

Digital Humanities on the Semantic Web: Making a Vision Come True

DARIAH Aalto Day, Nov 30, 2022

Eero Hyvönen

Professor, Department of Computer Science, Aalto University

Director, Helsinki Centre for Digital Humanities (HELDIG), University of Helsinki

<https://seco.cs.aalto.fi/u/eahyvone/>




Contents

1. **The Vision:** 4 Generations of Publishing Data for Humanities and Social Sciences
2. **From Vision to Reality:** How to Build a National Linked Open Data Infrastructure for Digital Humanities and to Use in on the Semantic Web
 - 20 years of work at the Aalto University and University of Helsinki (HELDIG)

A pair of hands is shown holding a clear crystal ball. The crystal ball reflects a bright blue sky with fluffy white clouds. The background of the entire image is a real sky with similar clouds and a glimpse of a blue ocean at the bottom. The text "1. Vision" is centered over the crystal ball.

1. Vision

We Aim at a Paradigm Change: 4 Generations of Publishing Data for Humanities

1. Texts (Engravings, Hand-written, and Printed)
 2. Online Systems for Searching and Exploring
 3. Publishing Content as Linked Data with Tools for DH
 4. Automatic Knowledge Discovery and Artificial Intelligence
- Our Focus
- 
- The background of the slide features a silhouette of human evolution, starting from a monkey-like creature on the left and progressing through three intermediate stages to a modern, upright human on the right. The scene is set against a vibrant sunset or sunrise sky with shades of orange, red, and purple. The silhouettes are black and are positioned in the lower half of the slide, behind the list of generations.

Problem of generations 1 and 2:
Content is available as texts for humans to
read but not as data for computers!

DATA

Our Approach:

Publishing and Using Cultural Heritage Linked Open **Data** in Digital Humanities



Why Linked Open Data?

- Enriching everybody's data collaboratively from separate silos
 - Everybody wins by collaboration!
- Creating Findable, Accessible, Interoperable, Re-usable data
 - The value of data increases!
- Creating more intelligent applications for the public, curators, and researchers
 - The machine “understands” linked data!



FAIR



<https://www.go-fair.org/fair-principles/>



A close-up, low-angle shot of the upper portion of Leonardo da Vinci's Vitruvian Man. The figure's face and long, flowing hair are illuminated by a warm, golden light, contrasting with the dark, shadowed background. The figure's head is tilted slightly to the right, and the eyes are looking forward. The texture of the hair and the contours of the face are clearly visible.

“Learn how to see. Realize that everything connects to everything else.”

Leonardo da Vinci (1452-1519)



3. From Vision to Reality

How to Build a National Linked Open Data
Infrastructure for Digital Humanities and to Use in on
the Semantic Web





Elements of a National Semantic Data Infrastructure for Open Cultural Heritage FAIR Data

Applications

Sampo-systems, Finna, Europeana, ARIADNEplus, TimeMachine...

Ontologies
Ontology services
Finto.fi

Datasets
Data services
LDF.fi

**Human
Infrastructure**



Linked Open Data Infrastructure for Finnish CH Cloud

Standards
Best practices
W3C, ...

Shared data models
DC, CIDOC-CRM, ...

Software tools

As of September 2010

Contents

- Human Infrastructure
- Ontology Infrastructure
- Linked Open Data Service Infrastructure
- Application Infrastructure: Sampo portals
- Tooling Infrastructure
- Dissemination of Results
- Organizational Infrastructure: Exit Plan

Human Infrastructure

Human Infrastructure Support

- Open Learning Materials
 - “Linked Data Technologies for Cultural Heritage and Digital Humanities: Introducing the Semantic Web in Video Lectures”
 - » 23 learning videos on LOD technology and CH applications for DH people
 - » <https://seco.cs.aalto.fi/teaching/sw-introduction/index.html>
 - Used at University of Helsinki DH course, Aalto University Semantic Web course, possibly at Digital Humanities at Oxford Summer School 2023, ...

Linked Data Technologies for Cultural Heritage and Digital Humanities: Introducing the Semantic Web in Video Lectures

Eero Hyvönen, Aalto University and University of Helsinki
Semantic Computing Reserch Group (SeCo), Helsinki Centre for Digital Humanities (HELDIG)

Introduction: Paradigm Change in Publishing and Using Cultural Heritage Data

A paradigm shift is taking place in publishing and using Cultural Heritage (CH) content on the Web: in addition to publishing materials for the humans to read as usual, also the underlying data is published as Linked Open Data (LOD) using the FAIR principles. This facilitates

1. publishing ever larger aggregated datasets,
2. enriching the Big Data with other publishers' data and by reasoning,
3. development of more intelligent interfaces and applications for the end users, and
4. using the data in Digital Humanities (DH) research for data analyses.

As a result, CH and DH have grown into and important application field of Semantic Web (SW) technologies.

Instead of a textual tutorial, this site provides an introduction based on videos, which hopefully lowers the barrier of getting grips with the opportunities and practical technical ideas underlying the paradigm change. The learning materials published in this online course are targeted to researches, collection managers and curators in memory organizations, students, and applications interested in the opportunities and challenges related to creating, aggregating, publishing, and using CH LOD on the SW.

The course first presents the basic ideas underlying the Semantic Web as a megatrend on the World Wide Web evolving into a Web of Data (Lecture 1). After this the fundamental idea of representing data as semantic networks using the RDF model and language are presented (Lecture 2). The topic of publishing LOD on the Semantic Web and using it by querying is the topic of Lecture 3. At the heart of the Semantic Web lay ontologies, the "silver bullet" of the Semantic Web. They define the concepts and data models based on which the Web of Data is built (Lecture 4). Another key idea on the SW is the idea of using logical rules and reasoning for enriching the data (Lecture 6). From a practical point of view, the SW needs an ontology and data infrastructure that could be created in interoperable and re-usable forms in a cost efficient way. (Lecture 7) Finally, practical applications of SW in creating and using LOD services and semantic portals are discussed using work in Finland since 2001 as a case study. This includes issues on building a national semantic web infrastructure that is interoperable with infrastructure developments on the WWW based on Semantic web standards of the World Wide Web consortium (W3C) (Lecture 8).

The learning materials are published under the open Creative Commons CC BY 4.0 license; you can refer to the course as *Eero Hyvönen: Linked Data Technologies for Cultural Heritage and Digital Humanities: Introducing the Semantic Web in Video Lectures* [Online course materials] <https://seco.cs.aalto.fi/teaching/sw-introduction/>

Lecture 1: Introduction: From WWW of today to the Semantic Web

This lecture first gives an overview of the "traditional" view of the WWW and its technologies in use today:



[PDF Slides](#)

Based on the challenges of the traditional WWW identified, a motivating introduction to why the SW is needed is then presented:



[PDF Slides](#)

Lecture 2: Resource Description Framework: the Foundation of SW

Resource Description Framework, including the RDF and RDF Schema specifications standardized by the W3C organization, form the backbone for representing data and knowledge on the Semantic Web, the Web of Data.

RDF Resource Description Framework

The video below describes the RDF data model and language:



<https://seco.cs.aalto.fi/teaching/sw-introduction/index.html>

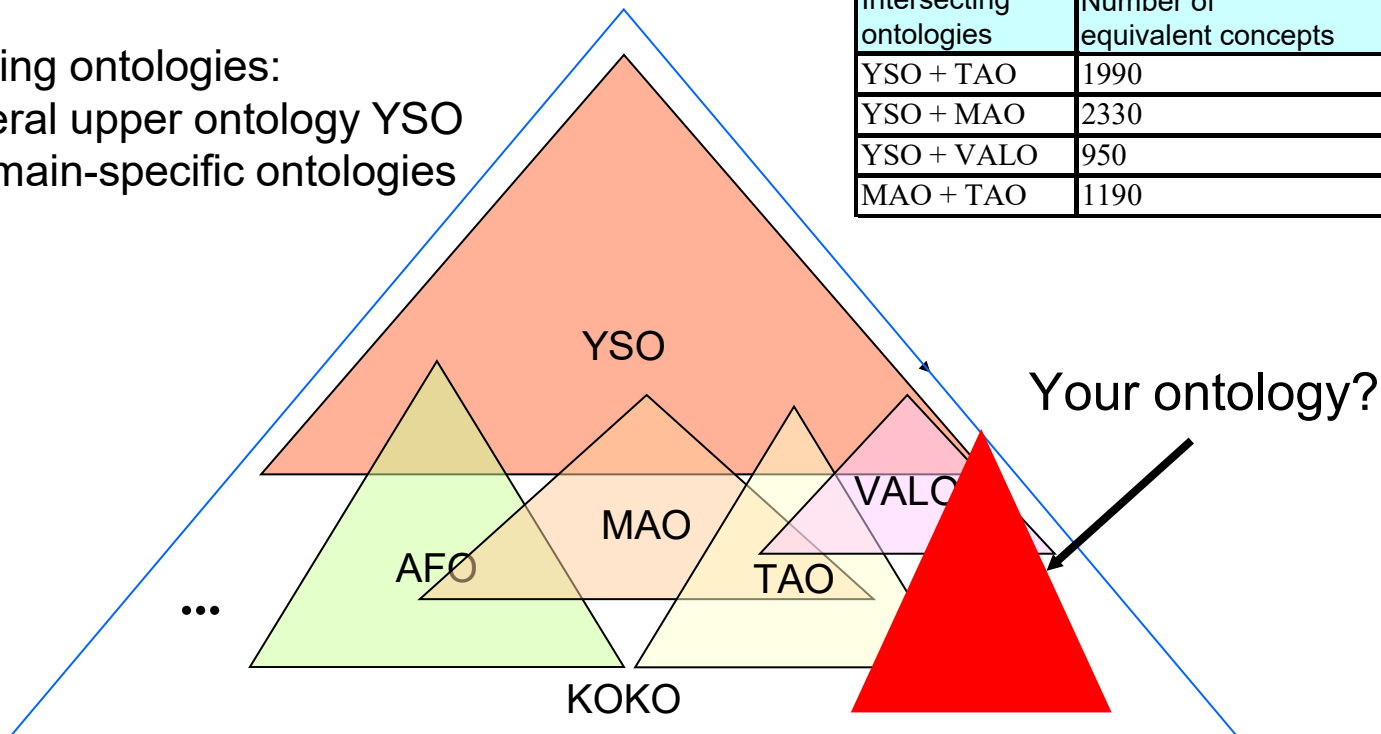
Ontology Infrastructure

National Linked Ontology Cloud:

Holistic Collaborative Finnish Ontology KOKO

Aligning ontologies:
General upper ontology YSO
+ domain-specific ontologies

Intersecting ontologies	Number of equivalent concepts
YSO + TAO	1990
YSO + MAO	2330
YSO + VALO	950
MAO + TAO	1190



Ontology Service Infrastructure Deployed

- Based on FinnONTO Research Initiative 2003-2012
- Ontology services developed: [ONKI.fi](http://onki.fi) & light.onki.fi
 - Many external users, e.g. memory institutions
- > National open ontology service Finto.fi by the National Library in 2014

The screenshot displays the ONKI ontology service interface. At the top, the 'ONKI' logo is on the left, and navigation links for 'Vocabularies', 'About', 'Feedback', and 'Help' are on the right, along with language options 'suomeksi' and 'på svenska'. Below the header, a teal bar contains 'EKS subject headings' and a search bar with 'Content language' set to 'English'. The main area is divided into a left sidebar and a right content pane. The sidebar has tabs for 'Alphabetical', 'Hierarchy', and 'Groups'. Under 'Alphabetical', there are two rows of letter links: 'A Å B C D E F G H I J K L M N' and 'O P Q R S T U V W X Y Z !*'. A scrollable list of subject headings is visible, including 'Aalto University', 'abandonment', 'abbreviations', 'abortion', 'absenteeism', 'abstract publications', 'Academic Carelian Society', 'Academy of Finland', 'Academy of Turku 1640', 'access to justice', 'access to public documents', 'accessibility', 'accident compensation', 'accident insurance', 'accident investigation', 'accident pension', and 'accidents'. The right pane shows details for the concept 'digital divide'. It includes a 'PREFERRED TERM' section with the term 'digital divide' and a copy icon. Below this is a 'BELONGS TO GROUP' section with a list: 'Political participation', 'Political science, politics', and '30.08.02 Structure of society and political change'. The 'TYPE' section shows 'eksuu03'. The 'DATE' section shows '10/14/03'. The 'IN OTHER LANGUAGES' section shows 'digitaalinen kuilu' (Finnish) and 'digital klyfta' (Swedish). The 'URI' section shows 'http://ldf.fi/eks/digitaalinen-kuilu' with a copy icon. At the bottom, the 'DOWNLOAD THIS CONCEPT' section shows 'RDF/XML TURTLE JSON-LD'.

ONKI Vocabularies About Feedback Help | suomeksi på svenska

EKS subject headings Content language English Search

Alphabetical Hierarchy Groups

A Å B C D E F G H I J K L M N
O P Q R S T U V W X Y Z !*

Aalto University
abandonment
abbreviations
abortion
absenteeism
abstract publications
Academic Carelian Society
Academy of Finland
Academy of Turku 1640
access to justice
access to public documents
accessibility
accident compensation
accident insurance
accident investigation
accident pension
accidents

PREFERRED TERM digital divide

BELONGS TO GROUP
Political participation
Political science, politics
30.08.02 Structure of society and political change

TYPE eksuu03

DATE 10/14/03

IN OTHER LANGUAGES
digitaalinen kuilu Finnish
digital klyfta Swedish

URI http://ldf.fi/eks/digitaalinen-kuilu

DOWNLOAD THIS CONCEPT: RDF/XML TURTLE JSON-LD

Linked Open Data Service Infrastructure: Linked Data Finland LDF.fi

Linked Data Finland

"7-star" model and LDF.fi data hotel

Goals: enhance re-usability and data quality

7-star Linked Data Service

However, in our opinion, providing 5-star Linked Data is just the beginning. To actually make use of the datasets, consumers need more support in getting to know and access them, as well as a better grasp of their quality and provenance. To this end, we extend the model with two additional stars:



Provide your data with a schema and documentation so that people can *understand and re-use* your data easily.



Validate your data and denote its provenance so that people can *trust the quality* of your data.

This added support should come with as little extra work as possible to the data publisher. Our hypothesis is that a lot of this can be done automatically, basing on the Linked Data core. A data publisher needs only to provide their data in the RDF format, and the LDF.fi portal will do the rest automatically. See the [overview paper](#) (in ESWC 2014 Proceedings, Springer-Verlag) for some more details about the underlying ideas.



Burj Al Arab

Application Infrastructure: Sampo Model, LOD Services & Portals

Cultural Heritage "Samos" in use on the Semantic Web 2004-

1. **MuseumFinland** – Finnish Museums on the Semantic Web (2004) [39 000 users]
2. **CultureSampo** – Finnish Culture on the Semantic Web (2008) [107 000 users]
3. **TravelSampo** - Mobile Contextualized Services of Cultural Tourism (2011)
4. **BookSampo** – Finnish Fiction Literature on the Semantic Web (2011) [2 million annual users today]
5. **WW1LOD** – World War I Linked Open Data (2014)
6. **WarSampo** – Finnish World War 2 on the Semantic Web (2015-19) [1 million users by now]
7. **Norssi Alumni on the Semantic Web** – Historical person registry using LOD (2017)
8. **U.S. Congress Prosopographer** – U.S. Congress Legislators 1789-2018
9. **BiographySampo** - Finnish Biographies on the Semantic Web (2018-20) [50 000 users]
10. **NameSampo** – Linked Data Workbench for Toponomastic Research (2019) [37 000 users]
11. **WarVictimSampo 1914-1922** – National War History [29 000 users]
12. **Mapping Manuscript Migrations (MMM)** – medieval and Renaissance manuscripts (2020) [9 100 users]
13. **AcademySampo** – Finnish Academic People 1640 – 1899 (2021) [8 200 users]
14. **FindSampo** – Archaeological Finds on the Semantic Web (2021) [3 100 users]
15. **WarMemoirSampo** – WW2 Veteran Stories (2021)
16. **LetterSampo** – Historical Letters on the Semantic Web (2022)
17. **CoinSampo, LawSampo, ParliamentSampo, ...** (2023)



Defense of Sampo, Ateneum,
A. Gallen-Kallela, 1896



E Hyvönen: [Digital Humanities on the Semantic Web: Sampo Model and Portal Series](#).
Semantic Web Journal, 2022. [pdf](#)

<https://seco.cs.aalto.fi/applications/sampo/>



New Use Cases & “Sampos” Underway/Planned in 2022



“BookSampo Revisited 2022”



“Operasampo”



National Audiovisual Institute

Tooling Infrastructure

Sampo-UI

Knowledge Extraction Tools

Dissemination of Results

Dissemination Presentations in 2022 in Conferences, Keynotes, EU Commission Work

- 26 presentations, 47 publications in 2022
- Keynotes
 - TEXT2KG 2022, 30.5., Hersonissos, Greece
 - MODELSEN 2022, 7.8.2022, Berlin, Germany
 - HOUSE Final Conference, 8.9.2022, Muenster, Germany
 - SKILLNET 2022 Conference, 17.11.2022, Utrecht, Netherlands
 - ANRDM 2022 Conference, 26.11.2022, Seoul, South Korea,
- Participation in the EU Commission Expert Group ECCCH
 - Pere Brunet, Livio de Luca, Eero Hyvönen, Adeline Joffres, Peter Plassmayer, Martijn Pronk, Roberto Scopigno and Gabor Sonkoly: [Report on a European Collaborative Cloud for Cultural Heritage. Ex-ante Impact Assessment](#). European Commission, Directorate-general for Research and Innovation, March, 2022. 108 p. [link](#)
 - 110 MEUR EU research programme? Call is open

Organizational Infrastructure: Exit Plan

Developing a Sustainable Solution for National Linked Open **Data Services** & **Sampo Systems**

- **Challenge:** Transition from research prototypes to legacy services of data owners

	Nimi	Julkaistu	Käyttäjiä	Koko	Datan pääomistajat
1	Kirjasampo	2011	2M vuosittain	4.4M	Yleiset kirjastot
2	Sotasampo	2015	1M	14M	Kansaliisarkisto, puolustusvoimat
3	Nimisampo	2019	40 000	26M	Kotus, Maanmittauslaitos
4	Biografiasampo	2019	50 000	5.6M	SKS
5	Sotasurmasampo	2019	30 000	10M	Kansallisarkisto
6	Akatemiasampo	2021	9 000	6.6M	HY:n arkisto, Kansallisarkisto
7	Sotamuistosampo	2021	3 000		Kansallisarkisto
8	Löytösampo	2021	4 000	1M	Museovirasto
9	Lakisampo	2022?			Oikeusministeriö
10	Parlamenttisampo	2022?			Eduskunta

Exit Plan: Linked Open Data Center Linda?

- National service provided by Sampo data owners
 - Data services : Linked Data Finland <https://ldf.fi>
 - Applications: Sampo series of applications
 - Development: Research projects with universities
- Model similar to deploying ontology services:
 - ONKI.fi -> Finto.fi
- Needed resources
 - Server infra: CSC – IT Centre for Science (Ministry of Education and Culture)
 - Linked Data Finland – maintenance:
 - LDF.fi & Finto.fi are based on the same technology
 - Sampo system maintenance
 - » Data maintenance: data owners' expertise is needed
- Discussions started with key stake holders
 - National Archives, National Library, FHA, SKS, Kotus, ...

More Information

- FIN-CLARIAH Linked Open Data Services
<https://seco.cs.aalto.fi/projects/fin-clariah/>
- Sampo Series of Linked Open Data Services and Portals
<https://seco.cs.aalto.fi/applications/sampo/>
- 500 Publications of SeCo Group online
<https://seco.cs.aalto.fi/publications/>