

Sotahistoriallisten aineistojen julkaiseminen semantisessa webissä

Prof. Eero Hyvönen

Aalto University and University of Helsinki
HELDIG – Helsinki Centre for Digital Humanities
Semantic Computing Research Group (SeCo)

*Sotasurmasammon 1914-1922 julkistustilaisuus
20.11.2019 Kansallisarkisto*

Sotasurmasammon 1914-1922 takojat

Aalto-yliopisto ja HELDIG-keskus

- Lia Gasbarra (ammattiontologia)
- Eero Hyvönen (hankkeen vetäjä)
- Esko Ikkala (arkkitehtuuri)
- Mikko Koho (ammattiontologia)
- Heikki Rantala (pääarkkitehti)
- Jouni Tuominen (infrastruktuuri)

Kansallisarkisto

- Päivi Happonen (ohj. ryhmä)
- Ilkka Jokipii (päättutkija)
- Vili Haukkovaara (ohj. ryhmä)
- Jarmo Nieminen (ohj. ryhmä)
- Markku Mäenpää (projektipäällikkö)

Sisältö: Sotasurmasammon uutuusarvo ja idea

- Visio 1: Historiallisten aineistojen tietoinfrastruktuuri ja "Sampo"-portaalit
- Visio 2: Tietoportaalien 4 sukupolvea ja Sotasurmasampo 1914-1922



Visio 1:

Historiallisten aineistojen tietoinfrastrukturi ja "Sampo"-portaalit

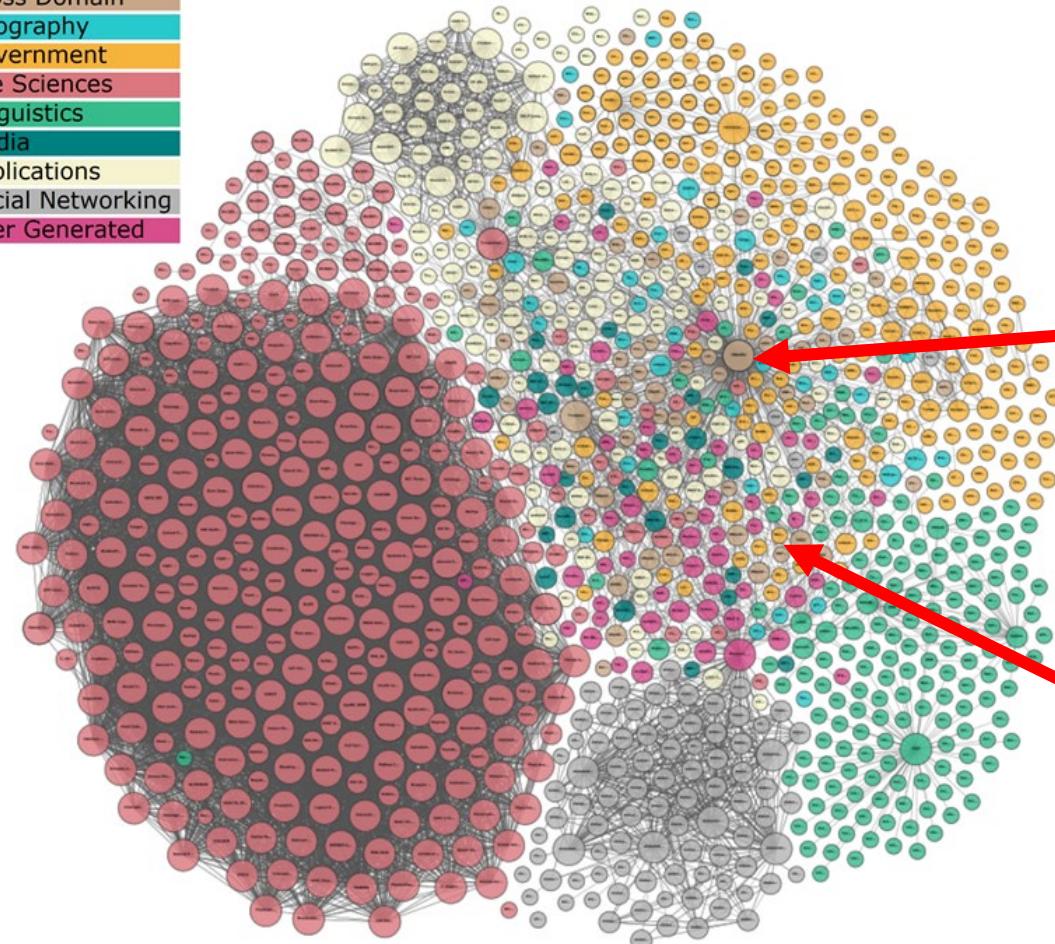


Web of Data: Linked Open Data Cloud

Human Knowledge on the Semantic Web

Legend

Cross Domain
Geography
Government
Life Sciences
Linguistics
Media
Publications
Social Networking
User Generated



Wikipedia

Your data

LODStats 2018 (<http://stats.lod2.eu/>):
10 000 datasets, 150 billion triples

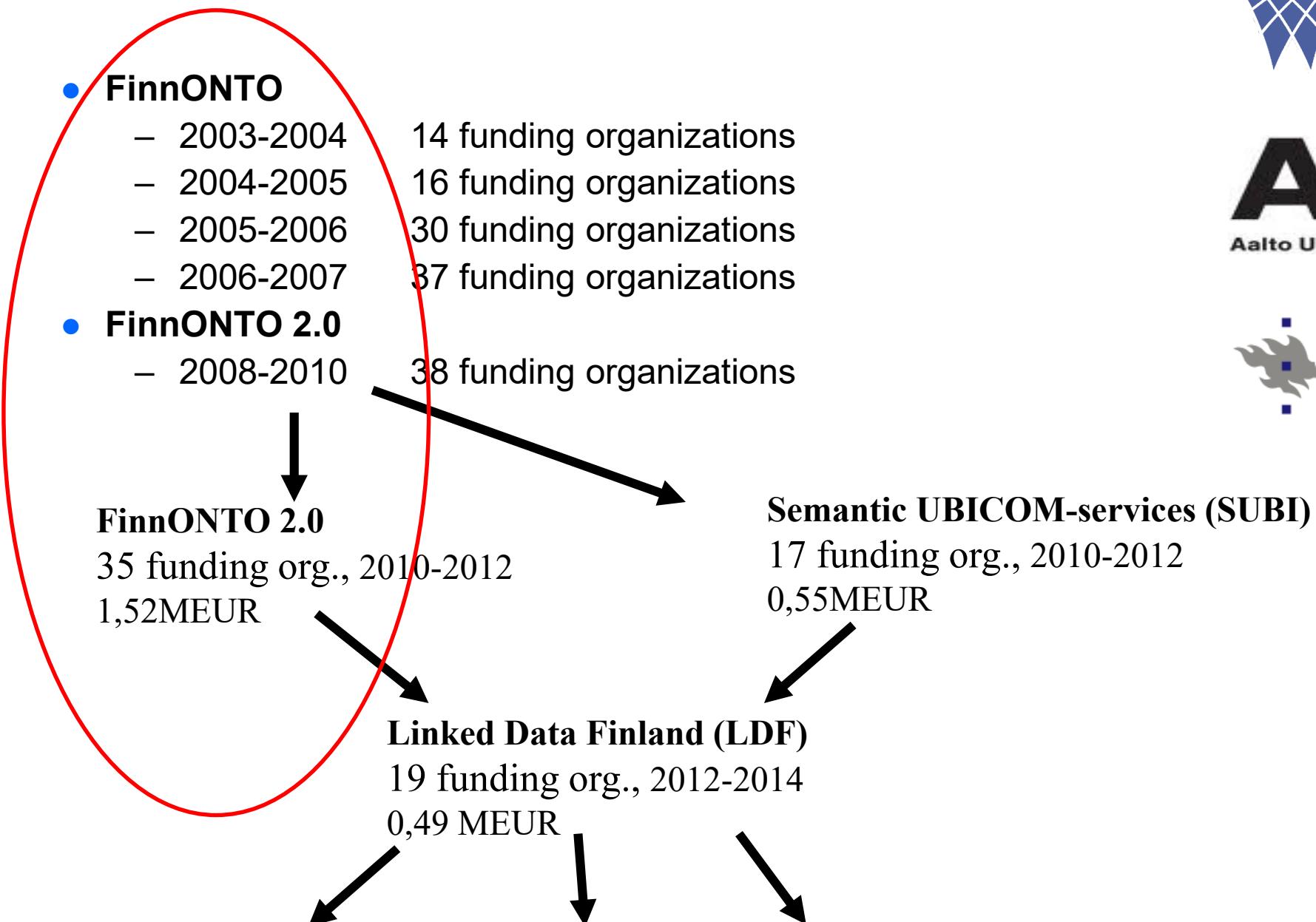
Big Boys Have Entered the Game: **Knowledge Graphs**

<http://schema.org>

- Google Knowledge Graph
- Microsoft Satori
- IBM Watson
- eBay Products
- Facebook Graph
- ...



National Linked Data Projects in Finland since 2003



Cultural Heritage "sampos" on the Semantic Web

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

- **CultureSampo** – Finnish Culture on the Semantic Web (2008)
- **TravelSampo** - Mobile Contextualized Services of Cultural Tourism (2011)
- **BookSampo** – Finnish Fiction Literature on the Semantic Web (2011) [2 million users in 2018]
- **WarSampo** – Finnish WW2 on the Semantic Web (2015) [230 000 users in 2018]
 - » + War Cemeteries application [2017]
 - » + War Prisoners application [to be published 29.11.2019]
- **BiographySampo** - Finnish Biographies on the Semantic Web (2018) [15 000 users]
- **NameSampo** – Linked Data Workbench for Toponomastic Research (2019) [30 000 users]
- **WarVictimSampo 1914-1922** – National War History, with National Archives of Finland

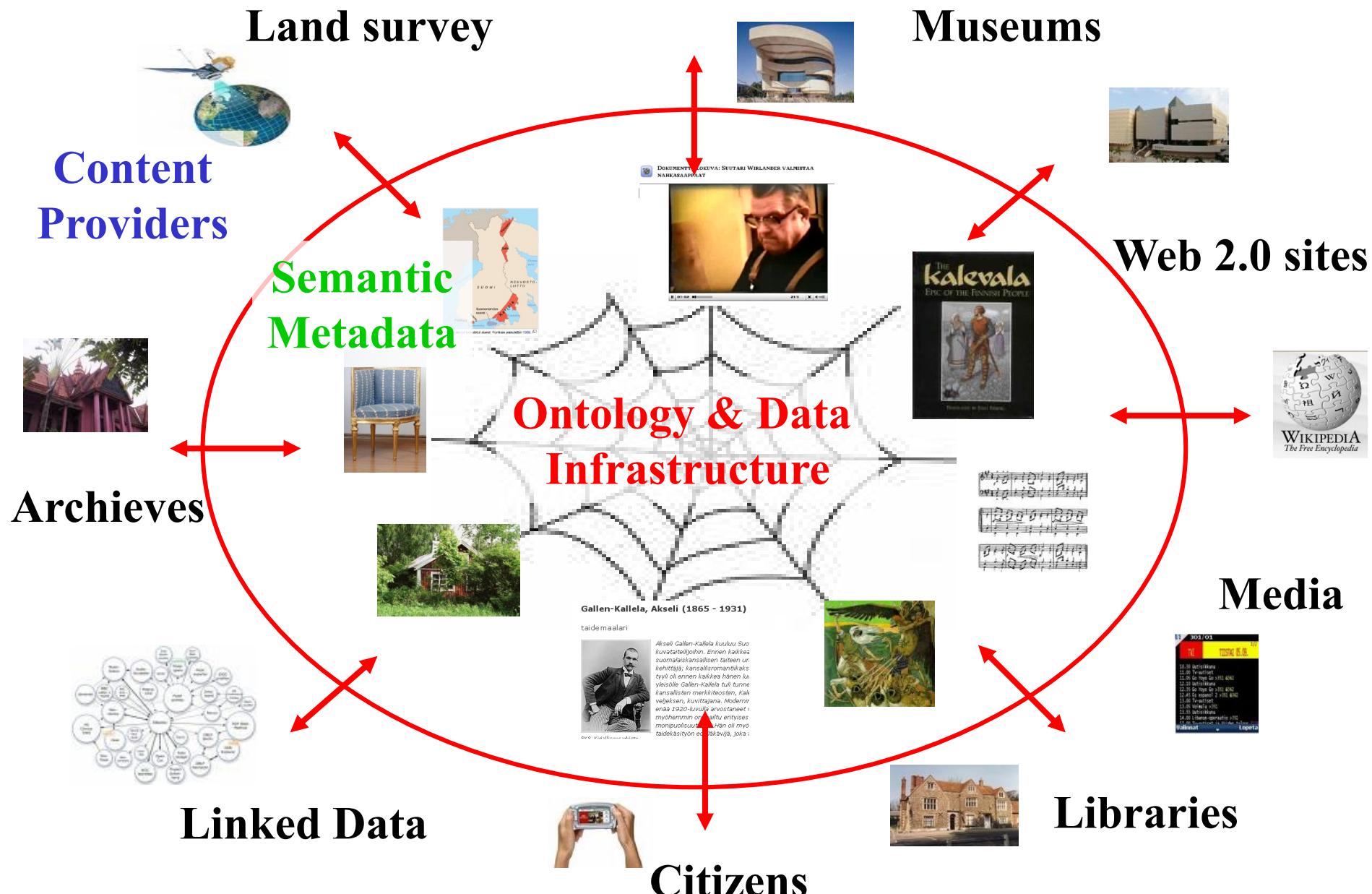
-
- **MMM – Manuscript Migration Mapper**, with Oxford, Pennsylvania, IRHT (Paris), 23.11.2019
 - **FindSampo** – Archaeology & Citizen Science of metal detectorists, with Finnish Heritage Agency (-2021)
 - **LawSampo** – Finnish Legislation and Case Law, with Ministry of Justice of Finland (2020)
 - **AcademySampo** – Historical Finnish Academic People in 1640–1899 (2020)
 - **ParliamentSampo** – Parliament of Finland data, new project, Academy of Finland (2020-2022)

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



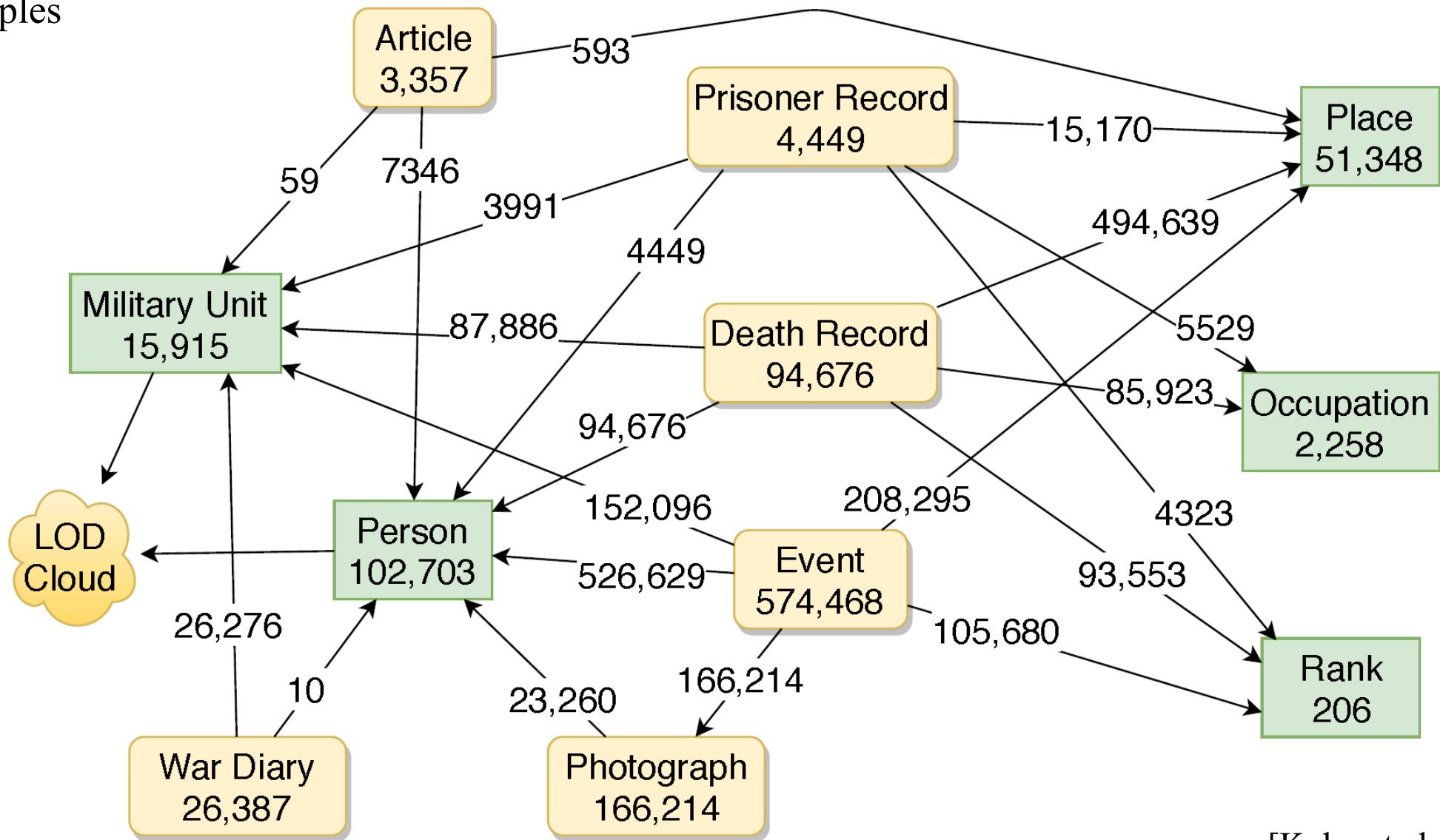
Defense of Sampo
A. Gallen-Kallela, 1896

"Sampo" Model for Publishing Cultural Heritage Contents: Everybody Wins!



WarSampo Infrastructure: Linked Open Data Cloud

14 million triples

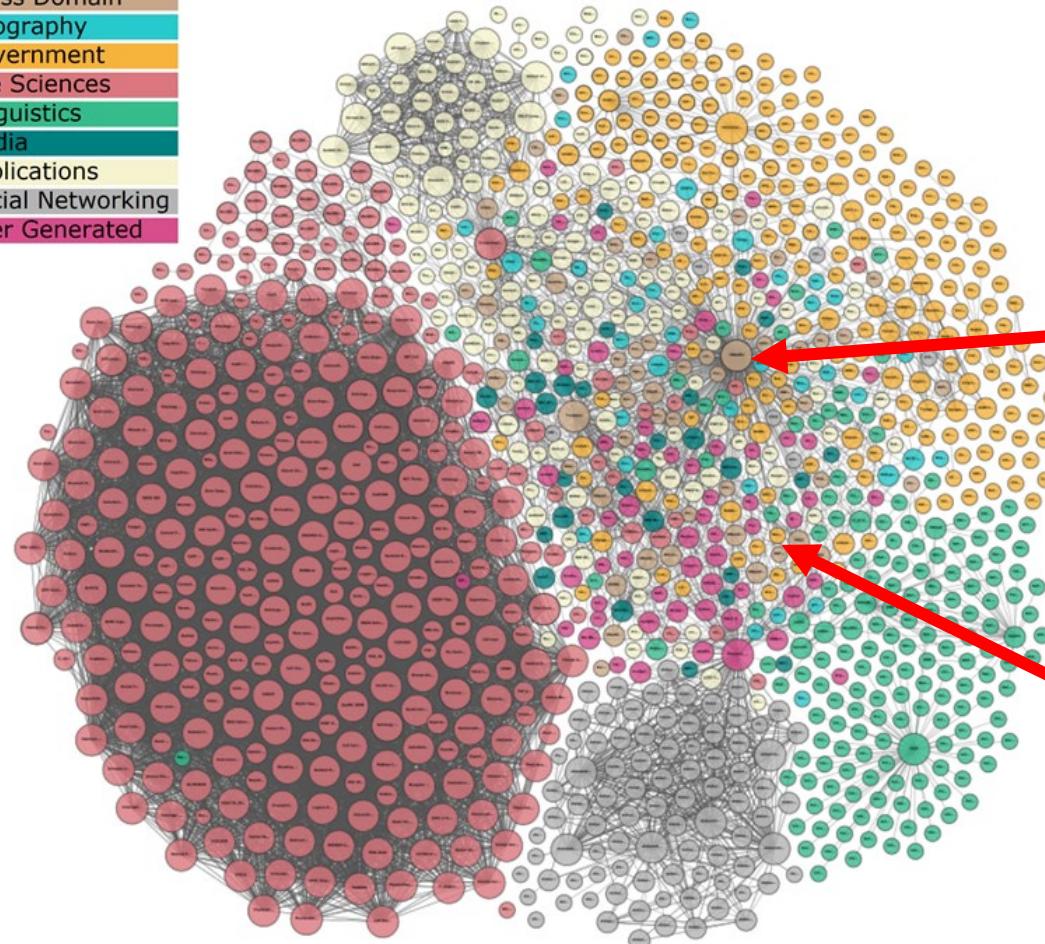


Web of Data: Linked Open Data Cloud

Human Knowledge on the Semantic Web

Legend

Cross Domain
Geography
Government
Life Sciences
Linguistics
Media
Publications
Social Networking
User Generated

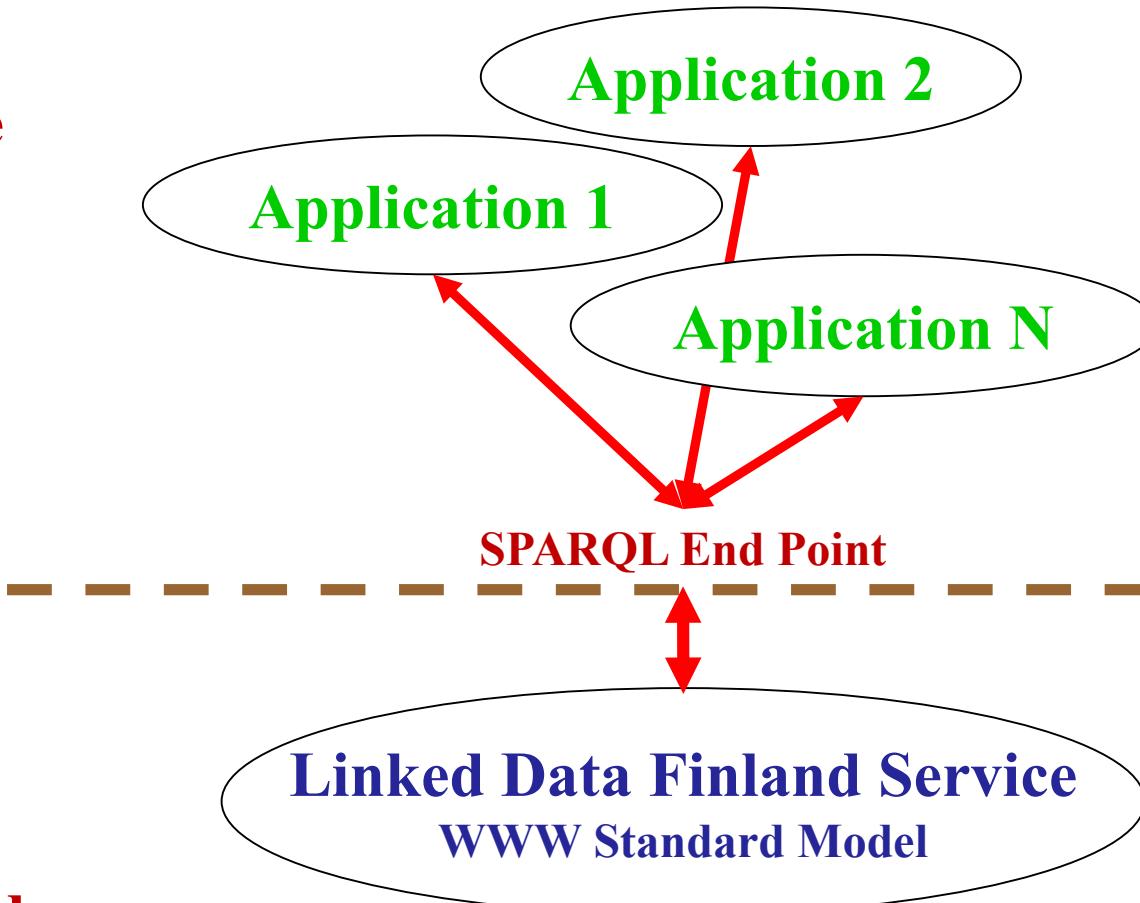


Wikipedia

WarSampo

Machines & Developers View: SPARQL Useful for Digital Humanities Researchers, too!

Client Side
(Browser)



Server Side
<http://ldf.fi>

The right side of the slide displays two screenshots of the Linked Data Finland Service interface. The top screenshot shows a map of Finland with various locations marked by colored dots, likely representing war cemeteries. The bottom screenshot shows a detailed view of a specific location, "Akaa, Kylmäkosken sankarihautausmaa", with a description, four small images of the cemetery, and a graph showing data over time.

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:



[HOMEPAGE](#)

[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](#)

KUTSU JULKISTUSTILAIKUUTEEN

Kansallisarkisto, Sotavainajien muiston vaalimisyhdistys ry sekä Aalto-yliopiston ja Helsingin yliopiston digitaalisten ihmistieteiden keskuksen (HELDIG) Semanttisen laskennan tutkimusryhmä (SeCo) kutsuvat Teidät avoimeen julkistustilaisuuteen:



Suomalaiset sotavangit 1939-1945 Neuvostoliitossa

Perjantaina 29.11.2019, klo 14:00–15:30

Kansallisarkisto, Vanha tutkijasali, Rauhankatu 17, Helsinki

Tilaisuudessa avataan ja esitellään talvisodan 80 vuotismuiston kunniaksi [Sotasammon](#) uusi, linkitettyyn avoimeen dataan perustuva sovellus "Suomalaiset sotavangit Neuvostoliitossa 1939-1945". Sotasampo on suomalaisia toisen maailman sodan aineistoja yhdistävä ja julkaiseva, semanttiseen webiin perustuva data- ja verkkopalvelu. Sotasampo julkaistiin syksyllä 2015 ja se sai kesällä 2017 kansainvälisen LODLAM Challenge Open Data Prize -palkinnon Venetsiassa. Palvelulla oli v. 2018 230 000 käyttäjää verkossa.

Pilottisovellus verkossa

Sotasammon uusi Sotavangit-sovellus on tarkoitettu sekä suuren yleisön että tutkijoiden käytettäväksi. Jokaiselle sotavangille on tehty automaattisesti "kotisivu", johon on linkitetty kaikki häneen liittyvät tiedonmuruset. Aineistot tarjoillaan käyttäjälle uutena, yhdeksäntenä sovellusnäkymänä Sotasammossa. Nämä tarjoaa älykkään hakuliittymän sotavankiaineistoon ja mahdollisuuden selata, analysoida ja visualisoida aineistoja digitaalisten ihmistieteiden keinoin. Sovellus avataan osoitteessa:

<https://sotasampo.fi/>

Tutkimushanke taustalla

Uusi sovellus on ideoitu, suunniteltu ja toteutettu Aalto-yliopiston ja Helsingin yliopiston HELDIG-keskuksen Semanttisen laskennan tutkimusryhmässä (SeCo) yhteistyössä Kansalliariston ja Sotavainajien muiston vaalimisyhdistyksen asiantuntijoiden kanssa Kaatuneiden muistosäätiön rahoituksella. Järjestelmä on uusi jäsen kansainvälisestä ja kotimaassa palkittujen semanttisen webin "[Sampo](#)-portaalien sarjassa" ja hyödyntää kansallista semanttisen webin infrastruktuuria [LODI4DH](#).

Visio 2:

Tietoportaalien neljä kehitysastetta ja Sotasurmasampo 1914-1922



Using the Semantic Web in Digital Humanities: Shift from Data Publishing to Data-analysis and Serendipitous Knowledge Discovery

Eero Hyvönen

*University of Helsinki, Helsinki Centre for Digital Humanities (HELDIG), Finland &
Aalto University, Department of Computer Science, Finland
Semantic Computing Research Group (SeCo) (<http://seco.cs.aalto.fi>)
E-mail: eero.lavonen@aalto.fi*

Abstract. This paper discusses a shift of focus in research on Cultural Heritage semantic portals, based on Linked Data, and envisions and proposes new directions of research. Three generations of portals are identified: Ten years ago the research focus in semantic portal development was on data harmonization, aggregation, search, and browsing ("first generation systems"). At the moment, the rise of Digital Humanities research has started to shift the focus to providing the user with integrated tools for solving research problems in interactive ways ("second generation systems"). This paper envisions and argues that the next step ahead to "third generation systems" is based on Artificial Intelligence: future portals not only provide tools for the human to solve problems but are used for finding research problems in the first place, for addressing them, and even for solving them automatically under the constraints set by the human researcher. Such systems should preferably be able to explain their reasoning, which is an important aspect in the source critical humanities research tradition. The second and third generation systems set new challenges for both computer scientists and humanities researchers.

Keywords: Digital Humanities, Linked Data, Semantic portals, Data analysis, Knowledge discovery

1. Introduction

Cultural Heritage (CH) has become a most active

and forms a substantial part of DBpedia³ and Wikidata⁴.

The availability of Big Data has boosted the rapid emerging new research area of Digital Humanit

1. GENERATION SYSTEMS

PRINTED TEXTS



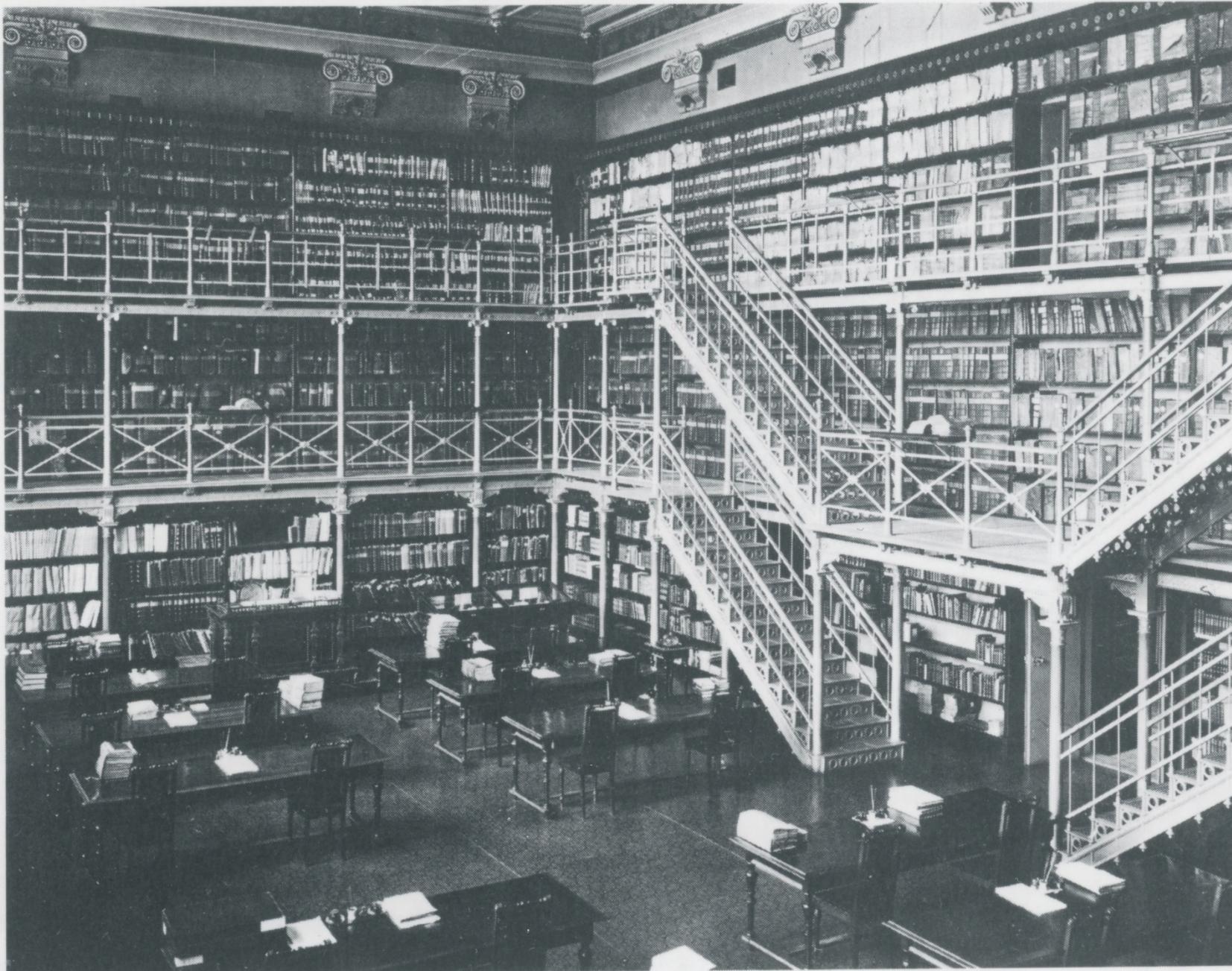
m
l
k
i
h

m
l
k
i
h

n
m
I
k
i
h

o
n
m
I
k
i
h
g

Tutkijasali vuosisadan alussa. Korokkeella päivystäjän pöytä. Pöydät ilman valaisimia.
Hyllyillä vanha kameraalinen aineisto ennen vuotta 1809.



2. GENERATION SYSTEMS

TEXTS ONLINE

Suomen sotasurmat 1914-1922 verkkopalvelu

<http://vesta.narc.fi/cgi-bin/db2www/sotasurmahaku/input>

<= Suomen sotasurmat 1914-1922 Suomeksi På Svenska In English

Haku

Sukunimi <input type="text"/>	Etunimi <input type="text"/>	Ammatti * <input type="text"/>			
Kirjoillaolokunta * <input type="text"/>	Kirjoillaolomaa * <input type="text"/>	Kuolintapa * <input type="text"/>			
Asuinkunta * <input type="text"/>	Kuolinkunta * <input type="text"/>				
Syntymäaika välillä Kuolinaika välillä		alkupäivämäärä <input type="text"/> <input type="text"/>	ja <input type="text"/> ja <input type="text"/>	loppupäivämäärä <input type="text"/> <input type="text"/>	(eg. 1890-12-31) (eg. 1918-12-31)
näytä enintään <input type="text"/> 20					
Etsi					

Haku

Sukunimi <input type="text"/>	Etunimi <input type="text"/>	Ammatti <input type="text"/>
Kirjoillaolokunta <input type="text"/>	Kirjoillaolomaa <input type="text"/>	
Asuinkunta <input type="text"/>	Kuolinkunta <input type="text"/>	Kuolintapa <input type="text"/>
Syntymäaika välillä <input type="text"/> alkupäivämäärä ja <input type="text"/> loppupäivämäärä (eg. 1890-12-31)		
Kuolinaika välillä <input type="text"/> ja <input type="text"/> (eg. 1918-12-31)		
näytä enintään <input type="text"/> 20		
Etsi		

1 - 20

Sukunimi	Etunimi	Ammatti	Syntymäaika	Kuolinaika	Kirjoillaolokunta	Asuinkunta	Kuolinkunta	Kuolintapa
Kuula	Toivo Timoteus	Säveltäjä	7.7.1883	18.5.1918	Helsinki e	Helsingin pitäjä	Viipuri	Murhattu
Kuula	Karl Vilhelm	Sekatyömiest	17.1.1890	12.7.1918	Lohja	Lohja	Hämeenlinna	Kuollut vankileirillä
Kuula	Vihti	Renki	22.3.1882	1918	Nastola			Kadonnut
Kuula	Einari Julius		17.12.1890	2.7.1918	Peräseinäjoki		Tampere	Vahingonlaukaus
Kuula	Väinö Valdemar Johan Gustafsson	Kondiittori	30.11.1896	3.4.1918	Siipyy		Tampere	Kaatunut

Hakuehdolla löytyi: (5)

Tiedot henkilöstä [Sulje ikkuna]

Sukunimi	Kuula (HKA) (PT) (RK) (SRK)	Osapuoli, johon luettiin kuollessaan	Valkoinen (PT)
Etunimi	Toivo Timoteus (HKA) (PT) (RK) (SRK)	Sotilasasema	
Sukupuoli	Mies (HKA) (PT) (SRK)	Sotilasarvo	Ei tietoa
Ikä	34	Asema (sotilas vai siviili)	Ei tietoa
Siviilisääty	Naimisissa (HKA) (PT) (SRK)	Aseenkanto	Kyllä (PT)
Lasten lukumäärä	1 (RK)	Asemahuomautukset	
Kansallisuus		Vangitsemispaiikka	
Syntymäpaikka	Vaasa (RK)	Vankileiri	
Syntymämaa		Vangitsemismotiivi	
Syntymävuosi	1883 (HKA) (PT) (SRK)	Vangitsemisaika	
Syntymäpäivä	7.7. (HKA) (PT) (SRK)	Vankilasta vapautumisen ajankohta	
Kirjoillaolunkunta	Helsinki e (HKA) (PT) (SRK)	Kuolinvuosi	1918 (HKA) (PT) (SRK)
Kirjoillaololääni	Uudenmaan lääni (HKA) (PT) (SRK)	Kuolinpäivä	18.5. (HKA) (PT) (SRK)
Kirjoillaolomaa		Kuolinkunta	Viipuri (HKA)
Asuinkunta	Helsingin pitäjä (HKA)	Kuolinlääni	Viipurin lääni (HKA)
Asuinlääni	Uudenmaan lääni (HKA)	Kuolinmaa	
Asuinmaa		Kuolinpaikka	
Perheen toimeentulo	Ei tietoa	Kuolintapa	Murhattu
Ammatti	Säveltäjä (HKA) (PT) (SRK)	Surmamotiivi	
Kieliryhmä	Suomi (RK)	Hautaustapa	Ei Tietoa
Uskontokunta			Merkity "muu tai
Henkilöhuomautukset			Surmahuomautukset tuntematon" ilman lisätietoja - PT,
Sotilasjärjestö			
Rekryointitapa sotilasjärjestöön	Ei Tietoa		Sotasurmatapauksen todennäköisyys (ei merkintää= todennäköinen, e=epävarma, l=todellisuudessa elossa selvinnyt)

Lisätieto

Kuolintapa	Ammuttu	todennäköinen	K
Kuolintapa	Ampumahaava päässä ja tulehdus aivoissa	todennäköinen	E



DAXIA

The image shows a large, ornate library with floor-to-ceiling bookshelves filled with books. A prominent red 'X' is drawn across the center of the image, covering the word 'DAXIA'. The letters are large and white, standing out against the dark background of the library.

3. GENERATION SYSTEMS

TEXTS AS (LINKED) DATA

TOOLS FOR DATA ANALYSIS

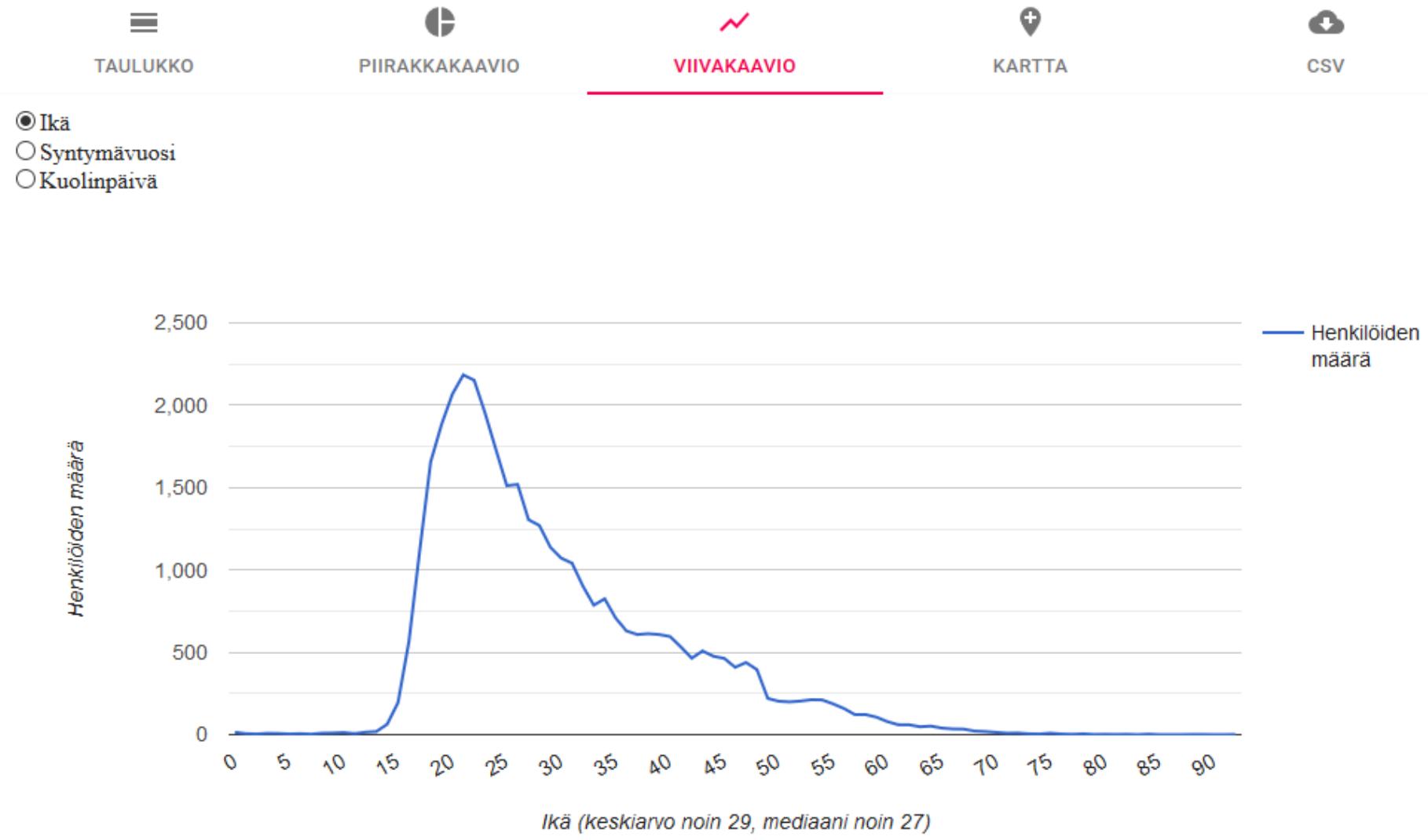
Sotasurmat

Tuloksia: 41521

Rajoita:

Nimi i

Search

Syntymäpäivä iKuolinpäivä iIkä iOsapuoli iAmmatti iKuolintapa iKirjoillaolokunta i

Sotasurmat

Tuloksia: 41521

Rajoita:

Nimi i

Search

Syntymäpäivä iKuolinpäivä iIkä iOsapuoli iAmmatti iKuolintapa iKirjoillaolokunta iKirjoillaololääni iKirjoillaolomaa iKuolinkunta i

TAULUKKO

PIIRAKKAKAAVIO

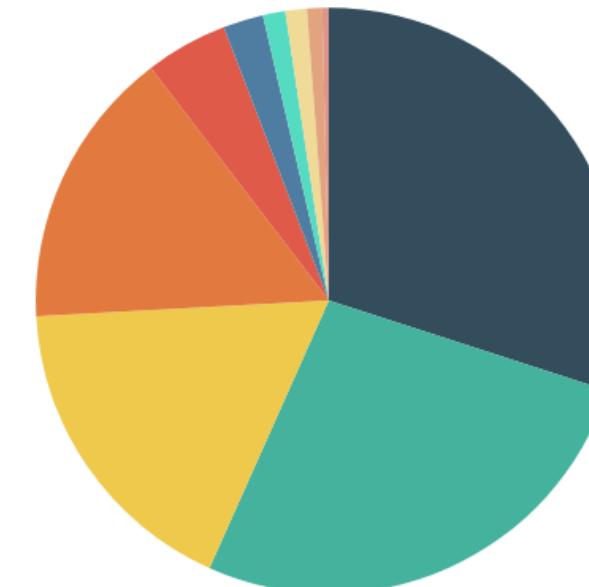
VIIVAKAAVIO

KARTTA



CSV

- Osapuoli
- Ammatti
- Sukupuoli
- Kirjoillaololääni
- Kirjoillaolokunta
- Kuolinlääni
- Kuolinkunta
- Kuolinsyy
- Siviilisäity



Kuolinlääni

- Hämeen lääni (12445)
- Ei merkintää datassa (11088)
- Uudenmaan lääni (7255)
- Viipurin lääni (6422)
- Turun ja Porin lääni (1887)
- Kuopion lääni (923)
- Mikkelin lääni (512)
- Oulun lääni (494)

Sotasurmat

Kuolinpäivä i

01.01.1914 — 31.12.1922

Ikä i

Osapuoli i

Ammatti i

Kuolintapa i

Kirjoillaolokunta i

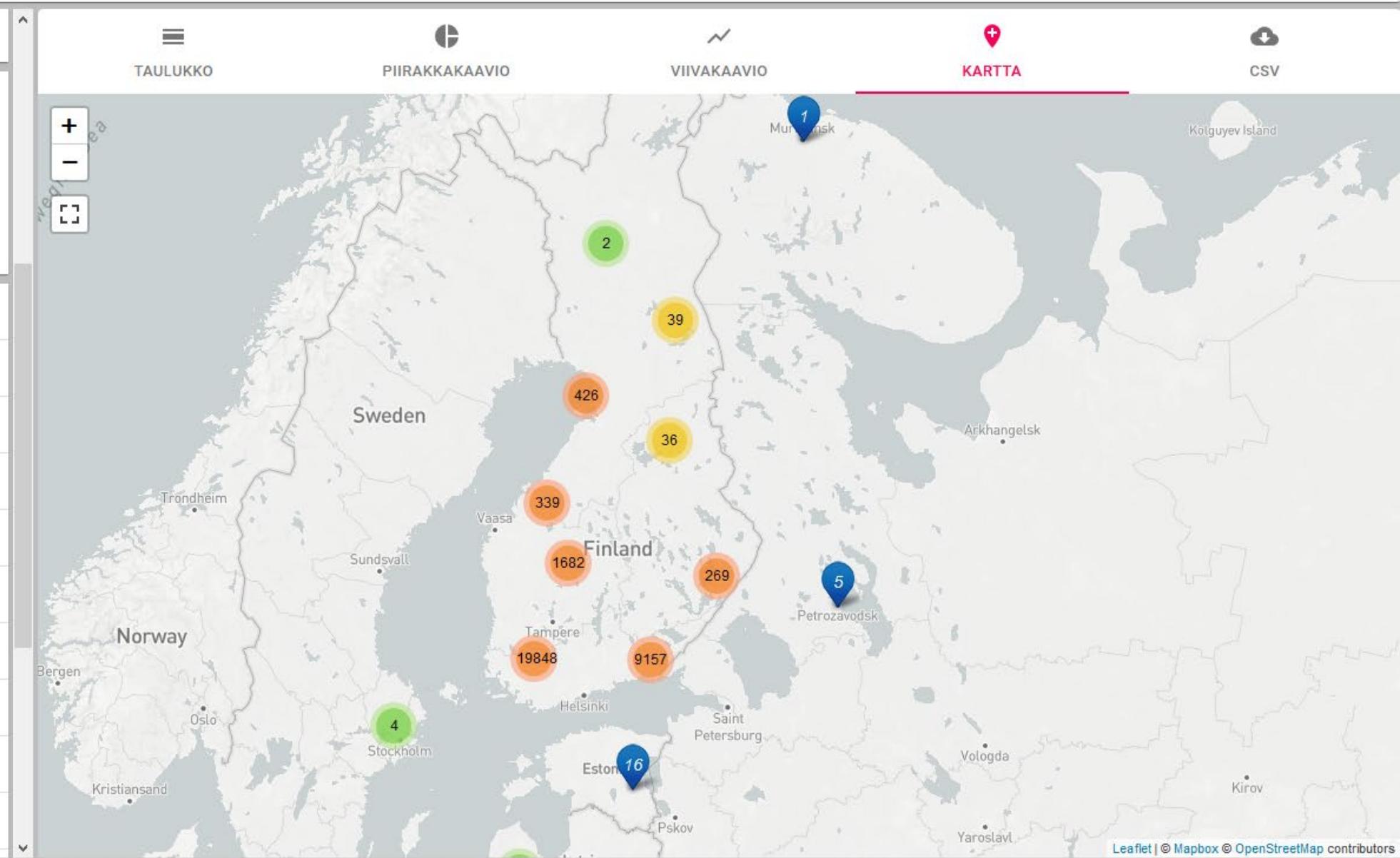
Kirjoillaololääni i

Kirjoillaolomaa i

Kuolinkunta i

Kuolinlänni i

Kuolinmaa i

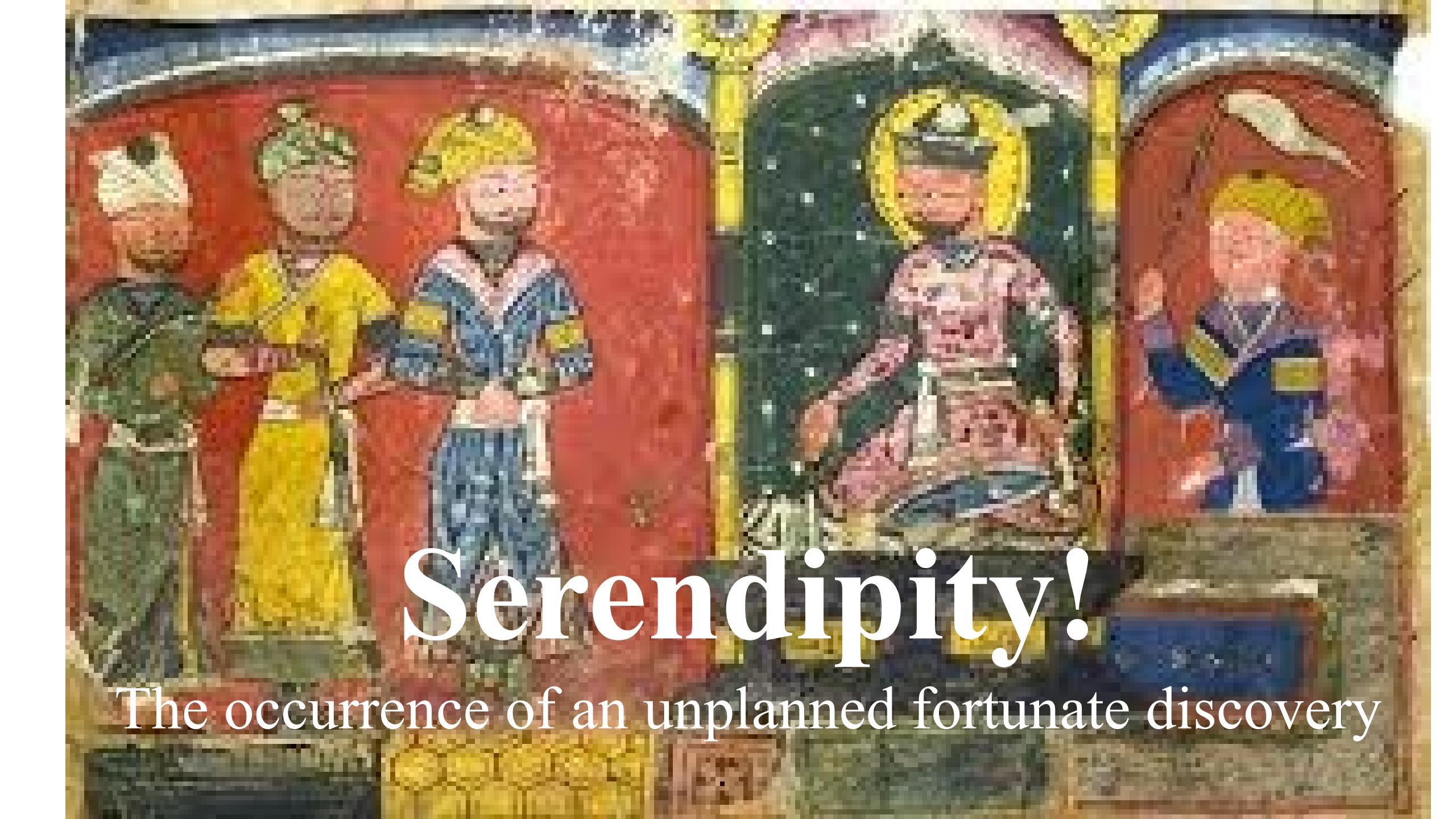


4. GENERATION SYSTEMS

AUTOMATIC KNOWLEDGE DISCOVERY & AI

Publication System as an Intelligent Agent

- Paradigm Shift: Passive Tools -> Active intelligent agents
- Agents **find** "interesting" research questions in the data
- Agents **solve** research questions
- Agents can **explain** their solutions



Serendipity!

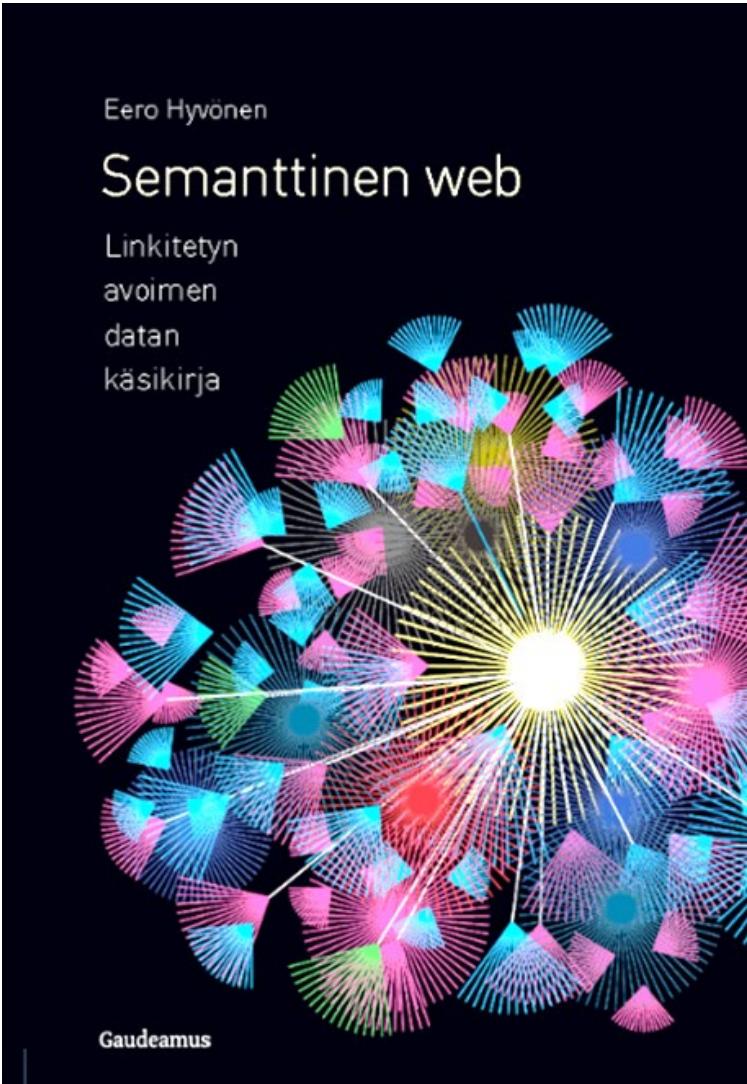
The occurrence of an unplanned fortunate discovery

Esimerkki serendipisyydestä

- **Sotasurmasammon mukaan Viipurin läänin sisällissotaan osallistuneet punaiset kuolivat n. 5 vuotta vanhempina kuin Turun ja Porin läänin punaiset. Miksi?**
- Kone voi etsiä mahdollisesti "kiinnostavia" ilmiöitä ja ehdottaa niitä tutkijalle tarkemmin tutkittavaksi
- Kone voi myös auttaa ongelmien ratkonnassa tuottaa jopa ratkaisuille perusteluja

**More Info
in Finnish:**

<https://seco.cs.aalto.fi/projects/sotasurmat-1914-1922/>



**More Info
in English:**

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

<https://seco.cs.aalto.fi/publications/>



HELDIG
Helsinki Centre for Digital Humanities