portable-antiquities.nl/`
[5] `https://vondsten.be/`
[6] `https://ariadne-infrastructure.eu/portal/`
[7] `https://numismatics.org/sco/`
[8] `https://numismatics.org/ocre/`
[9] `https://github.com/ewg118/numishare`.
[10] `http://sotapolku.fi`
[11] `http://www.1914-1918-online.net`
[12] `https://www.semantic-mediawiki.org/wiki/Semantic_MediaWiki`
[13] `https://www.memoiredeshommes.sga.defense.gouv.fr/conflits-et-operations-2/premiere-guerre-mondiale/morts-pour-la-france-de-la-premiere-guerre-mondiale/faire-une-recherche`.

3.2.2 Improving on the State of the Art

Publication II presents the WarVictimSampo portal. The portal was created to help users access WarVictimSampo KG. The portal allows users to search for and visualize war victims and battles in the KG in various ways. The user can search for victims with various facets and then access the individual records or visualize the results sets with various charts and maps. The portal is aimed at researchers and the general public. It replaced an existing web application for the war victims database of the National Archives of Finland.

The WarVictimSampo portal offers faceted search and more search functions in comparison to the old application. The old application also lacks visualizations and some data related to the victims was not accessible through it. Compared to the WarSampo casualties perspective, WarVictimSampo offers the ability to visualize the war casualties of a selected set on a map. There is also a feature that automatically calculates the average and median values of the age distribution in addition to showing the distribution as a chart. This is a technically simple feature, but makes it easier to see differences between groups, while WarSampo concentrates on getting information about individual victims. For example, in Figure 3.2 a user has used faceted search in the WarVictimSampo application[14] to select "reds" ("punainen") and registered province "Turun ja Porin lääni" when searching for war victims. This limits the result set to all the victims of Finnish Civil War who were on the Red side and live in the Turku province. A line chart visualization for ages of victims has been selected and shows that the median age of this group is relatively low compared to other victims from the Red side of the war.

Publication IV presents the FindSampo web application which allows users to search and visualize cataloged object finds from Finland. The user can search the objects with various facets and then access the individual records or visualize the result sets with various charts and maps. Facet ontologies for object types, materials, and periods offer hierarchies that make it easier to search for specific objects and select interesting groups to compare with visualizations. In comparison to existing state-of-the-art web applications, such as the ARIADNEplus portal, the PAS database, and Seleucid Coins Online, FindSampo offers a wider range of visualizations and hierarchical facets to make it easier to select and analyze groups of objects. For example, Fig. 3.3 is a screenshot of FindSampo where a user has selected all finds with the material silver ("hopea") and selected the "heatmap" visualization showing the relative numbers of coin finds in different places on a map[15], with a higher concentration of finds in red and a lower concentration in yellow. Quite a lot of silver objects

[14] The WarVictimSampo web application can be tested online at: `https://sotasurmat.narc.fi/en/`.
[15] The map in the screenshot is generated by the application using the Mapbox (`https://www.mapbox.com/about/maps`) service with data from OpenStreetMap (`http://www.openstreetmap.org/copyright`). FindSampo also uses map services provided by the National Land Survey of Finland and the FHA.

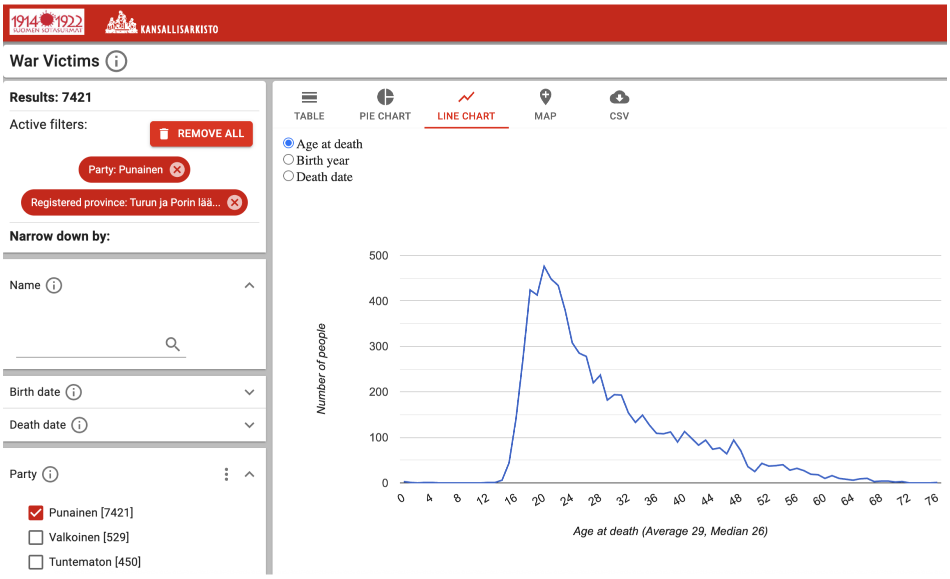

Figure 3.2. An example of using faceted search in WarVictimSampo: The user has selected "reds" ("Punainen") and registration province "Turun ja Porin lääni".

seem to be found around the town of Hämeenlinna.

Publication VII presents the CoinSampo web application, which can be used to access the CoinSampo KG. The user can explore the coins, the authorities (rulers), and the mints in the KG with various facets and then access the individual records or visualize the results sets with various charts and maps. Published in 2023 and being able to build on previous work, such as Publication IV and Publication II, the CoinSampo web application is more advanced and versatile than the WarVictimSampo and FindSampo applications and has more visualization options than, for example, the ARIADNEplus portal. The application demonstrates how Linked Data can be utilized for analysis purposes. For example, the minting places of the coins can be analyzed on a map by using the location data of the minting place. This is possible because the mint entities are connected to corresponding entities in Wikipedia that have coordinate information that is used to enrich the CoinSampo KG. The Finna[16] service has been used to show images of broadly similar coins, by constructing queries to Finna based on properties of the coins in the KG.

Publication V presents higher level ideas related to creating the CoinSampo semantic data service and web application, and generally using such applications for publishing CH data. The paper shows how web applications can be used by users with little or no training to quickly create simple analysis of the data by exploiting faceted search and visualizations. For example, Fig. 3.4 is a screenshot of the CoinSampo[17] application where the user has selected Viking Age from the period facet and Hämeenlinna from the municipality facet.

[16] `https://www.finna.fi/?lng=en-gb`

[17] The CoinSampo web application can be tested online at: `https://coinsampo.ldf.fi/`.

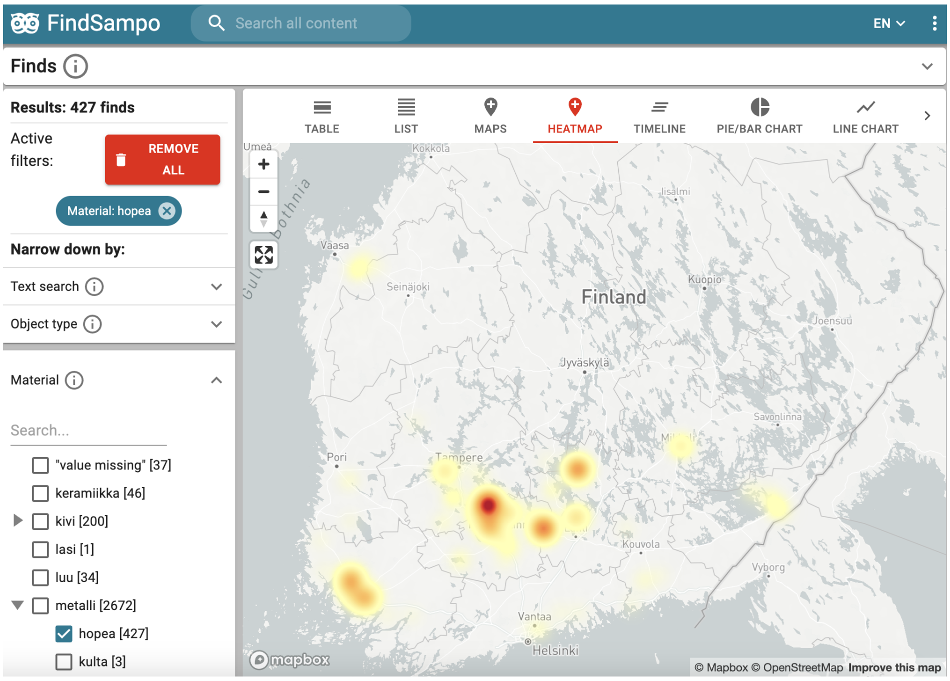

Figure 3.3. A screenshot from the FindSampo application: A heatmap visualization showing object finds from Finland that are made of silver.

Using the "migrations" visualization, the user can see arcs between the minting place of the coin and the find municipality, in this case limited to the town of Hämeenlinna. Such visualizations could be used by researchers, but also, for example, hobbyists or students interested in local history. Such visualization and search functions democratize the data by making it easier for the general public to analyze the data. This visualization also shows certain benefits of LD. The minting spots are visualized using coordinates that are added to the KG from Wikidata.

The paper identifies four distinct user types for application opening data about public archaeological finds: 1) Metal detectorists who find and report the objects as a hobby and who are interested in history and how the objects they find fit into larger picture. 2) Researchers, both academic and non-academic, who want to research historical phenomena. 3) Local historians who are interested in the history of a specific place, such as their hometown. 4) Collection managers who can use facets and visualizations to detect errors in their own data.

Publication I presents a method for using faceted search and web applications to search for relations or associations from CH KGs. The method is based on materializing the interesting connections in a KG as a new KG of interesting relations modeled as individual entities. These relation entities can be filtered using faceted search in a similar way as natural entities such as persons and places can be searched. Publication I applies this approach to BiographySampo KG and the relations between persons and places. In Publication VIII this approach is extended to connections between persons and persons, and case

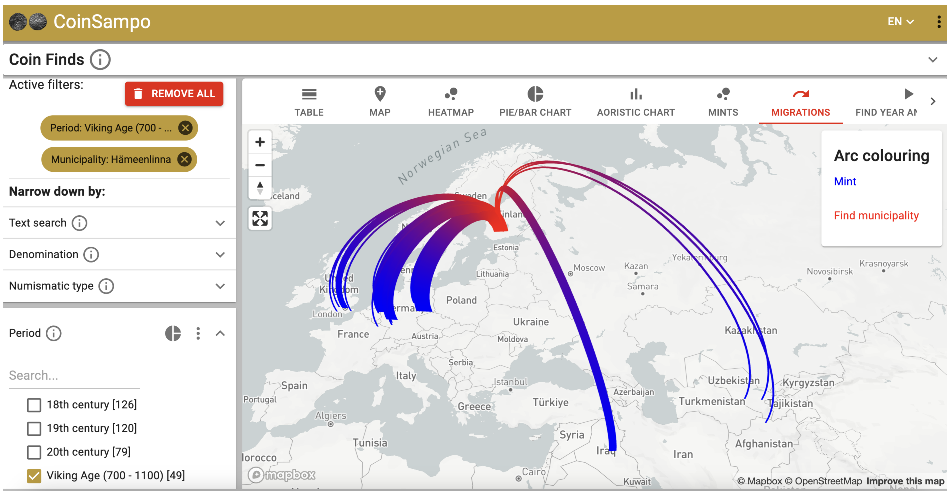

Figure 3.4. A screenshot from the CoinSampo application: A map showing arcs between minting place and find spot.

studies involving the Getty ULAN list of artists, with art history information, and the InTaVia collection of biographical information from multiple European countries.

Applying a faceted search to relations offers a different perspective to the data. In addition to results being relations and their explanations, hit counts of the relations are different from, for example, individual persons. This can offer unique insight into the data. Using a KG of interesting relations makes it convenient to create a faceted search system for relations and to show natural language explanations of each individual connection. Although precalculated relation entities make the performance of the faceted search more computationally efficient compared to calculating relations dynamically, the number of individual relation entities can grow so large that performance of the faceted search becomes slow.

3.3 Creating Semantic Applications (RQ3)

While data in an RDF KG can be accessed using SPARQL queries, in practice, for many user groups and use cases it is more convenient to access the data through a web application. Creating a complex web application with various different ways to search and visualize data, and making the application user-friendly and visually appealing, can be a daunting task that requires technical skills and effort. This research question deals with how the process of developing such applications can be supported.

3.3.1 State of the Art

Linked Data Reactor (LD-R) [90] aims to provide software developers with a starting point for a JavaScript web application that can be configured to view, browse, and edit RDF data. FERASAT [91] is a tool built on top of LD-R, and aims to provide faceted search and visualization tools to facilitate serendipitous knowledge discovery from LD. Metaphactory[18] [50] is a platform for building applications based on knowledge graphs and SPARQL queries. It is an extensive commercial platform with a wide range of features, ranging from managing knowledge graphs to visualizing them and exploring them with faceted search. ResearchSpace[19] [119] is a system designed at the British Museum and is based on the metaphactory platform. It is designed to support the creation and representation of CH data by building Semantic Web applications. It is designed to support LD and especially the CIDOC CRM ontology. Skosmos[20] [143] is a web publication platform for SKOS-based vocabularies and is widely adopted in publishing lightweight ontologies.

Although there are a number of tools that can be used to create applications based on LD, there are limitations in many of them. For example, LD-R[21] and SPARQL Faceter[22] are not currently actively updated. The metaphactory platform is actively updated and offers a wide variety of features but is not fully open source. Skosmos, on the other hand, offers great usability for its specific purpose of publishing lightweight ontologies. However, it has limited functionality and lacks, for example, different visualizations of the results.

A starting point from the point of view of this thesis has been the SPARQL Faceter tool. Multiple portals in the Sampo series have been built using it, including the WarSampo portal, and the demonstrator in Publication I. The benefit of SPARQL Faceter is that it requires only a minimal backend, which simplifies setup. Limitations include facet selections that are limited to one per facet and outdated technology.

3.3.2 Improving on the State of the Art

Publication III presents the Sampo-UI framework for creating web applications with faceted search and visualizations on top of SPARQL endpoints. Sampo-UI is essentially a working application that can be taken as a starting point

[18] Documentation is available at `https://help.metaphacts.com/resource/Help:Documentation`.

[19] GitHub page of the project with documentation is available at: `https://github.com/researchspace/researchspace`.

[20] `https://skosmos.org/`

[21] In September 2025 the latest update of LD-R GitHub repository `https://github.com/ali1k/ld-r` is from 2022.

[22] In September 2025 the latest update of SPARQL Faceter GitHub repository `https://github.com/SemanticComputing/angular-semantic-faceted-search` is from 2018.

when creating applications for RDF data. It makes creating a faceted search application faster and requires less programming experience. In comparison to SPARQL Faceter, Sampo-UI is a full implementation of a web application instead of a library that is only meant to implement faceted search. In comparison to LD-R/FERASAT and metaphactory, Sampo-UI is a simpler lower-level read-only implementation that is focused on faceted search and visualizations, based on the Sampo model concept of separating KG and applications built on top of it. Ideally, this means that the Sampo-UI framework is easier to apply to use cases where only faceted search and visualizations are required.

Publication VI presents a new version of Sampo-UI where the process is geared more toward using configuration files. In practice, it was observed that some things are almost always repeated in mostly similar ways when creating a faceted search application. Having a higher-level configuration approach for the most common functionalities means that less new code needs to be written. It can also be easier to learn to use for people without experience in web development.

The experience of using the initial version of Sampo-UI to build the WarVictimSampo portal is discussed in Publication III and Publication II. Developing the WarVictimSampo portal was considerably easier than it would have been without the starting point provided by Sampo-UI. However, adding configuration options to Sampo-UI, as discussed in Publication VI, made it much easier to create new applications so that simple demonstrators can be created in hours. This is apparent, for example, in Publication VIII which shows examples of multiple separate applications made with the new version of Sampo-UI.

3.4 Summary of Results

LD was applied in multiple example cases in the studies included in this thesis. The results of the thesis show that LD can be a useful tool for publishing CH. However, the examples in this thesis also show that LD can be converted to different models and that it is possible to publish the data in multiple models if necessary. It was also determined that CH data often have subtle nuances that can be difficult to fully cover in a general-purpose ontology, and it was often necessary to use at least partially specialized models to meet the requirements of DH researchers.

The results show that faceted search is a powerful tool for both the general public and CH researchers, as it allows users to explore the data without requiring technical expertise. Faceted search can be implemented with SPARQL queries; however, the implementation can be more or less complicated and the queries faster, depending on the data model used.

For creating faceted search applications on top of SPARQL endpoints, it proved useful to have an application that can be reused as a starting point. However, it was noted that in practice, it was useful to have declarative configuration files for the most common functions.

4. Discussion

The evaluation of research in the Semantic Web domain can be complex [17] as the effectiveness and usability of the systems depend on multiple factors [23]. This presents challenges for evaluating the methods, tools, and implementations presented in this thesis. The research in this thesis has been assessed based on theoretical and practical implications. In addition, reliability and validity are addressed. Finally, the lessons learned and recommendations for future research are discussed.

4.1 Theoretical and Practical Implications

4.1.1 Publishing Cultural Heritage Data (RQ1)

The research in this thesis shows that LD can be used to publish CH data. For the purposes of creating faceted search web applications, the process is made easier if the conceptualization of the data matches the conceptualization of the search application. This can sometimes mean somewhat different ontologies or models than the ontologies that aim to harmonize CH data such as CIDOC CRM. As shown, for example, in Publication I and Publication VIII, LD KGs can be converted to other models, and it is possible to have data in multiple forms. However, one model, perhaps the model that is the easiest to maintain, should be the primary form.

FindSampo and CoinSampo open data about archaeological object finds made by the public in Finland. Although data about the cataloged objects that FindSampo publishes is available on FHA web pages, the data was previously mostly available in PDF files. Also, previously, no proper controlled vocabularies were used to describe the objects. In FindSampo KG the objects are given individual URIs and they are described using concepts based on MAO/TAO ontology. The mapping of the concepts to the Getty AAT vocabulary was also created for the concepts used to describe object types in FindSampo. This made it easy to add information about objects in FindSampo KG to the international ARIADNEplus portal. CoinSampo opens data about all Finnish reported coin finds between 2013 and 2023. This makes it easier for researchers to make comparative analysis, and also allows metal detectorists to see their own finds online and compare them to other finds.

This work presents the FindSampo Core ontology for representing archaeological finds made by the public. This is a simple artifact-based ontology similar to Dublin Core. In contrast to, for example, event-based ontology, like the CIDOC CRM, which is meant to be general-purpose, the FindSampo ontology has only a small number of classes and properties, making it lightweight and simple to use. CoinSampo applies the FindSampo approach to numismatic object finds. Coins are a special type of object that require multiple classes and properties to describe that are not relevant to most objects. CoinSampo data model does not try to propose a generic model to represent coins in a way similar to the Nomisma.org ontology. Publication VII also presents an example of how often imprecise information about coin finds can be represented by using vague concepts that can represent multiple options. This was needed because of the nature of the Finnish coin find data.

The data published in WarVictimSampo improve and make the historical understanding of the Finnish Civil War more precise, especially the number of victims. The new data adds some 1600 new victims and corrects numerous errors in the information on previously recorded victims. Publishing the data as a LOD KG makes it easier to use the data in future research.

The research in this thesis presents a knowledge-based method for modeling and extracting relations and their explanations between resources in KGs. Representing relations as directed relations, similar to RDF reification, helps when implementing search, but the number of connections can increase memory and other computational requirements significantly. By modeling relations so that the direction is not semantically preserved, one can lower the number of connections but can make search more difficult to implement depending on what the desired search functions are. Compared to other approaches to relational search presented in chapter 2, the approach of using SPARQL CONSTRUCT queries to extract interesting relations based on defined forms can require much more work, and it can miss some serendipitous connections that could not be thought of when defining the forms of interesting connections. Using the predefined forms it is possible to create fully natural language explanations of connections, in comparison to the approaches that use combination of human-readable labels and visual representation to express the explanation of a connection. Using a knowledge-based approach, it is possible to strictly define the connections that are interesting in the data and get those connections reliably and consistently. The possibility of using rules and ontologies to create a human-readable explanation for semantic connections was originally identified in [151], but without showing actual implementation. Publication I and Publication VIII present a method for extracting such connections and creating human-readable explanations.

4.1.2 Application Features (RQ2)

The work in this thesis improves the state of the art for LD web applications by showing how applications can be used to compare properties of groups of individuals and discussing how the same functions in applications can be useful at the same time for presentation, research, and collection management. Compared to the WarSampo perspective, WarVictimSampo offers more robust tools for analyzing the data about the war victims. Similarly, FindSampo and CoinSampo are state-of-the-art web applications for analyzing CH data. Compared to state-of-the-art applications such as PAS, ARIADNE portal, and the applications of the American Numismatic Society, such as Seleucid Coins Online, they offer a wider selection of visualization tools. Notably, CoinSampo can offer wider functionalities compared to FindSampo, because of the more semantically rich data of the coin finds enriched from Wikidata, showing how often the functions of applications are less dependent on the application and more on the quality of the data.

WarVictimSampo pushes the state of the art in publishing war history on the Web internationally. The national importance of WarVictimSampo, for researchers and the general public, is attested by the numerous articles in national and local news outlets that reported its release. The new and more accurate data is the most important improvement, but the web application was also described by newspapers as being “on another planet” compared to the old web application in its usability and versatility [146], and “easier to use for the general public” [142]. WarVictimSampo is used by historians interested in the Finnish Civil War, for example, in a recent doctoral thesis on the effects of the Civil War on the town of Ruovesi [155]. It is also an important resource for genealogists. For example, the Geneological Society of Finland lists WarVictimSampo first in their list[1] of web pages to find information about historical persons in Finland. WarVictimSampo has also been used by journalists interested in local history [89].

By forming relation instances using the methods presented in this thesis, the relations can then be explored with a faceted search using tools such as Sampo-UI and visualized in different ways. Faceted search makes it possible to focus the search on different groups and reveals the relative numbers between them. These can be useful in answering certain research questions. In addition to this, Publication I and Publication VIII also show how the relative numbers of connections can be searched and visualized. A limitation of this is that faceted search of relations can be computationally challenging because the number of relations can be high.

According to Tunkelang [149], the implementation of the search user interface can be complex if the search is performed for multiple types of entities. In contrast to this, it was observed that it was more difficult to create an intuitive

[1]Available at the homepages of the Geneological Society of Finland at `https://www.genealogia.fi/henkilohakusivuja/`. Accessed on 3. December 2024.

user interface for relational search when the two entities were of the same type, such as persons. In relational search concept introduced in thesis, the search can be refined based on the properties of the relation itself, and of both resources that form the endpoints of the connection. When searching connections between the same type of entities, such as two persons, it can be difficult to make an intuitive distinction between the entities. For example, having two separate facets for the occupations of the two persons in the connection can be unintuitive to the user.

4.1.3 Creating Semantic Applications (RQ3)

Sampo-UI has been used in multiple internal and external projects. The previous Norwegian Place Names Portal [126], and the upcoming portal for Radio and Television Archives of Finland (RTVA) are based on Sampo-UI. Golem[2] lab research project [123] has used Sampo-UI when implementing their fan-fiction portal. The Sampo-UI design choices have also been credited as inspiration for the Knowledge Graph Explorer for the Virtual Record Treasury of Ireland [130, 129].

PM-Sampo[3] [137, 138] was created by Dutch researchers in cooperation with the author and others to help explore the provenance of objects in the collections of Dutch museums. A small-scale user study was conducted for PM-Sampo with five cultural heritage professionals, where all participants strongly agreed that the system supports exploratory analysis in a way that is relevant to their research.

Researchers at the University of Latvia conducted a test case for Sampo-UI where a new application was created to explore LD about Nobel Prizes and Nobel Laureates from Nobelprize.org[4]. This application, called Nobel Prize Sampo[5] [45] was created on top of existing SPARQL endpoint that is controlled by a third party and runs with a different type of triplestore than has been used in developing Sampo-UI. The application shows that Sampo-UI can be used to create a user interface on a KG that is not controlled by the developer of the UI. However, some difficulties were encountered that required altering the SPARQL queries created by Sampo-UI in minor ways. This is because SPARQL can be implemented in slightly, but potentially meaningfully, different ways on different triplestores. This does demonstrate a difficulty in creating a generic framework based on SPARQL queries.

[2] https://golemlab.eu/

[3] The demonstrator for PM-Sampo can be tested at https://pmsampo.demo.seco.cs.aalto.fi/en/.

[4] https://www.nobelprize.org/about/developer-zone-2/

[5] Nobel Prize Sampo can be tested at https://nobel-prize-sampo.lumii.lv.

4.2 Reliability and Validity

The research in this thesis consists of case studies in which specific CH datasets were converted to LOD. These datasets have their own history and idiosyncrasies. The research was also carried out mainly using data from a single country and focuses on a single technology stack. However, the case studies were conducted in cooperation with leading experts from Finnish memory organizations, so that they provide added value to them. The demonstrators also provide proof that the selected approach provides at least a working and useful way to publish CH.

4.3 Lessons Learned

Perhaps the most important lesson learned is how useful it is to immediately have robust exploration and visualization tools when starting to work with new data. A simple faceted search will immediately tell a lot about the terms used in the data: What terms are the most common and what terms are perhaps mistakes. Similarly, visualizations can be revealing. Maps can immediately show errors in coordinates that can be obvious on a map but might not have been noticed otherwise. For example, in CoinSampo, one mint of a gold coin had initially been mistakenly linked to Batavia in Indonesia. Because the minting spot was so obviously far away from the others on a map, it was easy to check the mint again and realize that the correct Batavia is in the Netherlands. Similarly in WarVictimSampo an error in a Python code creating machine readable dates, caused by a mistake in handling of the leap day, was spotted from death dates line chart showing a suspiciously high number of deaths at the end of February. Faceted search and visualization can therefore also be valuable to users who are experts in LD as well as lay users. This is true even when users have taken part in creating the data themselves.

It was also revealing that it was often difficult to use standard ontologies when publishing CH research datasets. Each CH dataset tends to have their own idiosyncrasies caused by its individual creation motives and history. Forcing them to a standard ontology would often cause something of the essence of the data to be lost, or require considerable amount of extra work. On the other hand, it is often useful to link used vocabularies to standard vocabularies, such as Wikidata. This makes it possible to enrich the data and also makes it easier for others to reuse.

4.4 Recommendations for Further Research

Maintaining and updating knowledge graphs with new information is a theme that has largely been left outside the scope of this thesis. However, it is relevant for the long-term viability of any data. For example, WarVictimSampo is continuously receiving requests from people who want to fix some perceived errors in the data, often related to their own family members. In this context, there are at least two different use cases for updating the data that may require different solutions: an option for the general public to add (suggestions for) new data and an option for experts to change and add data. In general, adding new data that is of good quality is challenging and requires work. How to make the process cost-efficient and convenient is an interesting research question.

Large language models (LLMs) offer new possibilities for knowledge discovery [68]. For example, in relational search, LLMs can potentially be used to create human-readable explanations of connections and can also potentially find serendipitous connections. With current technology, LLMs can plausibly be used to extract connections and generate explanations. However, there are questions about their reliability and consistency [42]. In comparison, knowledge-based methods that use KGs make it possible to obtain consistent and reliable results. Knowledge-based methods, similar to those employed in this thesis, could also be used to solve quantitative humanities research questions.

Web applications for opening specific datasets to exploration and visualization can be interesting and useful if a researcher is interested in that specific data. However, what CH researchers are often more interested in is their own research data. Based on feedback received for CoinSampo, having a tool that can be set up without too much effort and be used to explore data locally with faceted search and visualizations would be useful to many researchers. Simple RDF conversion and setting up a simple Sampo-UI portal for a data set can be done in a day if the user is familiar with the technology, but the required level of technical understanding is much to ask for from a CH scholar. A plan of future work that has already been started during this thesis includes a system that could, based on easy to do configuration, take data in CSV format, convert that to RDF, and also create a rudimentary Sampo-UI web application. This kind of process would create a limited RDF graph and a portal with limited functionality that could be run on users' computers with Docker Compose. This could be enough for many use cases that CH researchers have and could also be used as a starting point for more complex applications.

The work of converting the uncontrolled terms used in Finnish archaeology reports to match controlled vocabulary and to use URI identifiers was started in the Sualt project and in FindSampo. The work is planned to continue, but funding is limited. There are larger plans to make archaeological research data more FAIR in Finland. Although such projects are not currently funded, a new consortium[6] is being started to improve the archaeological research infrastructure in Finland. The consortium brings together the main archaeological actors in Finland.

Currently, CoinSampo only includes individual coin finds reported by the

[6]A home page for the consortium intially called "Arkeologia 2.0" is available at `https://www.museovirasto.fi/fi/kulttuuriymparisto/arkeologinen-kulttuuriperinto/arkeologia-2-0`.

public, and larger coin hoards from archaeological excavations are excluded. If the hoards were added to CoinSampo KG, information about all coins found in Finland would be openly available, which would improve the research situation for archeologists [38]. Most of the computational infrastructure required exists in CoinSampo, but some additional work would be needed.

The work done in projects related to this thesis includes work with ontologies and vocabularies that has not yet been fully documented in research articles. An example of this is adding FindSampo data to ARIADNEplus. The experiences of certain other international projects that contributed numismatic finds to ARIADNEplus are discussed in [160].

The difficulty of using the Nomisma.org ontology to represent coin finds, noted in Publication VII, is also recognized by the Nomisma.org project. There is ongoing work to create a simpler version of the Nomisma.org ontology and to create a web application based on the FindSampo and CoinSampo approaches.

4.5 Conclusions

This thesis presents multiple case studies where legacy CH data sets were converted to LOD, linked to external sources, and published on SPARQL endpoints. Web applications were built to explore and visualize the KQs. An open source framework called Sampo-UI was created to support the creation of web applications built on top of SPARQL endpoints.

The research and demonstrators created for this thesis show that LD can be a useful way to publish CH data. Using LD makes the data more aligned with the FAIR principles. The standardized format and global identifiers used in LD make the data more reusable and easier to enrich from outside sources, such as Wikidata, which also use LD. Data can be enriched from outside sources with information such as hierarchies, coordinates, and images.

The thesis shows that the faceted search is a useful tool for exploring CH data. It allows users to build their search incrementally and helps to see patterns in the data. Faceted search requires structured data that is of good quality, but can also help in improving the quality of the data by pointing out patterns and mistakes. The thesis also explores new ways to apply faceted search to search relations between resources in a CH KG. Relations can be explored with faceted search, which offers new possibilities for understanding the data.

This thesis also shows that web applications can be created efficiently on top of SPARQL endpoints by using an existing application that is meant to be used as a framework or template when creating new applications. Using configuration files to allow for the setting up of the most common functions in a declarative manner makes the process less complicated for application developers.

References

[1] Mariana Aguiar, Sérgio Nunes, and Bruno Giesteirad. A survey on user interaction with Linked Data. In *Proceedings of the Sixth International Workshop on the Visualization and Interaction for Ontologies and Linked Data (VOILA! 2021)*, volume 3023, pages 13–28. CEUR Workshop Proceedings, November 2021. https://ceur-ws.org/Vol-3023/paper5.pdf.

[2] Annastiina Ahola, Eero Hyvönen, Heikki Rantala, and Anne Kauppala. Historical opera and music theatre performances on the Semantic Web: OperaSampo 1830-1960. In *Proceedings of the 20th International Conference on Semantic Systems (SEMANTiCS 2024)*, Knowledge Graphs in the Age of Language Models and Neuro-Symbolic AI, Studies on the Semantic Web, pages 386–402, Amsterdam, Netherlands, September 2024. IOS Press. http://doi.org/10.3233/SSW240031.

[3] Annastiina Ahola, Lilli Peura, Rafael Leal, Heikki Rantala, and Eero Hyvönen. Using generative AI and LLMs to enrich art collection metadata for searching, browsing, and studying art history in Digital Humanities. In *Proceedings, 2nd International Conference on Data & Digital Humanities Generative Artificial Intelligence for Text and Multimodal Data 12th - 13th December 2024, University of Minho, Braga, Portugal*, December 2024. In press, preprint available online at https://seco.cs.aalto.fi/publications/2024/ahola-et-al-genai-2024.pdf.

[4] Wafa' Za'al Alma'aitah, Abdullah Zawawi Talib, and Mohd Azam Osman. Opportunities and challenges in enhancing access to metadata of cultural heritage collections: a survey. *Artificial Intelligence Review*, 53(5):3621–3646, 2020. https://doi.org/10.1007/s10462-019-09773-w.

[5] Kemafor Anyanwu, Angela Maduko, and Amit Sheth. SemRank: ranking complex relationship search results on the semantic web. In *Proceedings of the 14th International Conference on World Wide Web*, WWW '05, page 117–127, New York, NY, USA, 2005. Association for Computing Machinery. https://doi.org/10.1145/1060745.1060766.

[6] David E Avison, Robert M Davison, and Julien Malaurent. Information systems action research: Debunking myths and overcoming barriers. *Information & Management*, 55(2):177–187, 2018. https://doi.org/10.1016/j.im.2017.05.004.

[7] Murtha Baca and Melissa Gill. Encoding multilingual knowledge systems in the digital age: the Getty vocabularies. *Knowledge Organization (KO)*, 42(4):232–243, 2015. https://doi.org/10.5771/0943-7444-2015-4-232.

[8] Thomas Baker, Sean Bechhofer, Antoine Isaac, Alistair Miles, Guus Schreiber, and Ed Summers. Key choices in the design of Simple Knowledge Organization

System (SKOS). *Journal of Web Semantics*, 20:35–49, 2013. https://doi.org/10.1016/j.websem.2013.05.001.

[9] Alessia Bardi, Miriam Baglioni, Michele Artini, Andrea Mannocci, and Gina Pavone. The ARIADNEplus Knowledge Base: a Linked Open Data set for archaeological research. In *Proceedings of the 32nd Symposium on Advanced Database Systems*, pages 91–100. CEUR Workshop Proceedings, vol. 3741, 2024. https://ceur-ws.org/Vol-3741/paper16.pdf.

[10] Wolfram Bartussek, Hermann Bense, Thomas Hoppe, Bernhard G. Humm, Anatol Reibold, Ulrich Schade, Melanie Siegel, and Paul Walsh. *Introduction to Semantic Applications*, pages 1–12. Springer Berlin Heidelberg, Berlin, Heidelberg, 2018. https://doi.org/10.1007/978-3-662-55433-3_1.

[11] Richard L. Baskerville. Investigating information systems with action research. *Communications of the Association for Information Systems*, 2(1), 1999. https://doi.org/10.17705/1CAIS.00219.

[12] Francesco Beretta. A challenge for historical research: making data FAIR using a collaborative ontology management environment (OntoME). *Semantic Web*, 12(2):279–294, 2021. https://journals.sagepub.com/doi/10.3233/SW-200416.

[13] Eleanora Bernasconi, Miguel Ceriani, and Massimo Mecella. Linked Data interfaces: A survey. In *Proceedings of the 19th Conference on Information and Research Science Connecting to Digital and Library Science (IRCDL 2023)*, volume 3365, pages 1–16, Bari, Italy, February 2023. CEUR Workshop Proceedings. https://ceur-ws.org/Vol-3365/paper1.pdf.

[14] Tim Berners-Lee. Linked Data - Design Issues, 2006. http://www.w3.org/DesignIssues/LinkedData.html, [Accessed 24.4.2025].

[15] Tim Berners-Lee, James Hendler, and Ora Lassila. The Semantic Web. *Scientific American*, 284(5):28–37, 2001. https://doi.org/10.1038/scientificamerican0501-34.

[16] Abraham Bernstein, James Hendler, and Natalya Noy. A New Look at the Semantic Web. *Communications of the ACM*, 59(9):35–37, Aug 2016. http://doi.acm.org/10.1145/2890489.

[17] Abraham Bernstein and Natasha Noy. Is this really science? the semantic webber's guide to evaluating research contributions. Technical report, University of Zurich, Department of Informatics (IFI), 2014. https://capuana.ifi.uzh.ch/publications/PDFs/9417_Report.pdf.

[18] Maria Teresa Biagetti. Ontologies as knowledge organization systems. *KO Knowledge Organization*, 48(2):152–176, 2021. https://doi.org/10.5771/0943-7444-2021-2-152.

[19] Federico Bianchi, Matteo Palmonari, Marco Cremaschi, and Elisabetta Fersini. Actively learning to rank semantic associations for personalized contextual exploration of knowledge graphs. In Eva Blomqvist, Diana Maynard, Aldo Gangemi, Rinke Hoekstra, Pascal Hitzler, and Olaf Hartig, editors, *The Semantic Web: 14th International Conference, ESWC 2017*, Lecture Notes in Computer Science, pages 120–135. Springer, 2017. https://doi.org/10.1007/978-3-319-58068-5_8.

[20] Ceri Binding, Keith May, and Douglas Tudhope. Semantic interoperability in archaeological datasets: Data mapping and extraction via the CIDOC CRM. In *Research and Advanced Technology for Digital Libraries: 12th European Conference, ECDL 2008, Aarhus, Denmark, September 14-19, 2008. Proceedings 12*, pages 280–290. Springer, 2008. https://doi.org/10.1007/978-3-540-87599-4_30.

[21] Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked Data - the story so far. *International Journal on Semantic Web and Information Systems*, 5(3):1–22, 2009. https://doi.org/10.4018/JSWIS.2009081901.

[22] Toby Burrows, Doug Emery, Mitc Fraas, Eero Hyvönen, Esko Ikkala, Mikko Koho, David Lewis, Andrew Morrison, Kevin Page, Lynn Ransom, Emma Thomson, Jouni Tuominen, Athanasios Velios, and Hanno Wijsman. Mapping Manuscript Migrations knowledge graph: data for tracing the history and provenance of medieval and renaissance manuscripts. *Journal of Open Humanities Data*, 6(3), 2020. https://doi.org/10.5334/johd.14.

[23] Frada Burstein and Shirley Gregor. The systems development or engineering approach to research in information systems: An action research perspective. In *Proceedings of the 10th Australasian Conference on Information Systems*, volume 8 of *Asian Libraries*, pages 122–134. Emerald Group Publishing Limited, 1999. https://doi.org/10.1108/al.1999.17308dab.006.

[24] Carlotta Capurro and Gertjan Plets. Europeana, EDM, and the europeanisation of cultural heritage institutions. *Digital Culture & Society*, 6(2):163–190, 2021. https://doi.org/10.14361/dcs-2020-0209.

[25] Ana Carballo-Garcia and Juan-José Boté-Vericad. Fair data: History and present context. *Central European Journal of Educational Research*, 4(2):45–53, December 2022. https://doi.org/10.37441/cejer/2022/4/2/11379.

[26] Frank Carpentier. The Portable Antiquities of the Netherlands: A review. *Advances in Archaeological Practice*, 10(3):347–353, 2022. https://doi.org/10.1017/aap.2022.25.

[27] Valentine Charles and Antoine Isaac. Enhancing the Europeana data model (EDM). Technical report, Europeana, May 2015. https://pro.europeana.eu/files/Europeana_Professional/Publications/EDM_WhitePaper_17062015.pdf.

[28] Gong Cheng, Fei Shao, and Yuzhong Qu. An empirical evaluation of techniques for ranking semantic associations. *IEEE Transactions on Knowledge and Data Engineering*, 29(11):2388–2401, November 2017. https://doi.org/10.1109/TKDE.2017.2735970.

[29] Gong Cheng, Yanan Zhang, and Yuzhong Qu. Explass: exploring associations between entities via top-k ontological patterns and facets. In *The Semantic Web - ISWC 2014*, volume 8797 of *Lecture Notes in Computer Science*, pages 422–437. Springer, 2014. https://doi.org/10.1007/978-3-319-11915-1_27.

[30] Richard Cyganiak, David Wood, Markus Lanthaler, Graham Klyne, Jeremy Carroll, and Brian McBride. RDF 1.1 Concepts and Abstract Syntax. W3C recommendation, World Wide Web Consortium (W3C), February 2014. http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/, [Accessed 20.12.2024].

[31] Aba-Sah Dadzie and Matthew Rowe. Approaches to visualising Linked Data: A survey. *Semantic Web*, 2(2):89–124, 2011. https://doi.org/10.3233/SW-2011-0037.

[32] Edie Davis and Bahareh Heravi. Linked Data and Cultural Heritage: A systematic review of participation, collaboration, and motivation. *Journal on Computing and Cultural Heritage (JOCCH)*, 14(2), May 2021. https://doi.org/10.1145/3429458.

[33] Pieterjan Deckers, Lizzy Bleumers, Sanne Ruelens, Bert Lemmens, Nastasia Vanderperren, Clémence Marchal, Jo Pierson, and Dries Tys. MEDEA: Crowdsourcing the recording of metal-detected artefacts in Flanders (Belgium). *Open Archaeology*, 2(1), 2016. https://doi.org/10.1515/opar-2016-0019.

[34] Martin Doerr. The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine*, 24(3):75–92, September 2003. https://doi.org/10.1609/aimag.v24i3.1720.

[35] Martin Doerr. Ontologies for Cultural Heritage. In Steffen Staab and Rudi Studer, editors, *Handbook on Ontologies*, pages 463–486. Springer, Berlin, Heidelberg, 2009. https://doi.org/10.1007/978-3-540-92673-3_21.

[36] Martin Doerr, Maria Theodoridou, Edeltraud Aspöck, and Anja Masur. Mapping archaeological databases to CIDOC-CRM. In *CAA2015 Keep the revolution going–Proceedings of 43rd Annual Conference of Computer Applications and Quantitative Methods in Archaeology*, pages 443–451. Archaeopress Publishing Ltd, 2016. https://www.archaeopress.com/Archaeopress/download/9781784913373.

[37] Senka Drobac, Johanna Enqvist, Petri Leskinen, Muhammad Faiz Wahjoe, Heikki Rantala, Mikko Koho, Ilona Pikkanen, Iida Jauhiainen, Jouni Tuominen, Hanna-Leena Paloposki, Matti La Mela, and Eero Hyvönen. The laborious cleaning: Acquiring and transforming 19th-century epistolary metadata. In *Digital Humanities in the Nordic and Baltic Countries Publication, DHNB2023 Conference Proceeding*, volume 5, pages 248–262. University of Oslo Library, Norway, 2023. https://journals.uio.no/dhnbpub/article/view/10669.

[38] Frida Ehrnsten. Myntfynd i Finland årgång 2023. *Nordisk Numismatik Unions Medlemsblad*, 3:69–78, 2024. https://numismatik.dk/wp-content/uploads/NNUM_2024_03.pdf.

[39] Fredo Erxleben, Michael Günther, Markus Krötzsch, Julian Mendez, and Denny Vrandečić. Introducing Wikidata to the Linked Data Web. In Peter Mika, Tania Tudorache, Abraham Bernstein, Chris Welty, Craig Knoblock, Denny Vrandečić, Paul Groth, Natasha Noy, Krzysztof Janowicz, and Carole Goble, editors, *The Semantic Web – ISWC 2014: 13th International Semantic Web Conference*, volume 8796 of *Lecture Notes in Computer Science*, pages 50–65, Riva del Garda, Italy, October 2014. Springer. https://doi.org/10.1007/978-3-319-11964-9_4.

[40] Richard Gartner. What metadata is and why it matters. In *Metadata: Shaping Knowledge from Antiquity to the Semantic Web*, pages 1–13. Springer, Cham, 2016. https://doi.org/10.1007/978-3-319-40893-4_1.

[41] Lia Gasbarra, Mikko Koho, Ilkka Jokipii, Heikki Rantala, and Eero Hyvönen. An Ontology of Finnish Historical Occupations. In Pascal Hitzler, Sabrina Kirrane, Olaf Hartig, Victor de Boer, Maria-Esther Vidal, Maria Maleshkova, Stefan Schlobach, Karl Hammar, Nelia Lasierra, Steffen Stadtmüller, Katja Hose, and Ruben Verborgh, editors, *The Semantic Web: ESWC 2019 Satellite Events*, volume 11762 of *Lecture Notes in Computer Science*, Portorož, Slovenia, June 2019. Springer, Cham. https://doi.org/10.1007/978-3-030-32327-1_13.

[42] Manas Gaur and Amit Sheth. Building trustworthy neurosymbolic AI systems: Consistency, reliability, explainability, and safety. *AI Magazine*, 45(1):139–155, February 2024. https://doi.org/10.1002/aaai.12149.

[43] Victor Gay and Pauline Grosjean. Morts Pour la France: A database of French fatalities of the Great War. *Explorations in Economic History*, 90, 2023. https://doi.org/10.1016/j.eeh.2023.101550.

[44] Fausto Giunchiglia and Ilya Zaihrayeu. Lightweight Ontologies. In Ling Liu and M. Tamer Özsu, editors, *Encyclopedia of Database Systems*, pages 1613–1619. Springer, US, 2009. https://doi.org/10.1007/978-0-387-39940-9_1314.

[45] Niks Kristofers Grislis, Kārlis Čerāns, Mikus Grasmanis, Heikki Rantala, and Eero Hyvönen. How to add a user interface on top of an external SPARQL endpoint: Case Nobel Prize Sampo. In *Joint Proceedings of Posters, Demos, Workshops, and Tutorials of the 21st International Conference on Semantic Systems*,

volume 4064. CEUR Workshop Proceedings, Aachen, Germany, September 2025. https://ceur-ws.org/Vol-4064/PD-paper17.pdf.

[46] Ethan Gruber. Using XForms to create, publish, and manage Linked Open Data. In *XML London 2016 – Conference Proceedings*, pages 99–107. XML London, 2016. https://xmllondon.com/2016/xmllondon-2016-proceedings.pdf.

[47] Thomas Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220, 1993. https://doi.org/10.1006/knac.1993.1008.

[48] Nicola Guarino, Daniel Oberle, and Steffen Staab. What is an Ontology. In Steffen Staab and Rudi Studer, editors, *Handbook on Ontologies*, pages 1–17. Springer, Berlin, Heidelberg, 2nd edition, April 2009. https://doi.org/10.1007/978-3-540-92673-3_0.

[49] Cristóbal Torres Gutiérrez and Aidan Hogan. WoolNet: Finding and visualising paths in knowledge graphs. In *Proceedings of the 16th Alberto Mendelzon International Workshop on Foundations of Data Management (AMW 2024), Mexico City, Mexico*, volume 3954. CEUR Workshop Proceedings, 2024. https://ceur-ws.org/Vol-3954/paper3781.pdf.

[50] Peter Haase, Daniel Herzig, Artem Kozlov, Andriy Nikolov, and Johannes Trame. metaphactory: A platform for knowledge graph management. *Semantic Web*, 10(6):1109–1125, 2019. https://doi.org/10.3233/SW-190360.

[51] Mordechai Haklay and Patrick Weber. OpenStreetMap: User-generated street maps. *IEEE Pervasive computing*, 7(4):12–18, 2008. https://doi.org/10.1109/MPRV.2008.80.

[52] Mordechai (Muki) Haklay, Daniel Dörler, Florian Heigl, Marina Manzoni, Susanne Hecker, and Katrin Vohland. What is Citizen Science? the challenges of definition. In Katrin Vohland, Anne Land-Zandstra, Luigi Ceccaroni, Rob Lemmens, Josep Perelló, Marisa Ponti, Roeland Samson, and Katherin Wagenknecht, editors, *The Science of Citizen Science*, pages 13–33. Springer, Cham, 2021. https://doi.org/10.1007/978-3-030-58278-4_2.

[53] Patricia Harpring. Development of the Getty Vocabularies: AAT, TGN, ULAN, and CONA. *Art Documentation: Journal of the Art Libraries Society of North America*, 29(1):67–72, 2010. https://www.jstor.org/stable/27949541.

[54] Steve Harris and Andy Seaborne. SPARQL 1.1 Query Language. W3C recommendation, World Wide Web Consortium (W3C), March 2013. https://www.w3.org/TR/sparql11-query/ [accessed 15.10.2025].

[55] Ali Hasnain and Dietrich Rebholz-Schuhmann. Assessing FAIR data principles against the 5-star open data principles. In Aldo Gangemi, Anna Lisa Gentile, Andrea Giovanni Nuzzolese, Sebastian Rudolph, Maria Maleshkova, Heiko Paulheim, Jeff Z. Pan, and Mehwish Alam, editors, *The Semantic Web: ESWC 2018 Satellite Events, Crete, Greece*, Lecture Notes in Computer Science, pages 469–477. Springer, Cham, 2018. https://doi.org/10.1007/978-3-319-98192-5_60.

[56] Sebastian Heath. SPARQL as a first step for querying and transforming numismatic data: Examples from Nomisma.org. In *Alexander the Great. A Linked Open World*, pages 35–52. Ausonius éditions, Bordeaux, 2018.

[57] Alan R Hevner, Salvatore T March, Jinsoo Park, and Sudha Ram. Design Science in Information Systems Research. *MIS quarterly*, 28(1):75–105, 2004. https://doi.org/10.2307/25148625.

[58] Pascal Hitzler. A review of the semantic web field. *Communications of the ACM*, 64(2):76–83, 2021. https://doi.org/10.1145/3397512.

[59] Eero Hyvönen. *Publishing and Using Cultural Heritage Linked Data on the Semantic Web*, volume 2 of *Synthesis Lectures on the Semantic Web: Theory and Technology*. Morgan & Claypool Publishers, 2012.

[60] Eero Hyvönen, Esko Ikkala, Mikko Koho, Jouni Tuominen, Toby Burrows, Lynn Ransom, and Hanno Wijsman. Mapping Manuscript Migrations on the Semantic Web: A semantic portal and Linked Open Data service for premodern manuscript research. In et al. Hotho, A., editor, *The Semantic Web–ISWC 2021: 20th International Semantic Web Conference, ISWC 2021*, Lecture Notes in Computer Science, pages 615–630. Springer, Cham, 2021. https://doi.org/10.1007/978-3-030-88361-4_36.

[61] Eero Hyvönen, Eetu Mäkelä, Tomi Kauppinen, Olli Alm, Jussi Kurki, Tuukka Ruotsalo, Katri Seppälä, Joeli Takala, Kimmo Puputti, Heini Kuittinen, Kim Viljanen, Jouni Tuominen, Tuomas Palonen, Matias Frosterus, Reetta Sinkkilä, Panu Paakkarinen, Joonas Laitio, and Katariina Nyberg. CultureSampo – Finnish culture on the Semantic Web 2.0. Thematic perspectives for the end-user. In J. Trant and Bearman D., editors, *Museums and the Web 2009, Proceedings*. Archives & Museum Informatics, 2009. https://seco.cs.aalto.fi/publications/2009/hyvonen-et-al-culsa-mw-2009.pdf.

[62] Eero Hyvönen, Eetu Mäkelä, Mirva Salminen, Arttu Valo, Kim Viljanen, Samppa Saarela, Miikka Junnila, and Suvi Kettula. MuseumFinland—Finnish museums on the semantic web. *Journal of Web Semantics*, 3(2-3):224–241, 2005. https://doi.org/10.1016/j.websem.2005.05.008.

[63] Eero Hyvönen, Laura Sinikallio, Petri Leskinen, Senka Drobac, Rafael Leal, Matti La Mela, Jouni Tuominen, Henna Poikkimäki, and Heikki Rantala. Publishing and using parliamentary Linked Data on the Semantic Web: ParliamentSampo system for Parliament of Finland. *Semantic Web*, 2025. https://journals.sagepub.com/doi/full/10.3233/SW-243683.

[64] Eero Hyvönen, Jouni Tuominen, Miika Alonen, and Eetu Mäkelä. Linked Data Finland: A 7-star Model and Platform for Publishing and Re-using Linked Datasets. In Valentina Presutti, Eva Blomqvist, Raphael Troncy, Harald Sack, Ioannis Papadakis, and Anna Tordai, editors, *The Semantic Web: ESWC 2014 Satellite Events*, volume 8798 of *Lecture Notes in Computer Science*, pages 226–230, Crete, Greece, May 2014. Springer, Cham. https://doi.org/10.1007/978-3-319-11955-7_24.

[65] Eero Hyvönen. Semantic Portals for Cultural Heritage. In Steffen Staab and Rudi Studer, editors, *Handbook on Ontologies*, pages 757–778. Springer, Berlin, Heidelberg, 2nd edition, April 2009. https://doi.org/10.1007/978-3-540-92673-3_34.

[66] Eero Hyvönen. Digital Humanities on the Semantic Web: Sampo Model and Portal series. *Semantic Web*, 14(4):729–744, 2022. https://doi.org/10.3233/SW-223034.

[67] Eero Hyvönen. How to create and use a national cross-domain ontology and data infrastructure on the Semantic Web. *Semantic Web*, 15(4):1499–1513, 2024. https://doi.org/10.3233/SW-243468.

[68] Eero Hyvönen. Serendipitous knowledge discovery on the Web of Wisdom based on searching and explaining interesting relations in knowledge graphs. *Journal of Web Semantics*, 85, 2025. https://doi.org/10.1016/j.websem.2024.100852.

[69] Eero Hyvönen, Patrik Boman, Heikki Rantala, Annastiiina Ahola, and Petri Leskinen. ConfermentSampo - a knowledge graph, data service, and semantic portal for intangible academic Cultural Heritage 1643-2023 in Finland. In *Proceedings of the 6th International Knowledge Graph and Semantic Web Conference, Paris, France*, Lecture Notes in Computer Science. Springer, Cham, 2024. https://doi.org/10.1007/978-3-031-81221-7_1.

[70] Eero Hyvönen, Hien Cao, Rafael Leal, Heikki Rantala, and Aki Hietanen. A model and case study for searching and reading cross-border multilingual legislation on the Semantic Web. In *Knowledge Graphs in the Age of Language Models and Neuro-Symbolic AI. Proceedings of the 20th International Conference on Semantic Systems, 17–19 September 2024, Amsterdam, The Netherlands*. IOS Press, 2024. `https://doi.org/10.3233/SSW240021`.

[71] Eero Hyvönen, Erkki Heino, Petri Leskinen, Esko Ikkala, Mikko Koho, Minna Tamper, Jouni Tuominen, and Eetu Mäkelä. WarSampo data service and semantic portal for publishing linked open data about the Second World War history. In *The Semantic Web—Latest Advances and New Domains (ESWC 2016)*, Lecture Notes in Computer Science. Springer, Cham, 2016. `https://doi.org/10.1007/978-3-319-34129-3_46`.

[72] Eero Hyvönen, Petri Leskinen, Henna Poikkimäki, Heikki Rantala, Jouni Tuominen, Senka Drobac, Ossi Koho, Ilona Pikkanen, and Hanna-Leena Paloposki. Searching, exploring, and analyzing historical letters and the underlying networks: LetterSampo Finland (1809–1917) data service and semantic portal. In *Digital Humanities in Nordic and Baltic Countries (DHNB 2025). Book of Abstracts.*, 2025. `https://dhnb.eu/wp-content/uploads/2025/03/DHNB-2025_abstracts.pdf`.

[73] Eero Hyvönen, Petri Leskinen, Minna Tamper, Heikki Rantala, Esko Ikkala, Jouni Tuominen, and Kirsi Keravuori. BiographySampo - Publishing and Enriching Biographies on the Semantic Web for Digital Humanities Research. In et al. Hitzler, P., editor, *The Semantic Web: ESWC 2019*, volume 11762 of *Lecture Notes in Computer Science*. Springer, Cham, 2019. `https://doi.org/10.1007/978-3-030-21348-0_37`.

[74] Eero Hyvönen, Petri Leskinen, Minna Tamper, Heikki Rantala, Esko Ikkala, Jouni Tuominen, and Kirsi Keravuori. Demonstrating BiographySampo in solving digital humanities research problems in biography and prosopography. In *The Fourth Digital Humanities in the Nordic Countries 2019 (DHN2019), Book of Abstracts*. University of Copenhagen, March 2019. `https://cst.dk/DHN2019Pro/abstracts/hyvonen-et-al-dhn-2019-bs.pdf`.

[75] Eero Hyvönen, Petri Leskinen, Minna Tamper, Heikki Rantala, Esko Ikkala, Jouni Tuominen, and Kirsi Keravuori. Linked data – a paradigm shift for publishing and using biography collections on the semantic web. In *Proceedings of the Third Conference on Biographical Data in a Digital World (BD 2019)*, pages 16–23. CEUR-WS Proceedings, vol. 3152, 2022. `https://ceur-ws.org/Vol-3152/BD2019_paper_3.pdf`.

[76] Eero Hyvönen, Petri Leskinen, Minna Tamper, Jouni Tuominen, and Kirsi Keravuori. Semantic National Biography of Finland. In *Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference (DHN 2018)*, pages 372–385. CEUR Workshop Proceedings, Vol-2084, March 2018. `https://ceur-ws.org/Vol-2084/short12.pdf`.

[77] Eero Hyvönen, Eetu Mäkelä, Tomi Kauppinen, Olli Alm, Jussi Kurki, Tuukka Ruotsalo, Katri Seppälä, Joeli Takala, Kimmo Puputti, Heini Kuittinen, Kim Viljanen, Jouni Tuominen, Tuomas Palonen, Matias Frosterus, Reetta Sinkkilä, Panu Paakkarinen, Joonas Laitio, and Katariina Nyberg. CultureSampo – A National Publication System of Cultural Heritage on the Semantic Web 2.0. In Lora Aroyo, Paolo Traverso, Fabio Ciravegna, Philipp Cimiano, Tom Heath, Eero Hyvönen, Riichiro Mizoguchi, Eyal Oren, Marta Sabou, and Elena Simperl, editors, *The Semantic Web: Research and Applications, 6th European Semantic Web Conference, ESWC 2009*, volume 5554 of *Lecture Notes in Computer Science*, Crete, Greece, May 31 - June 4 2009. Springer, Berlin, Heidelberg. `https://doi.org/10.1007/978-3-642-02121-3_69`.

[78] Eero Hyvönen, Heikki Rantala, Esko Ikkala, Mikko Koho, Jouni Tuominen, Babatunde Anafi, Suzie Thomas, Anna Wessman, Eljas Oksanen, Ville Rohiola, Jutta Kuitunen, and Minna Ryyppö. Citizen science archaeological finds on the Semantic Web: The FindSampo framework. *Antiquity, A Review of World Archaeology*, 95(382):e24, August 2021. https://doi.org/10.15184/aqy.2021.87.

[79] Eero Hyvönen and Jouni Tuominen. 8-star Linked Open Data model: Extending the 5-star model for better reuse, quality, and trust of data. In *Posters, Demos, Workshops, and Tutorials of the 20th International Conference on Semantic Systems (SEMANTiCS 2024)*, volume 3759. CEUR Workshop Proceedings, September 2024. https://ceur-ws.org/Vol-3759/paper4.pdf.

[80] Eero Hyvönen, Kim Viljanen, Jouni Tuominen, and Katri Seppälä. Building a national Semantic Web Ontology and Ontology service infrastructure - The FinnONTO approach. In *Proceedings of the European Semantic Web Conference ESWC 2008*, volume 5021 of *Lecture Notes in Computer Science*, pages 95–109. Springer, Berlin, Heidelbergr, June 2008. https://doi.org/10.1007/978-3-540-68234-9_10.

[81] Juhani Iivari and John R Venable. Action research and design science research-seemingly similar but decisively dissimilar. In *European COnference on Information Systems (ECIS 2009), proceedings*, 2009. https://aisel.aisnet.org/ecis2009/73.

[82] Esko Ikkala, Jouni Tuominen, Jaakko Raunamaa, Tiina Aalto, Terhi Ainiala, Helinä Uusitalo, and Eero Hyvönen. NameSampo: A Linked Open Data infrastructure and workbench for toponomastic research. In Patricia Murrieta and Bruno Martins, editors, *GeoHumanities'18: Proceedings of the 2nd ACM SIGSPATIAL Workshop on Geospatial Humanities*, Seattle, WA, USA, November 2018. ACM, New York, NY, USA. http://doi.acm.org/10.1145/3282933.3282936.

[83] Antoine Isaac. Europeana Data Model Primer. Technical report, Europeana, 2013. https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf, [Accessed 6.1.2025].

[84] Antoine Isaac and Bernhard Haslhofer. Europeana Linked Open Data – data.europeana.eu. *Semantic Web*, 4(3):291–297, 2013. https://doi.org/10.3233/SW-120092.

[85] Annika Jacobsen, Ricardo de Miranda Azevedo, Nick Juty, Dominique Batista, Simon Coles, Ronald Cornet, Mélanie Courtot, Mercè Crosas, Michel Dumontier, Chris T Evelo, et al. FAIR principles: interpretations and implementation considerations. *Data intelligence*, 2(1-2):10–29, 2020. https://doi.org/10.1162/dint_r_00024.

[86] Pertti Järvinen. Action research is similar to design science. *Quality & Quantity*, 41:37–54, February 2007. http://dx.doi.org/10.1007/s11135-005-5427-1.

[87] Constantia Kakali, Irene Lourdi, Thomais Stasinopoulou, Lina Bountouri, Christos Papatheodorou, Martin Doerr, and Manolis Gergatsoulis. Integrating Dublin Core metadata for cultural heritage collections using ontologies. In *DCMI '07: Proceedings of the 2007 international conference on Dublin Core and Metadata Applications*, pages 128–139. DCPapers, 2007. https://dcpapers.dublincore.org/files/articles/952108714/dcmi-952108714.pdf.

[88] Eric Kansa. Openness and archaeology's information ecosystem. *World Archaeology*, 44(4):498–520, 2012. https://doi.org/10.1080/00438243.2012.737575.

[89] Juha Karilainen. Valkeakosken alueelta löytyi uusia tietoja vuoden 1918 tapahtumissa surmansa saaneista – katso koko nimilista: he olivat synkän ajan paikalliset uhrit. Valkeakosken Sanomat, newspaper article is available online at: https://www.valkeakoskensanomat.fi/uutiset/art-2000006968065.html. [Accessed 29.12.2024].

[90] Ali Khalili, Antonis Loizou, and Frank van Harmelen. Adaptive Linked Data-driven Web components: Building flexible and reusable Semantic Web interfaces. In Harald Sack, Eva Blomqvist, Mathieu d'Aquin, Chiara Ghidini, Simone Paolo Ponzetto, and Christoph Lange, editors, *The Semantic Web. Latest Advances and New Domains. ESWC 2016.*, volume 9678 of *Lecture Notes in Computer Science*, pages 677–692. Springer, Cham, 2016. https://doi.org/10.1007/978-3-319-34129-3_41.

[91] Ali Khalili, Peter Van den Besselaar, and Klaas Andries de Graaf. FERASAT: A serendipity-fostering faceted browser for Linked Data. In Aldo Gangemi, Roberto Navigli, Maria-Esther Vidal, Pascal Hitzler, Raphaël Troncy, Laura Hollink, Anna Tordai, and Mehwish Alam, editors, *The Semantic Web. ESWC 2018*, volume 10843 of *Lecture Notes in Computer Science*, pages 351–366. Springer, Cham, 2018. https://doi.org/10.1007/978-3-319-93417-4_23.

[92] Mikko Koho. *Representing, Using, and Maintaining Military Historical Linked Data on the Semantic Web*. Doctoral dissertation, Aalto University, School of Science, 2020. https://urn.fi/URN:ISBN:978-952-60-3869-8.

[93] Mikko Koho, Erkki Heino, and Eero Hyvönen. SPARQL Faceter – Client-side faceted search based on SPARQL. In *Joint Proceedings of the 4th International Workshop on Linked Media and the 3rd Developers Hackshop*. CEUR Workshop Proceedings, 2016. http://ceur-ws.org/Vol-2187/paper5.pdf.

[94] Mikko Koho, Eero Hyvönen, Erkki Heino, Jouni Tuominen, Petri Leskinen, and Eetu Mäkelä. Linked death—representing, publishing, and using Second World War death records as Linked Open Data. In Alessandro Adamou, Enrico Daga, and Leif Isaksen, editors, *Proceedings of the 1st Workshop on Humanities in the Semantic Web - WHiSe (at ESWC 2016)*, pages 3–14. CEUR Workshop Proceedings, May 2016. https://ceur-ws.org/Vol-1608/paper-02.pdf.

[95] Mikko Koho, Esko Ikkala, Petri Leskinen, Minna Tamper, Jouni Tuominen, and Eero Hyvönen. WarSampo knowledge graph: Finland in the second world war as Linked Open Data. *Semantic Web*, 12(2):265–278, 2021. https://journals.sagepub.com/doi/10.3233/SW-200392.

[96] Mikko Koho, Rafael Leal, Esko Ikkala, Minna Tamper, Heikki Rantala, and Eero Hyvönen. Building lightweight ontologies for faceted search with named entity recognition: Case WarMemoirSampo. In *Proceedings of the 1st International Workshop on Knowledge Graph Generation From Text and the 1st International Workshop on Modular Knowledge co-located with 19th Extended Semantic Conference (ESWC 2022)*, volume 3184, pages 19–35. CEUR Workshop Proceedings, 2022. https://ceur-ws.org/Vol-3184/TEXT2KG_Paper_2.pdf.

[97] Mikko Koho, Heikki Rantala, and Eero Hyvönen. Digital humanities and military history: Analyzing casualties of the WarSampo knowledge graph. In Karl Berglund, Matti La Mela, and Inge Zwart, editors, *Proceedings of the 6th Digital Humanities in the Nordic and Baltic Countries Conference (DHNB 2022)*, volume 3232, pages 308–316. CEUR Workshop Proceedings, 2022. https://ceur-ws.org/Vol-3232/paper29.pdf.

[98] Lukas Koster and Saskia Woutersen-Windhouwer. FAIR principles for library, archive and museum collections: A proposal for standards for reusable collections. *Code4Lib Journal*, 40, 2018. https://journal.code4lib.org/articles/13427.

[99] Bill Kules, Robert Capra, Matthew Banta, and Tito Sierra. What do exploratory searchers look at in a faceted search interface? In *Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries*, JCDL '09, page 313–322, New York, NY, USA, 2009. Association for Computing Machinery. https://doi.org/10.1145/1555400.1555452.

[100] Åsa M Larsson and Daniel Löwenborg. The digital future of the past-research potential with increasingly FAIR archaeological data. In Charlotta Hillerdal and Kristin Ilves, editors, *Re-imagining Periphery: Archaeology and Text in Northern Europe from Iron Age to Viking and Early Medieval Periods*, pages 61–70. Oxbow Books, 2020. https://doi.org/10.2307/j.ctv138wt08.

[101] Jens Lehmann, Jörg Schüppel, and Sören Auer. Discovering unknown connections–the DBpedia Relationship Finder. In *The Social Semantic Web 2007 - Proceedings of the 1st Conference on Social Semantic Web (CSSW)*, volume 113 of *Lecture Notes in Informatics*, pages 99–110. GI, 2007. http://subs.emis.de/LNI/Proceedings/Proceedings113/gi-proc-113-010.pdf.

[102] Petri Leskinen, Goki Miyakita, Mikko Koho, and Eero Hyvönen. Combining faceted search with data-analytic visualizations on top of a SPARQL endpoint. In Valentina Ivanova, Patrick Lambrix, Steffen Lohmann, and Catia Pesquita, editors, *Proceedings of the Fourth International Workshop on Visualization and Interaction for Ontologies and Linked Data (VOILA 2018)*, volume 2187, pages 53–63. CEUR Workshop, October 2018. https://ceur-ws.org/Vol-2187/paper5.pdf.

[103] Michael Lewis. A detectorist's utopia? Archaeology and metal-detecting in England and Wales. *Open Archaeology*, 2(1), 2016. https://doi.org/10.1515/opar-2016-0009.

[104] Michael Lewis, Eljas Oksanen, Frida Ehrnsten, Heikki Rantala, Jouni Tuominen, and Eero Hyvönen. The impact of human decision-making on the research value of archaeological data. *Journal on Computing and Cultural Heritage (JOCHH)*, 2025. Accepted.

[105] Steffen Lohmann, Philipp Heim, Timo Stegemann, and Jürgen Ziegler. The RelFinder user interface: Interactive exploration of relationships between objects of interest. In *Proceedings of the 14th International Conference on Intelligent User Interfaces (IUI 2010)*, pages 421–422. ACM, 2010. http://doi.acm.org/10.1145/1719970.1720052.

[106] Janet C. Luhmann and Manuel Burghardt. Digital humanities—A discipline in its own right? An analysis of the role and position of digital humanities in the academic landscape. *Journal of the Association for Information Science and Technology*, 73:148 – 171, 2021. https://doi.org/10.1002/asi.24533.

[107] Eetu Mäkelä, Kaisa Hypén, and Eero Hyvönen. BookSampo—lessons learned in creating a semantic portal for fiction literature. In *The Semantic Web – ISWC 2011*, pages 173–188. Springer, Berlin, Heidelberg, 2011. https://doi.org/10.1007/978-3-642-25093-4_12.

[108] Eetu Mäkelä, Eero Hyvönen, Samppa Saarela, and Kim Viljanen. ONTOVIEWS–A tool for creating semantic web portals. In S.A. McIlraith, D. Plexousakis, and F. van Harmelen, editors, *The Semantic Web – ISWC 2004*, volume 3298 of *Lecture Notes in Computer Science*, pages 797–811. Springer, Berlin, Heidelberg, 2004. https://doi.org/10.1007/978-3-540-30475-3_55.

[109] Eetu Mäkelä, Tuukka Ruotsalo, and Eero Hyvönen. How to deal with massively heterogeneous cultural heritage data—lessons learned in CultureSampo. *Semantic Web*, 3(1):85–109, 2012. https://doi.org/10.3233/SW-2012-0049.

[110] Gary Marchionini. Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4):41–46, April 2006. https://doi.org/10.1145/1121949.1121979.

[111] Carlo Meghini, Roberto Scopigno, Julian Richards, Holly Wright, Guntram Geser, Sebastian Cuy, Johan Fihn, Bruno Fanini, Hella Hollander, Franco Niccolucci, Achille Felicetti, Paola Ronzino, Federico Nurra, Christos Papatheodorou, Dimitris Gavrilis, Maria Theodoridou, Martin Doerr, Douglas Tudhope, Ceri Binding, and Andreas Vlachidis. ARIADNE: A research infrastructure for archaeology. *Journal on Computing and Cultural Heritage (JOCCH)*, 10(3), August 2017. https://doi.org/10.1145/3064527.

[112] Alistair Miles, Brian Matthews, Michael Wilson, and Dan Brickley. SKOS Core: Simple knowledge organisation for the Web. In *International Conference on Dublin Core and Metadata Applications (DC-2005)*, pages 3–10. DCPapers, 2005. https://dcpapers.dublincore.org/files/articles/952107985/dcmi-952107985.pdf.

[113] Eetu Mäkelä, Juha Törnroos, Thea Lindquist, and Eero Hyvönen. WW1LOD: An application of CIDOC-CRM to World War 1 linked data. *International Journal on Digital Libraries*, 18(4):333–343, 2017. https://doi.org/10.1007/s00799-016-0186-2.

[114] Jakob Nielsen. *Usability engineering*. Morgan Kaufmann, 1994.

[115] Xi Niu, Xiangyu Fan, and Tao Zhang. Understanding faceted search from data science and human factor perspectives. *ACM Transactions on Information Systems (TOIS)*, 37(2), January 2019. https://doi.org/10.1145/3284101.

[116] Philipp Nussbaumer and Bernhard Haslhofer. Putting the CIDOC CRM into practice-experiences and challenges. Technical report, University of Vienna, 2007. https://eprints.cs.univie.ac.at/404/1/covered.pdf [Accessed 15.10.2025].

[117] Eljas Oksanen, Heikki Rantala, Jouni Tuominen, Michael Lewis, David Wigg-Wolf, Frida Ehrnsten, and Eero Hyvönen. Digital Humanities solutions for pan-european numismatic and archaeological heritage based on Linked Open Data. In Karl Berglund, Matti La Mela, and Inge Zwart, editors, *Proceedings of the 6th Digital Humanities in the Nordic and Baltic Countries Conference (DHNB 2022)*, volume 3232, pages 352–360. CEUR Workshop Proceedings, 2022. http://ceur-ws.org/Vol-3232/paper34.pdf.

[118] Eljas Oksanen and Anna Wesman. New horizons in understanding Finnish Iron Age material culture through metal-detected finds. *Internet Archaeology*, 68, 2025. https://doi.org/10.11141/ia.68.4.

[119] Dominic Oldman and Diana Tanase. Reshaping the knowledge graph by connecting researchers, data and practices in ResearchSpace. In D. Vrandečić et al., editors, *The Semantic Web – ISWC 2018*, volume 11137 of *Lecture Notes in Computer Science*, pages 325–340. Springer, Cham, 2018. https://doi.org/10.1007/978-3-030-00668-6_20.

[120] Eyal Oren, Renaud Delbru, and Stefan Decker. Extending faceted navigation for RDF data. In Isabel Cruz, Stefan Decker, Dean Allemang, Chris Preist, Daniel Schwabe, Peter Mika, Mike Uschold, and Lora M. Aroyo, editors, *The Semantic Web – ISWC 2006*, volume 4273 of *Lecture Notes in Computer Science*, pages 559–572, Athens, GA, USA, November 2006. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11926078_40.

[121] Tuğba Özacar, Övünç Öztürk, Lobaba Salloutah, Fulya Yüksel, Baraa Abdülbaki, and Elif Bilici. A Semantic Web case study: Representing the Ephesus museum collection using Erlangen CRM Ontology. In Emmanouel Garoufallou,

Sirje Virkus, Rania Siatri, and Damiana Koutsomiha, editors, *Metadata and Semantic Research MTSR 2017*, Communications in Computer and Information Science, pages 202–210. Springer, Cham, 2017. https://doi.org/10.1007/978-3-319-70863-8_19.

[122] Brenda O'Neill and Larry Stapleton. Digital cultural heritage standards: from silo to semantic web. *AI & Society*, 37(3):891–903, 2022. https://doi.org/10.1007/s00146-021-01371-1.

[123] Franziska Pannach, Luotong Cheng, and Federico Pianzola. The GOLEM-knowledge graph and search interface: Perspectives into narrative and fiction. In *Proceedings of the Computational Humanities Research Conference 2024*, pages 462–471. CEUR Workshop Proceedings, vol. 3834, 2024. https://ceur-ws.org/Vol-3834/paper80.pdf.

[124] Jung-ran Park and Eric Childress. Dublin Core metadata semantics: An analysis of the perspectives of information professionals. *Journal of Information Science*, 35(6):727–739, 2009. https://doi.org/10.1177/0165551509337871.

[125] Ken Peffers, Tuure Tuunanen, Marcus Rothenberger, and Samir Chatterjee. A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(January):45–77, 2008. https://doi.org/10.2753/MIS0742-1222240302.

[126] Alexandra Petrulevich, Henrik Askjer, Sara Ellis Nilsson, Peder Gammeltoft, Andreas Lecerof, Emily Lethbridge, and Øyvind Liland Gjesdal. Nordic spatial humanities: Ups and downs in LOD implementation across humanities' digital spatial research infrastructures in the nordic countries. 2024. Under review, preprint is available at https://www.semantic-web-journal.net/system/files/swj3754.pdf.

[127] Giuseppe Pirrò. Building relatedness explanations from knowledge graphs. *Semantic Web*, 10(6):963–990, 2019. https://doi.org/10.3233/SW-190348.

[128] Tapio Pitkäranta and Eero Hyvönen. Semantic Web – a forgotten wave of artificial intelligence? In *Proceedings of the 14th International Joint Conference on Knowledge Graphs (IJCKG 2025)*. Springer Verlag, 2025. In press.

[129] Alex Randles, Lucy McKenna, Lynn Kilgallon, Beyza Yaman, Peter Crooks, and Declan O'Sullivan. Evaluating the knowledge graph editor of the Virtual Record Treasury of Ireland. In Bo Fu, Patrick Lambrix, Huanyu Li, Susana Nunes, and Catia Pesquita, editors, *Proceedings of the 9th International Workshop on the Visualization and Interaction for Ontologies, Linked Data and Knowledge Graphs (VOILA! 2024)*, pages 62–77. CEUR Workshop Proceedings, vol. 3773, November 2024. https://ceur-ws.org/Vol-3773/paper5.pdf.

[130] Alex Randles, Lucy McKenna, Lynn Kilgallon, Beyza Yaman, Peter Crooks, and Declan O'Sullivan. The knowledge graph explorer for the Virtual Record Treasury of Ireland. In Bo Fu, Patrick Lambrix, Huanyu Li, Susana Nunes, and Catia Pesquita, editors, *Proceedings of the 9th International Workshop on the Visualization and Interaction for Ontologies, Linked Data and Knowledge Graphs (VOILA! 2024)*, pages 47–61. CEUR Workshop Proceedings, vol. 3773, November 2024. https://ceur-ws.org/Vol-3773/paper4.pdf.

[131] Babak Ranjgar, Abolghasem Sadeghi-Niaraki, Maryam Shakeri, Fatema Rahimi, and Soo-Mi Choi. Cultural heritage information retrieval: Past, present and future trends. *IEEE Access*, 2024.

[132] Julian Richards. Joined up thinking: Aggregating archaeological datasets at an international scale. *Internet Archaeology*, 64, 2023. https://doi.org/10.11141/ia.64.3.

[133] Jesús Robledano-Arillo, Diego Navarro-Bonilla, and Julio Cerdá-Díaz. Application of Linked Open Data to the coding and dissemination of Spanish Civil War photographic archives. *Journal of Documentation*, 76(1):67–95, 10 2019. https://doi.org/10.1108/JD-06-2019-0112.

[134] Stefanie Rühle, Francesca Schulze, and Michael Büchner. Applying a Linked Data compliant model: The usage of the Europeana Data Model by the Deutsche Digitale Bibliothek. In *Proceedings of the International Conference on Dublin Core and Metadata Applications*. DCPapers, 2014. https://dcpapers.dublincore.org/files/articles/952136503/dcmi-952136503.pdf.

[135] Oula Seitsonen. Crowdsourcing cultural heritage: public participation and conflict legacy in Finland. *Journal of Community Archaeology & Heritage*, 4(2):115–130, 2017. https://doi.org/10.1080/20518196.2016.1252129.

[136] Amith Sheth, Boanerges Aleman-Meza, I. Budak Arpinar, Clemens Bertram, Yashodhan Warke, Cartic Ramakrishnan, Chris Halaschek, Kemafar Anyanwu, David Avant, F. Sena Arpinar, and Krys Kochut. Semantic association identification and knowledge discovery for national security applications. *Journal of Database Management*, 16(1):33–53, 2005. https://doi.org/10.4018/jdm.2005010103.

[137] Sarah Binta Alam Shoilee, Annastiina Ahola, Heikki Rantala, Eero Hyvönen, Victor de Boer, Jacco van Ossenbruggen, and Susan Legene. Enhancing provenance research with linked data: A visual approach to knowledge discovery. In *Proceedings of the Second International Workshop of Semantic Digital Humanities (SemDH 2025)*, volume 4009. CEUR Workshop Proceedings, June 2025. https://ceur-ws.org/Vol-4009/paper_6.pdf.

[138] Sarah Binta Alam Shoilee, Victor de Boer, Annastiina Ahola, Heikki Rantala, Eero Hyvönen, Jacco van Ossenbruggen, and Susan Legene. Bridging data gaps: Harnessing semantic associations for knowledge discovery in colonial heritage. In *The Thirteenth International Conference on Knowledge Capture (K-CAP 2025)*, 2025. Accepted.

[139] Ana Luísa Silva and Ana Lúcia Terra. Cultural heritage on the Semantic Web: The Europeana Data Model. *IFLA journal*, 50(1):93–107, 2024. https://doi.org/10.1177/03400352231202506.

[140] Marcus Smith. Linked open data and aggregation infrastructure in the cultural heritage sector: A case study of SOCH, a linked data aggregator for Swedish open cultural heritage. In *Information and knowledge organisation in digital humanities*, pages 64–85. Routledge, 2021.

[141] Peter Spyns, Robert Meersman, and Mustafa Jarrar. Data modelling versus ontology engineering. *ACM SIGMOD Record*, 31(4):12–17, December 2002. https://doi.org/10.1145/637411.637413.

[142] STT. Sotasurmat-sivuston päivityksessä löytyi 1600 sisällissodan ennestään tuntematonta vainajaa, lähes kaikki uhreista punaisia. Keskisuomalainen, Newspaper article available online at https://www.ksml.fi/paikalliset/2370084 [accessed 20.11.2024].

[143] Osma Suominen, Sini Pessala, Jouni Tuominen, Mikko Lappalainen, Susanna Nykyri, Henri Ylikotila, Matias Frosterus, and Eero Hyvönen. Deploying national ontology services: From ONKI to Finto. In Axel Polleres, Alexander Garcia, and Richard Benjamins, editors, *Proceedings of the Industry Track at the International Semantic Web Conference 2014*. CEUR Workshop Proceedings, vol. 1383, October 2014. https://www.ceur-ws.org/Vol-1383/paper6.pdf.

[144] Osma Suominen, Kim Viljanen, and Eero Hyvänen. User-centric faceted search for semantic portals. In *The Semantic Web: Research and Applications. ESWC 2007*, volume 4519 of *Lecture Notes in Computer Science*, pages 356–370. Springer, Berlin, Heidelberg, 2007. https://doi.org/10.1007/978-3-540-72667-8_26.

[145] Patrik Svensson. The landscape of digital humanities. *Digital Humanities Quarterly*, 4(1), 2010. https://digitalhumanities.org/dhqdev/vol/4/1/000080/000080.html.

[146] Ilari Tapio. Sotasurmasampo vahvistaa isosetien kohtalot – amerikan nestori ammuttiin vankileirillä, nahkuri nestori katosi ruoanhakumatkalla. Kaleva, Newspaper article available online at https://www.kaleva.fi/sotasurmasampo-vahvistaa-isosetien-kohtalot-amerik/1669471 [accessed 10.11.2024].

[147] Gonzalo Tartari and Aidan Hogan. WiSP: Weighted shortest paths for RDF graphs. In Valentina Ivanova, Patrick Lambrix, Steffen Lohmann, and Catia Pesquita, editors, *Proceedings of the Fourth International Workshop on Visualization and Interaction for Ontologies and Linked Data (VOILA 2018)*, pages 37–52. CEUR Workshop Proceedings, vol. 2187, 2018. https://ceur-ws.org/Vol-2187/paper4.pdf.

[148] Suzie Thomas, Anna Wessman, Esko Ikkala, Jouni Tuominen, Mikko Koho, and Eero Hyvönen. (Co-)Creating a sustainable platform for Finland's archaeological chance finds: The Story of SuALT. In Ethan Watrall and Lynne Goldstein, editors, *Digital Heritage and Archaeology in Practice*, pages 165–184. University Press of Florida, 2022. https://doi.org/10.2307/j.ctv2pfq2jj.12.

[149] Daniel Tunkelang. *Faceted search*, volume 1 of *Synthesis Lectures on Information Concepts, Retrieval, and Services*. Morgan & Claypool Publishers, San Rafael, CA, USA, 2009. https://doi.org/10.2200/S00190ED1V01Y200904ICR005.

[150] Jouni Tuominen, Eero Hyvönen, and Petri Leskinen. Bio CRM: A data model for representing biographical data for prosopographical research. In Antske Fokkens, Serge ter Braake, Ronald Sluijter, Paul Arthur, and Eveline Wandl-Vogt, editors, *Proceedings of the Second Conference on Biographical Data in a Digital World 2017 (BD2017)*, pages 59–66, Linz, Austria, 2017. CEUR Workshop Proceedings, vol. 2119. http://ceur-ws.org/Vol-2119/paper10.pdf.

[151] Kim Viljanen, Teppo Kansala, Eero Hyvönen, and Eetu Makela. ONTODELLA–a projection and linking service for semantic web applications. In *17th International Workshop on Database and Expert Systems Applications (DEXA'06)*, pages 370–376. IEEE, 2006. https://doi.org/10.1109/DEXA.2006.105.

[152] Kim Viljanen, Jouni Tuominen, Teppo Känsälä, and Eero Hyvönen. Distributed semantic content creation and publication for cultural heritage legacy systems. In *Proceedings of the 2008 IEEE International Conference on Distributed Human-Machine Systems*. IEEE, March 9-12 2008.

[153] V. Viswanathan and K. Ilango. Ranking semantic relationships between two entities using personalization in context specification. *Information Sciences*, 207:35–49, 2012. https://doi.org/10.1016/j.ins.2012.04.024.

[154] Nikos Voskarides, Edgar Meij, Manos Tsagkias, Maarten De Rijke, and Wouter Weerkamp. Learning to explain entity relationships in knowledge graphs. In Chengqing Zong and Michael Strube, editors, *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 564–574, 2015. https://doi.org/10.3115/v1/P15-1055.

[155] Anna Warsell. *Sisällissodan arvet maalaispitäjässä: Ruovesi vuosina 1918 - 1930*. Doctoral dissertation, University of Tampere, Faculty of Social Sciences, June 2024. https://urn.fi/URN:NBN:fi:tuni-202405316535.

[156] Anna Wessman, Suzanne Thomas, and Ville Rohiola. Digital archaeology and citizen science:: Introducing the goals of FindSampo and the SuALT project. *SKAS*, 2019(1):2–17, 2019.

[157] Anna Wessman, Suzie Thomas, Pieterjan Deckers, Andres S. Dobat, Stijn Heeren, and Michael Lewis. Hobby metal-detecting as citizen science. background, challenges and opportunities of collaborative archeological finds recording schemes. *Heritage & Society*, 16(2):89–108, 2023. https://doi.org/10.1080/2159032X.2022.2098654.

[158] Anna Wessman, Suzie Thomas, Ville Rohiola, Mikko Koho, Esko Ikkala, Jouni Tuominen, Eero Hyvönen, Jutta Kuitunen, Helinä Parviainen, and Marianna Niukkanen. Citizen science in archaeology: Developing a collaborative web service for archaeological finds in Finland. In John Jameson and Sergiu Musteață, editors, *Transforming Heritage Practice in the 21st Century: Contributions from Community Archaeology*, pages 337–352. Springer, Cham, July 2019. https://doi.org/10.1007/978-3-030-14327-5_23.

[159] Lars Westerlund, editor. *Sotaoloissa vuosina 1914-22 surmansa saaneet: tilastoraportti*. Valtioneuvoston kanslian julkaisusarja. Valtioneuvoston kanslia, Helsinki, 2004. http://urn.fi/URN:ISBN:952-5354-52-0.

[160] David Wigg-Wolf, Anna-Lisa Pfeiffer, and Karsten Tolle. Aggregating coin find data to the ARIADNE portal. challenges of using the Getty AAT for a specialist domain. *Internet Archaeology*, 64, 2023. https://doi.org/10.11141/ia.64.6.

[161] Mark Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip Bourne, et al. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1):1–9, 2016. https://doi.org/10.1038/sdata.2016.18.

[162] Florian Windhager, Paolo Federico, Günther Schreder, Katrin Glinka, Marian Dörk, Silvia Miksch, and Eva Mayr. Visualization of cultural heritage collection data: State of the art and future challenges. *IEEE Transactions on Visualization and Computer Graphics*, 25(6):2311–2330, 2018. https://doi.org/10.1109/TVCG.2018.2830759.

[163] Marcia Lei Zeng. Knowledge organization systems (KOS). *Knowledge Organization (KO)*, 35(2-3):160–182, 2008. https://doi.org/10.5771/0943-7444-2008-2-3-160.