

Kirjasampo 2.0:n julkistustilaisuus, Aalto-yliopisto ja Zoom, 10.10.2023, Helsinki

Miksi Kirjasampo 2.0 taottiin: Sampo-mallin esittely

Eero Hyvönen

Aalto University, Dept. of Computer Science, professor
University of Helsinki, Helsinki Centre for Digital Humanities (HELDIG), director

Semantic Computing Research Group (SeCo)

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



Sisältö

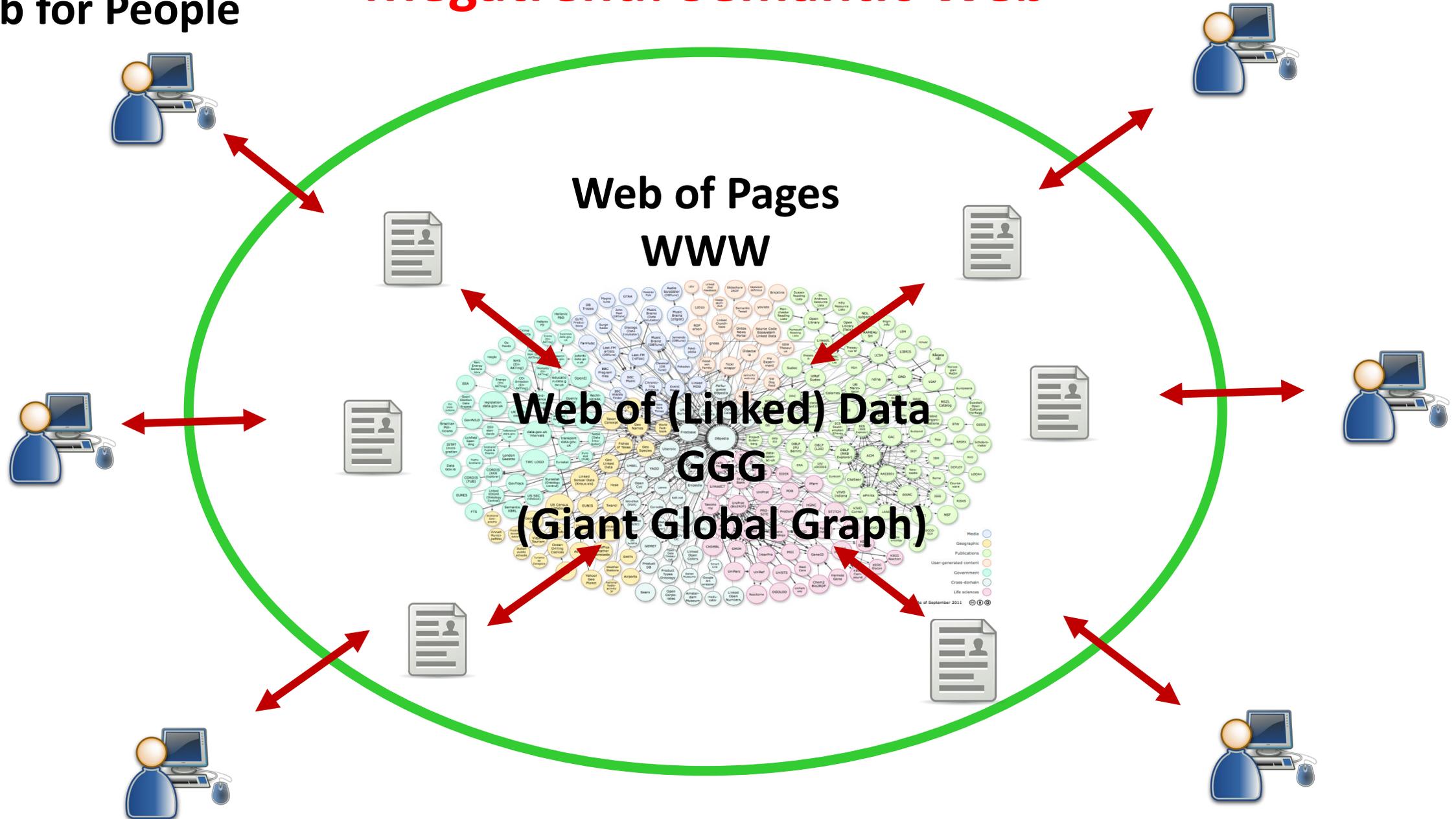
- 1. Historiaa 2003-2023: FinnONTOsta Sampo-sarjaan**
- 2. Sampo-malli: datapalvelut ja semanttiset portaalit**
- 3. Päivän ohjelma**

1. Historiaa 2003-2023:

FinnONTO:sta Sampo-sarjaan

Megatrend: Semantic Web

Web for People



Development of Semantic Web and Reactions in Finland

Phases of SW

Ontologies

Linked Data

Knowledge Graphs

AI

Infrastructure

- Domain ontology services
- Linked open Data services

Semantic Web Kick-off
in Finland

ONKI.fi released

Finto.fi released

LDF.fi released

Sampo-UI Framework

FinnONTO

Sampo-projects

DARIAH-FI

Major projects

Linked Data Finland

2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023

Building a national **ontology infrastructure**

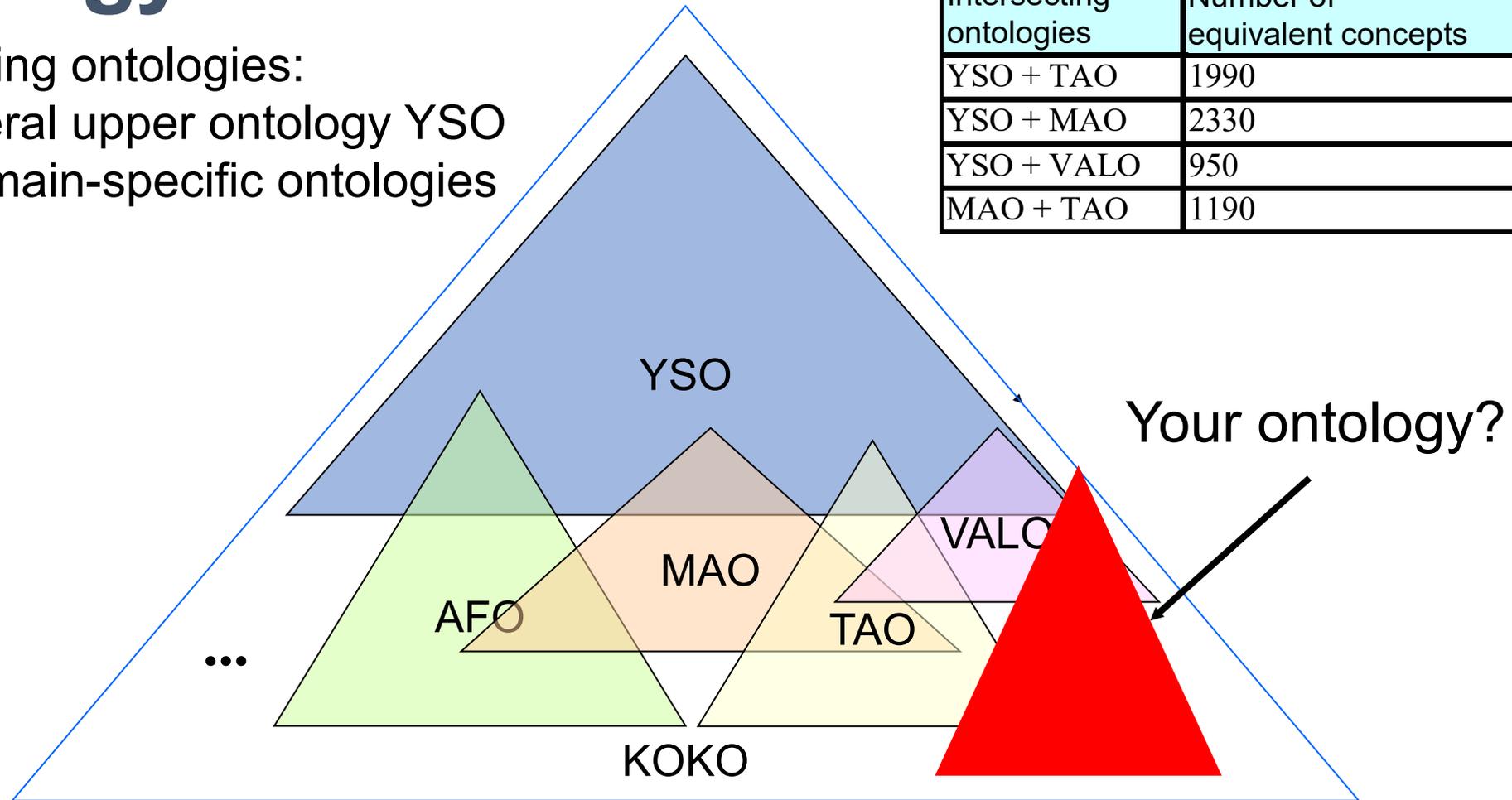
FinnONTO 2003-2012

	Name	Ontology domain	Underlying thesaurus	Size	Maintaining Organization
1	YSO	General domain	General Finnish Thesaurus, YSA, Allärs	23700	National Library, Åbo Academy
2	MUSO	Music	Thesaurus of Music, MUSA/CILLA	1000	National Library
3	MAO	Museum domain	Thesaurus of Museum Domain, MASA	6800	National Board of Antiquities
4	AFO	Agriculture, forestry	Agriforest Thesaurus	5500	Viikki Science Library
5	TAO	Applied arts	Thesaurus of Applied Arts	2600	University of Eastern Finland and Library of Aalto University
6	VALO	Photography	Thesaurus of Photography Literature, Thesaurus of Photography Technology	1900	Finnish Museum of Photography
7	MERO	Seafaring, shipping	Thesaurus of Seafaring	1400	Finnish Transport Agency
8	KAUNO	Literature subjects	Thesaurus of Literature, Bella	4900	Finnish Public Libraries, Kirjastot.fi
9	JUHO	Public government	Thesaurus of Finnish Government, VNAS	6400	Ministry of Finance
10	TERO	Health promotion	YSA, TESA, MoSH, Stameta	22000	Various organizations
11	KITO	Literature research	Thesaurus of Literature Research	900	Finnish Literature Society
12	KULO	Culture research	Thesaurus for Folk Culture Studies	1600	Finnish Literature Society
13	KTO	Linguistics	Thesaurus of Linguistics	1000	Research Institute for the Languages in
14	PUHO	Defense	Thesaurus of Defence Administration	2000	Finnish Defence Forces
15	POIO	Points of interest	TGN, Geonames, LDG, SUO	4600	Various organizations
	TOTAL			86300	

Case: 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



[Hyvönen et al., ESWC 2009]

KOKO ontologies and ONKI service deployed January 2014 by the National Library as Finto

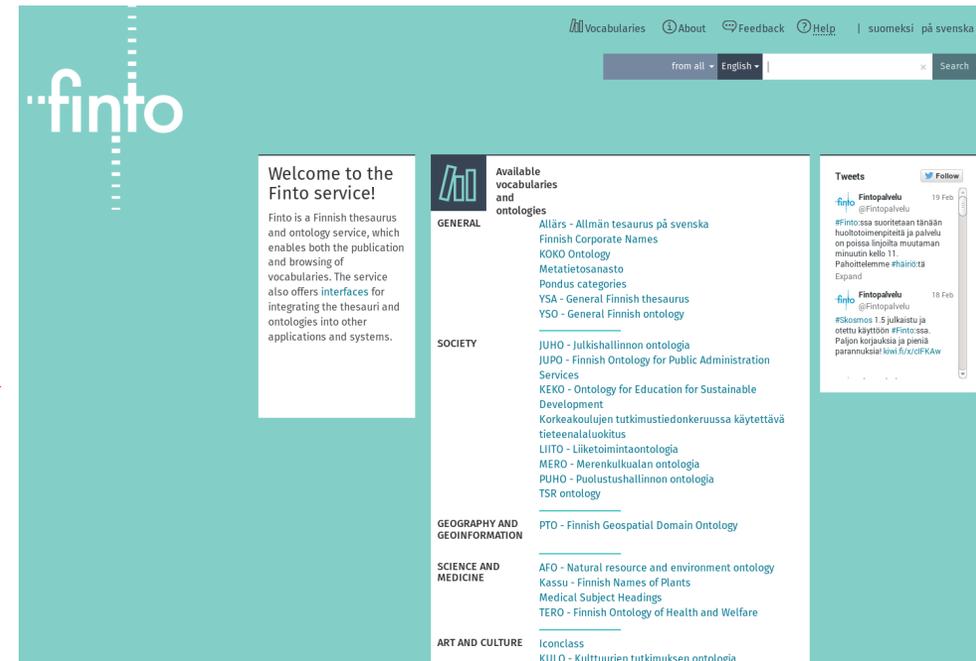
Permanent free national service funded by Finnish ministries

2019: 32 million API calls
2020: 2,1 million page visits
2020: 12% increase in use



Available vocabularies and ontologies

- General concepts**
 - General Finnish thesaurus
 - General Finnish upper ontology
 - KOKO Ontology
 - Allmän tesauros på svenska - Allärs
 - AFO Ontology
 - Ontology of Applied Arts - TAO
 - Ontology for Museum Domain
 - Library of Congress Subject Headings - LCSH
- Health**
 - Finnish Ontology of Health and Welfare
- Culture**
 - Kaunokki Ontology
 - The Finnish Ontology of Photography VALO
 - Kulttuurien tutkimuksen ontologia - KULO
 - Finnish Music Thesaurus - MUSA/CILLA
 - Kielitieteen ontologia - KTO
 - Kirjallisuudentutkimuksen ontologia - KITO
- Public Administration**
 - Puolustushallinnon ontologia - PUHO
 - Julkishallinnon ontologia - JUHO
 - Kunnat 2011
 - Julkisten palveluiden ontologia JUPO
 - Schools Online Thesaurus (ScOT)
- Business**
 - Merenkulkualan ontologia - MERO
- Science**
 - Kassu - Kasvien suomenkieliset nimet



Building a **Linked Open Data** Infrastructure

Linked Data Finland 2012-2014



[Home](#)
[Project](#)
[Datasets](#)
[Search Data](#)
[Schemas](#)
[Services](#)
[Policies](#)
[Documentation](#)
[Validation](#)
[Linked Data Science](#)
[Applications](#)
[Your Data?](#)
[Linked Data School](#)

Linked Data Finland

Living Laboratory Data Service for the Semantic Web

This site is the Living Laboratory of the [Linked Data Finland](#) research initiative, conducted by the [Semantic Computing Research Group](#) at [Aalto University](#) in collaboration with University of Helsinki and a large consortium of Finnish public organizations and companies.

Our goal is to make life easier for both publishers as well as consumers of structured data on the Web. We base our work on the [Linked Data](#) paradigm and stack of standards, which combines an expressive, semantic data model ([RDF](#)) with standardized access mechanisms ([SPARQL](#) and [live HTTP URIs](#)).

5-star Linked Data

The baseline of our work is the [5-star Linked Data model](#), proposed [originally](#) by Tim Berners-Lee.

- ★ Make data available on the Web in whatever format.
- ★★ Make data available as structured data (e.g., Excel instead of an image scan of a table).
- ★★★ Use non-proprietary formats (e.g., CSV instead of Excel format).
- ★★★★ Use URIs to denote things, so that people can point at your data.
- ★★★★★ Link your data to other data to provide context.

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.

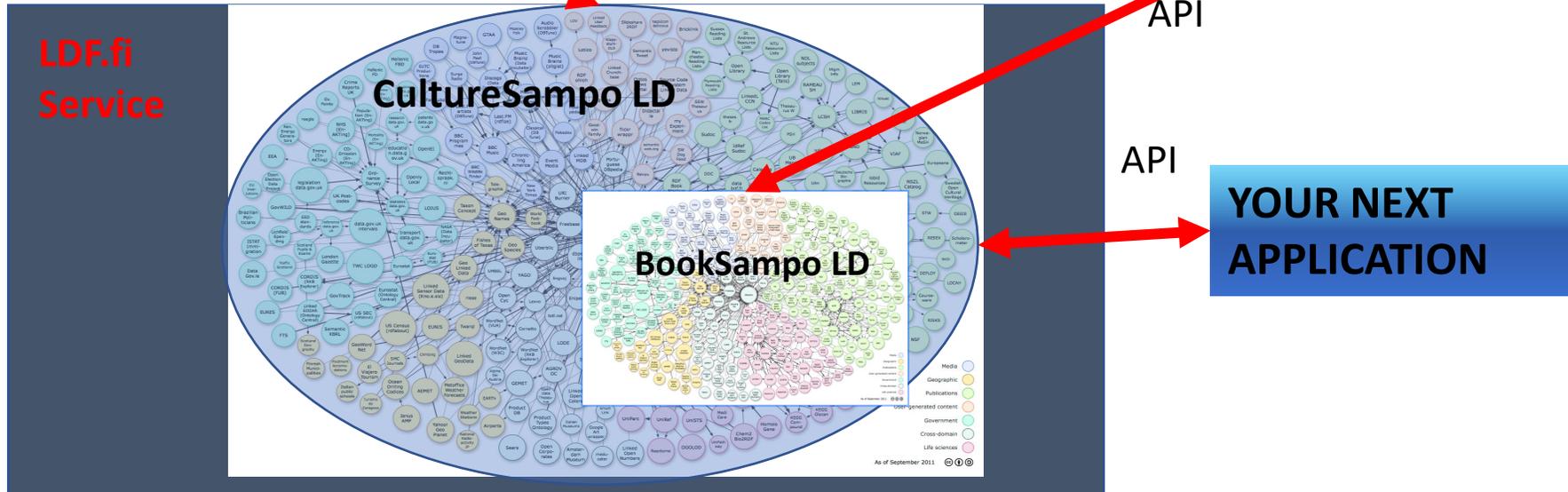
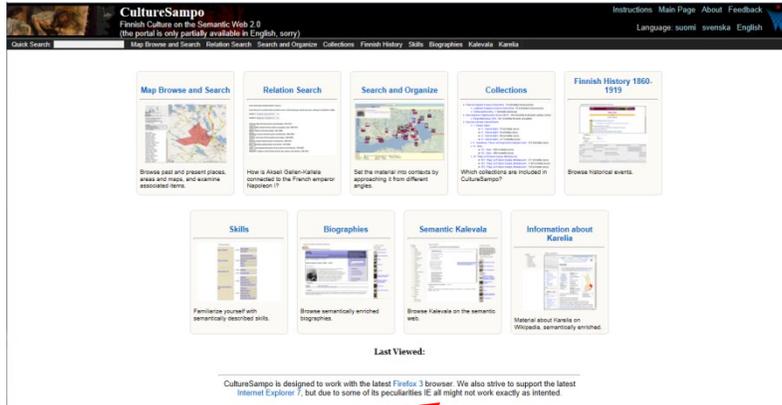
This added support should come with as little extra work as possible to the data publisher. Our hypothesis is that a lot of this can be done automatically, basing on the Linked Data core. A data publisher needs only to provide their data in the RDF format, and the LDF.fi portal will do the rest automatically.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

Further Information

On the left, you find more information about the project and the datasets we are working on. By selecting a

<https://ldf.fi>

Key Idea and Goal is Very Simple: Provide Linked Open Data Infra & Services for Applications



First Deployments by Libraries

- 2011 BookSampo deployed by Public Libraries
 - **Kirjasampo.fi**
 - 1.6 million annual users today
- 2014 ONKI ontology services deployed by National Library
 - **Finto.fi**
 - Tens of millions of annual API calls
- => 20+ Sampo Series of data services and semantic portals

Sampo Series of LOD Services & Portals



A. Gallen-Kallela: Defense of Sampo, 1896, Turku Art Museum,

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) [1.6 million users in 2022]
5. **WW1LOD** – World War I Linked Open Data (2014)
6. **WarSampo** – Finnish World War 2 on the Semantic Web (2015-19) [1 100 000 users]
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) [381 000 users]
10. **NameSampo** – Linked Data Workbench for Toponomastic Research (2019) [55 000 users]
11. **WarVictimSampo 1914-1922** – National War History [80 000 users]
12. **Mapping Manuscript Migrations (MMM)** – medieval manuscripts (2020) [9100 users]
13. **AcademySampo** – Finnish Academic People 1640 – 1899 (2021) [26 000 users]
14. **FindSampo** – Archaeological Finds on the Semantic Web (2021) [7 000 users]
15. **WarMemoirSampo** – Memoirs of Finnish WW2 veterans (2021) [3 800 users]
16. **LetterSampo** – Early Modern Letters on the Semantic Web (2022)
17. **LawSampo** – Finnish Legislation and Case Law on the Semantic Web (2023)
18. **ParliamentSampo** – Parliament of Finland on the Semantic Web (2023)
19. **BookSampo 2.0, OperaSampo, ConfermentSampo, CoinSampo, ArtSampo, ...** (2023-)

Sampo-UI

IN USE

Sampo = Mythical artifact of the Finnish Epic Kalevala that gives to its owner riches and good fortune. A *metaphore of amazing ancient technology*.

Links to all Sampos: <https://seco.cs.aalto.fi/applications/sampo/>

Linked Open Data Infrastructure for Cultural Heritage FAIR Data: FIN-CLARIAH



Applications

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

Ontologies
Ontology services
Finto.fi

Datasets
Data services
LDF.fi

Linked Open Data Infrastructure for Finnish CH Cloud

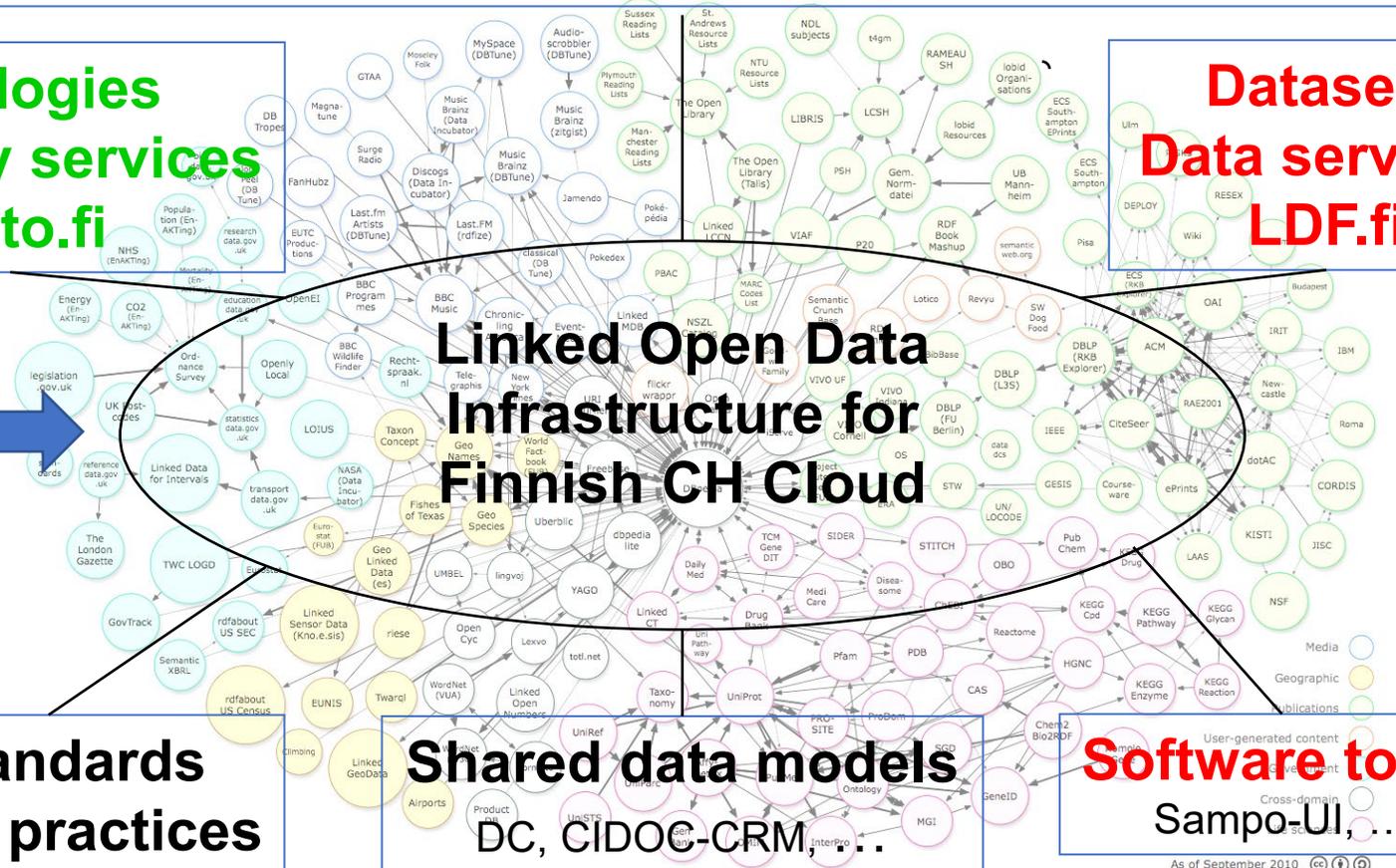
Human Infrastructure



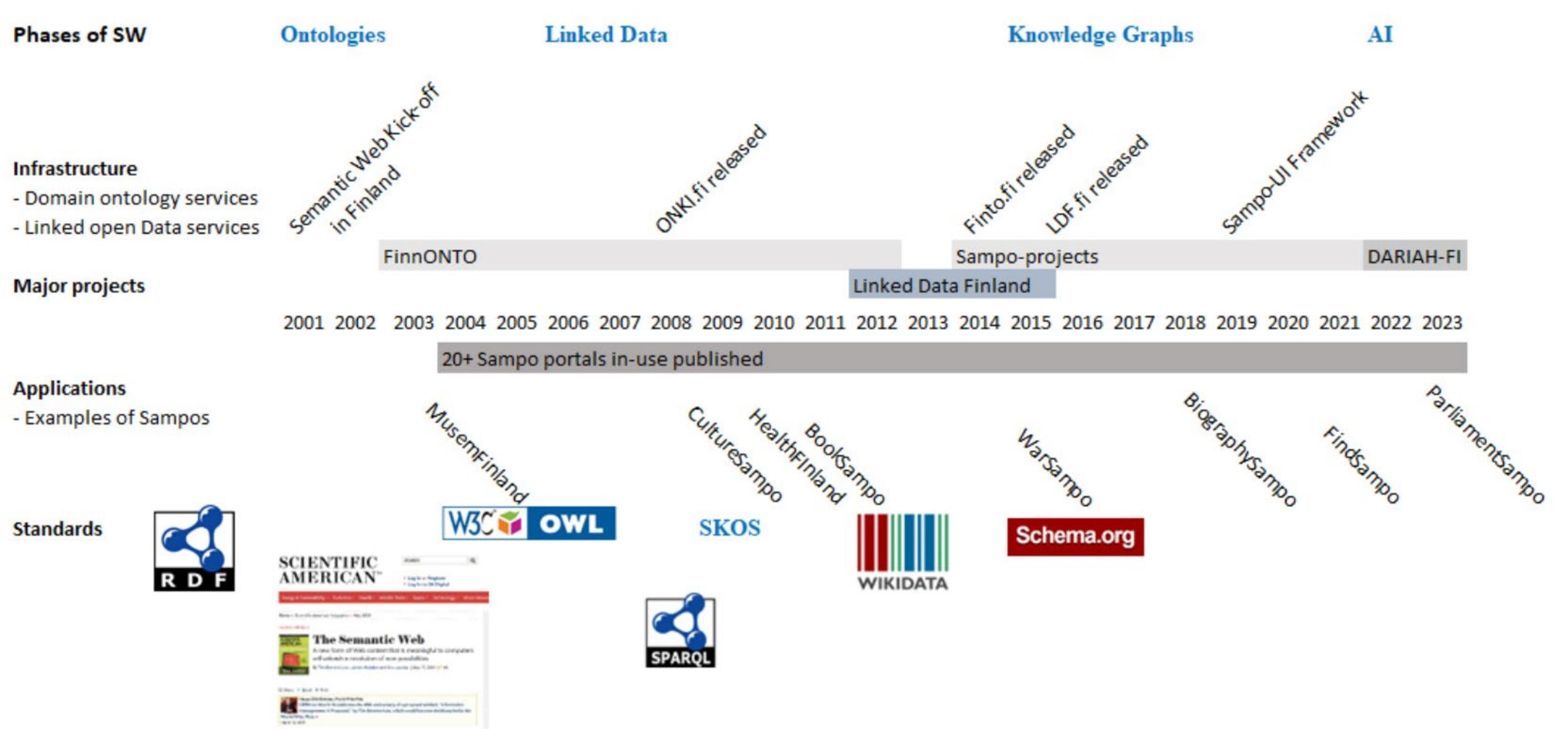
Standards
Best practices
W3C, ...

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

Software tools
Sampo-UI, ...



As of September 2010 © ⓘ ⓘ



Why infrastructure?

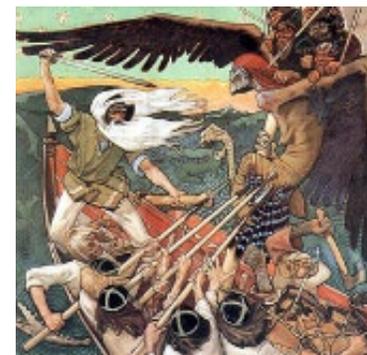
”Intellectuals solve problems
- geniuses prevent them”

Albert Einstein

2. Sampo-malli: datapalvelut ja semanttiset portaalit

Sampo Model Principles

Sampo Model Principles P1–P6



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 (UI)
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

Principles for creating LOD services

How to create a KG & data service

Principles for creating portal interfaces

How to create a KG –based portal UI

[Hyvönen, *Semantic Web*, 2023]

[Ikkala et al., *Semantic Web*, 2022]

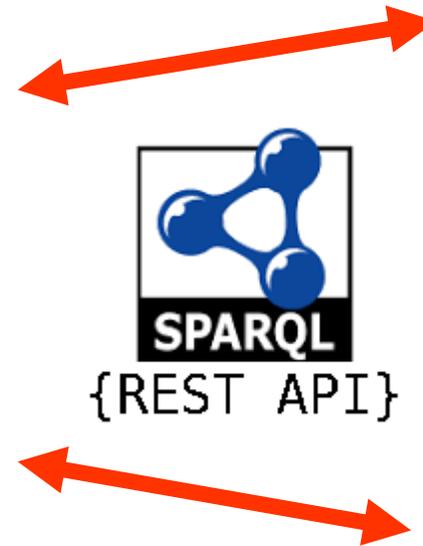
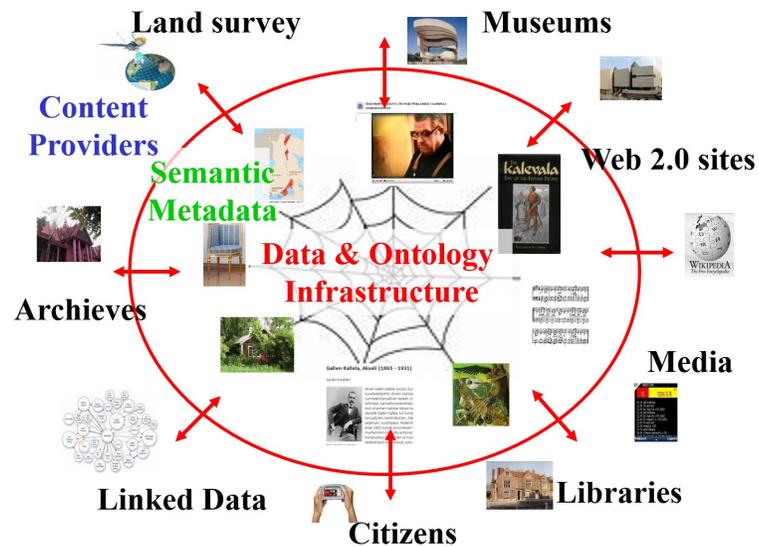


Eero Hyvönen: [Digital Humanities on the Semantic Web: Sampo Model and Portal Series](#).
Semantic Web – Interoperability, Usability, Applicability, vol. 14, no. 4, pp. 729-744, 2023.

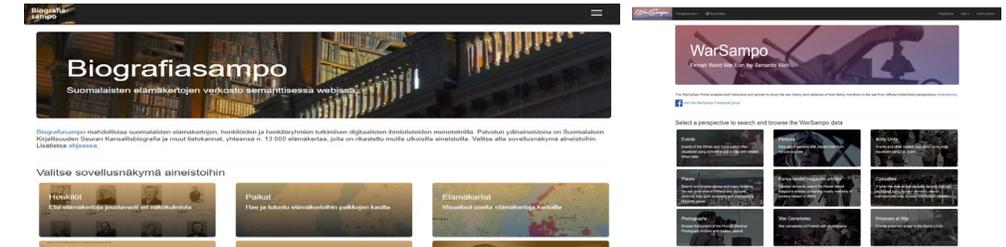
Using Sampo Model for Cultural Heritage Data by 1) human users and 2) machines



Sampo LOD Service



Sampo Portals & Applications



Digital Humanities Research Tools

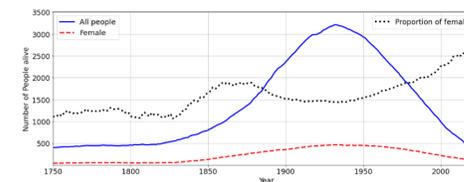


Fig. 3: Amount of people alive on timeline



Fig. 4: Average lifespan of biographees



Linked Data Finland Service:
<https://ldf.fi>

Kirjasampo 2.0

Semanttinen haku, selailu ja data-analyysit

Valitse näkymä etsiäksesi ja selataksesi aineistoja:

Romaanit

Näkymä, jossa voi hakea romaaneja.

Julkaisut

Näkymä, jossa voi hakea teosten julkaisujen tietoja.

Henkilöt

Näkymä, jossa voi hakea kirjallisuuteen liittyviä henkilöitä.

Kansikuvat

Näkymä, jossa voi hakea teosten kansikuvia.

Tietokirjat

Näkymä, jossa voi hakea tietokirjoja.

Kuvat käytössä Shutterstock.com:in lisensillä

<https://analyysi.kirjasampo.fi>

Miksi Kirjasampo 2.0?

- Täydentää Kirjasampo.fi:n käyttöliittymää Sampo-mallilla
 - Semanttinen fasettihaku
 - Semanttinen selailu
 - Data-analyttiset työkalut
- Julkaisee Kirjasammon linkitetyn datan palvelun tutkimuskäyttöön
 - Suomen Akatemian digitaalisten ihmistieteiden FIN-CLARIAH-ohjelmassa
 - Suomalaisen kirjallisuuden data-lähtöinen tutkimus
- Sampo-mallin ja Sampo-UI-järjestelmän edelleen kehittäminen
 - Uusia data-analyttisiä komponentteja
 - Opetusaineistot sampojen takojille
- Datan laadun validointi

Video: Semantic Web and AI for Digital Humanities

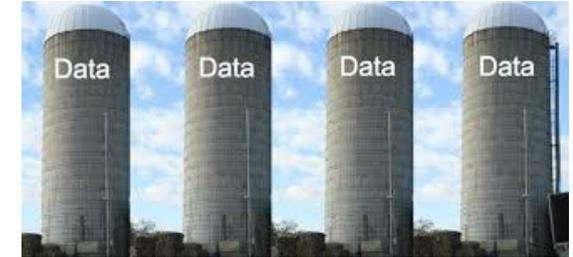
<https://vimeo.com/470313703>

Tuottanut Suomen Akatemia



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!

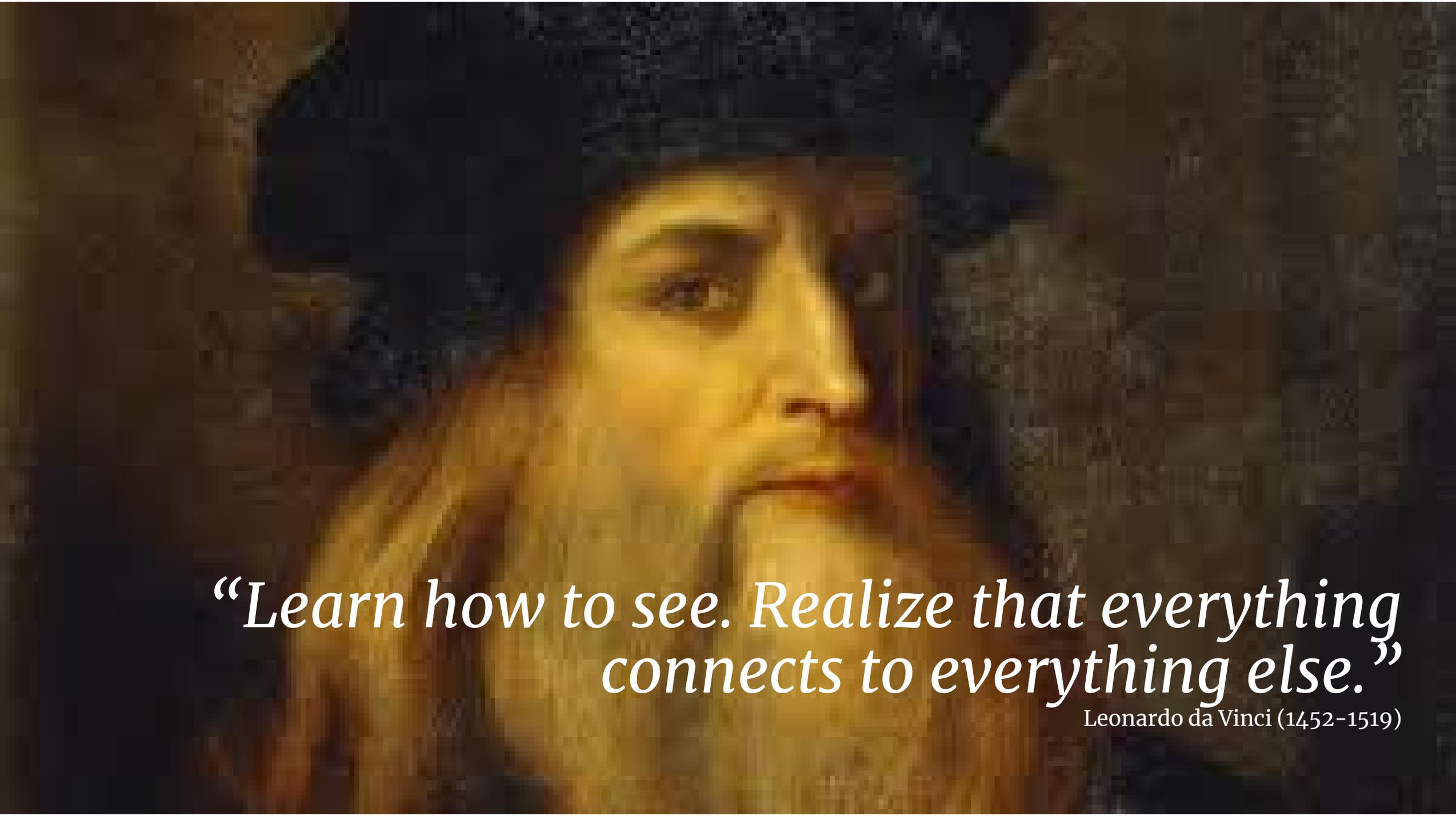


FAIR



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



A close-up, low-angle view of the upper portion of Leonardo da Vinci's 'Vitruvian Man' drawing. The figure's head and shoulders are visible, rendered in a golden-brown hue against a dark, textured background. The figure's right arm is raised, and the left arm is partially visible. The overall composition is dramatic and emphasizes the figure's form.

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

Leonardo da Vinci (1452-1519)

**We Have Created Sampos to prove
that Leonardo Da Vinci was right!**

"Sampo Series" Demonstrates a Paradigm Shift: 4 Generations of Publishing Data for Humanities

1. Printed Texts
2. Online Systems for Searching and Exploring
3. Publishing Content as Linked Data with Tools for DH
4. Automatic Knowledge Discovery and Artificial Intelligence

Kirjasampo 3.0?

- Suosittelujärjestelmän kehittäminen
- Datan rikastaminen muilla sammoilla, Wikipedialla jne.
- Knowledge discovery ja tekoäly
 - Kiinnostavien ilmiöiden automaattinen tunnistaminen
 - Tutkimusongelmien ratkaisemisen tukeminen, explainable AI
- Linkitys Helmet- ym. kirjastojärjestelmiin
 - Kun löytyy kiinnostava teos



HOME PAGE

MISSION

NEWS

PROJECTS

EVENTS

PEOPLE

PARTNERS

CAREERS

TEACHING

CONTACT

PUBLICATIONS

[APPLICATIONS](#)

ONTOLOGIES

LINKED DATA

INFRA FOR DH

SERVICES

TOOLS

SOFTWARE

DOWNLOAD

Wikipedia: [Semantic web suomeksi](#)

Google this site

Search!



Booksampo has been deployed by the Finnish public libraries

BookSampo - Finnish Fiction Literature on the Semantic Web

What is BookSampo?

BookSampo is a Linked Open Data (LOD) service and a semantic portal on top of it in use Finland. It is a member of the [Sampo series](#) of LOD services and semantic portals based on the Finnish Semantic Web infrastructure [FIN-CLARIAH](#).

The Booksampo Knowledge Graph (KG) covers metadata about practically all Finnish fiction literature available at Finnish public libraries on a work level. After its initial publication in autumn 2011, the Booksampo KG has been extended with additional content including, e.g., non-fiction literature, too. The system introduces a variety of semantic web novelties deployed into practise: The underlying data model is based on the functional, content-centered metadata indexing paradigm using RDF. Linked Data (LD) principles are used for mapping the metadata with tens of interlinked ontologies in the national [FinnONTO](#) ontology infrastructure. The contents are also linked with the large LD metadata repository of related cultural heritage content of CultureSampo. BookSampo was originally based on using the [CultureSampo - Finnish Culture on the Semantic 2.0](#) as its data service, demonstrating the idea of re-using semantic content from multiple perspectives without a need for modifications.

Public Portal on the Semantic Web

The portal has been online since autumn 2011 at: <http://www.kirjasampo.fi/>.

Booksampo was developed as part of the national [FinnONTO research project series 2003-2012](#) of Aalto University and University of Helsinki together with the help of partnering Finnish organizations in the research consortium, including the Finnish public libraries. The original portal interface was implemented at the Finnish public library consortium [Kirjasto.fi](#). The system is based on the fiction literature data [KAMINO](#) (developed from the Kaunokki thesaurus) and other ontologies developed in the FinnONTO-project, and metadata from library databases, biographies, titles, articles and other resources. The contents search and recommending services of the portal came through the APIs of the [CultureSampo system](#), that contained all semantic content of the system interlinked in RDF with other cultural materials.

Most of the BookSampo content was transformed automatically from existing databases, with the help of ontologies derived from thesauri in use in Finland, but in addition tens of volunteered librarians have participated in a Web 2.0 fashion in annotating and correcting the metadata using the [SAHA metadata editor](#) connected to the [ONKI](#) ontology services of FinnONTO, especially regarding older literature.

BookSampo Deployment by the Finnish Public Libraries

The BookSampo research prototype was deployed the public libraries and has been maintained independently by them since 2011. At the same time, the Booksampo KG was separated from the larger CultureSampo KG and is today hosted by a Fuseki triplestore SPARQL endpoint. The BookSampo portal has had lately some 1.6 million annual users and has become an important part of the web services provided by the national public library consortium [Kirjastot.fi](#).

BookSampo 2.0: Semantic Search, Browsing, and Data-analyses

In 2022 we started a new project pertaining to BookSampo with two major aims:

1. Develop a prototype user interface for the BookSampo knowledge graph based on the [Sampo Model](#) and [Sampo UI](#) tools that facilitate semantic search and browsing integrated seamlessly with tools for data analysis.
2. Study Finnish fiction literature using Digital Humanities methods based on the BookSampo knowledge graph, LOD services, and the portal.



Kiitos yhteistyöstä!

[Tutkimuksen kotisivu:](#)

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

Latest News

[FIN-CLARIAH Research Infrastructure](#)

A new national research infrastructure initiative FIN-CLARIAH for...
8.12.2021 8:12 by eahyvone

[WarMemoirSampo published on December 3, 2021](#)

A new "Sampo" application, "WarMemoirSampo"...
8.12.2021 8:04 by eahyvone

[Five new SeCo papers accepted for the ISWC 2021](#)

The 20th International Semantic Web Conference (ISWC 2021), the...
2.8.2021 6:53 by eahyvone

Latest Publications

Frida Ehrnsten, Eljas Oksanen, Heikki Rantala and Eero Hyvönen: [DigiNUMA ja Rahasampo – uusi digitaalinen palvelu rahallöydöistä kiinnostuneille](#)

Mehwish Alam, Victor de Boer, Enrico Daga, Marieke van Erp, Eero Hyvönen and Albert Meriño-Peñuela: [Editorial of Special Issue on Cultural Heritage and Semantic Web](#)

Eljas Oksanen and Michael Lewis: [Evaluating Transformations in Small Metal Finds Following the Black Death](#)

Heikki Rantala, Eero Hyvönen and Petri Leskinen: [Finding and explaining relations in a biographical knowledge graph based on life events: Case BiographySampo](#)

SeCo on Twitter

Tweets from @secoresearch

Follow on Twitter

Semantic Computing Research ... @secorese... · Sep 24

AVOIN KUTSU KIRJASAMPO 2.0 -JÄRJESTELMÄN JULKISTUSTILAISUUTEEN

10.10.2023, klo 13:00–15:00
Aalto-yliopisto ja Zoom
seco.cs.aalto.fi/events/2023/20...

@HelsinkiDH @AaltoUniversity @CSAalto @helsinkiuni @SocSciHelsinki

2

Semantic Computing Research ... @secorese... · Aug 8

Parlamenttisampo.fi tarjoaa työkalun kansanedustajien analysoimiseen. Alla Helsingin Sanomien artikkelit mm. ministerien ja kansanedustajien rasistisista puheista. facebook.com/groups/1697627...

@HelsinkiDH @AaltoUniversity @CSAalto @helsinkiuni @SocSciHelsinki

Päivän ohjelma



- 13:00–13:15 **Miksi Kirjasampo 2.0 taottiin: Sampo-mallin esittely**
Eero Hyvönen, professori, Aalto-yliopisto, johtaja, Helsingin yliopiston HELDIG-keskus
- 13:15–13:45 **Kirjasampo.fi: tausta, tavoitteet ja esittely**
Matti Sarmela, kehittäispäällikkö, Valtakunnalliset kehittämispalvelut, Helsingin kaupunginkirjasto,
Kaisa Hypén, päätoimittaja, Kirjasampo.fi, Turun kaupunginkirjasto ja
Tuomas Aitonurmi, Kirjasammon toimittaja, Valtakunnalliset kehittämispalvelut, Helsingin kaupunginkirjasto
- 13:45-14:00 **Sampo-UI: malli ja työkalu sampojen käyttöliittymiä varten**
Heikki Rantala, väitöskirjatutkija, Aalto-yliopisto
- 14:00-14:15 **Kirjasampo 2.0:n esittely**
Annastiina Ahola, DI, Semanttisen laskennan tutkimusryhmä (SeCo), Aalto-yliopisto
- 14:15-14:30 **Kirjasampo-datan käyttö kirjallisuuden tutkimuksessa**
Telma Peura, FM, Semanttisen laskennan tutkimusryhmä (SeCo), Helsingin yliopisto (HELDIG-keskus)
- 14:30-15:00 **Malja suomalaiselle kirjallisuudelle: verkostoitumistilaisuus aulassa**