



How to Create a National Cross-domain Ontology and Linked Data Infrastructure and Use It on the Semantic Web

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/>



Contents

- 1. Semantic Web**
 - Extending the Layer Cake Model
- 2. How to Build a National LOD Infrastructure?**
 - Lessons learned in Finland
- 3. How to use SW Infra for Applications?**
 - Sampo Model and series of systems
- 4. Paradigm Shifts in Web Publishing**

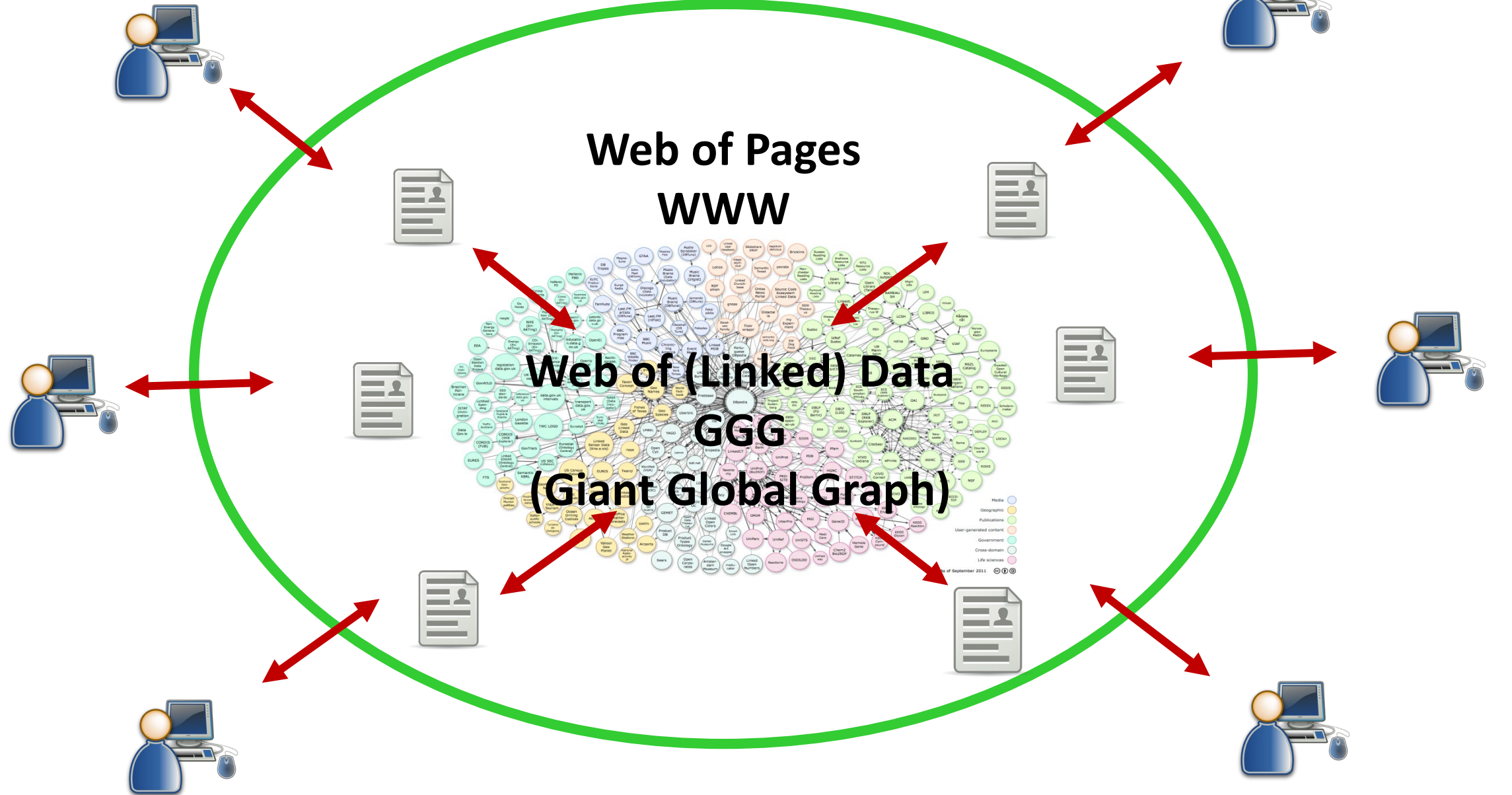


1. SEMANTIC WEB

EXTENDING THE LAYER CAKE MODEL

Web for People

Semantic Web



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/>





Evolving Layer Cake Model of W3C

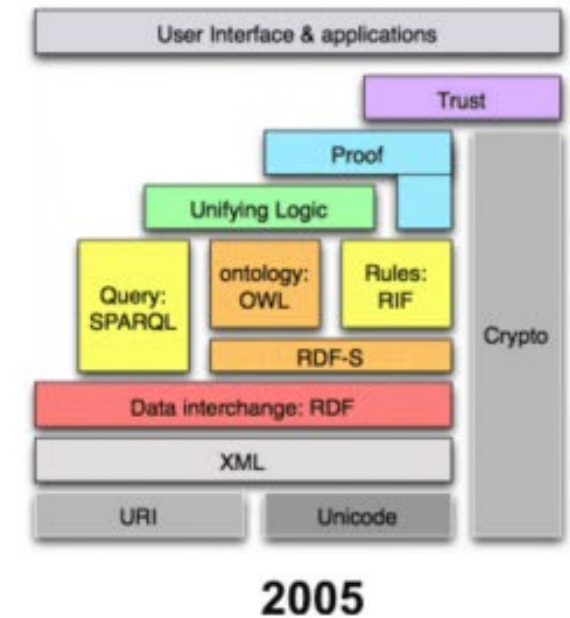
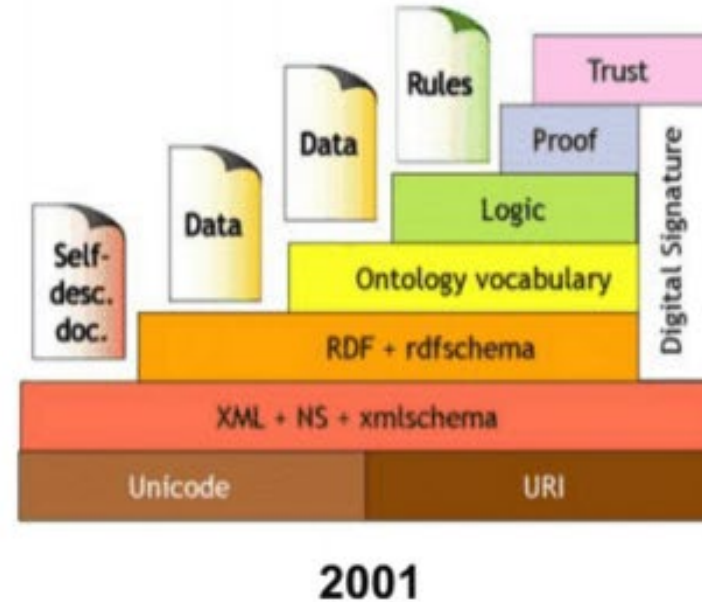
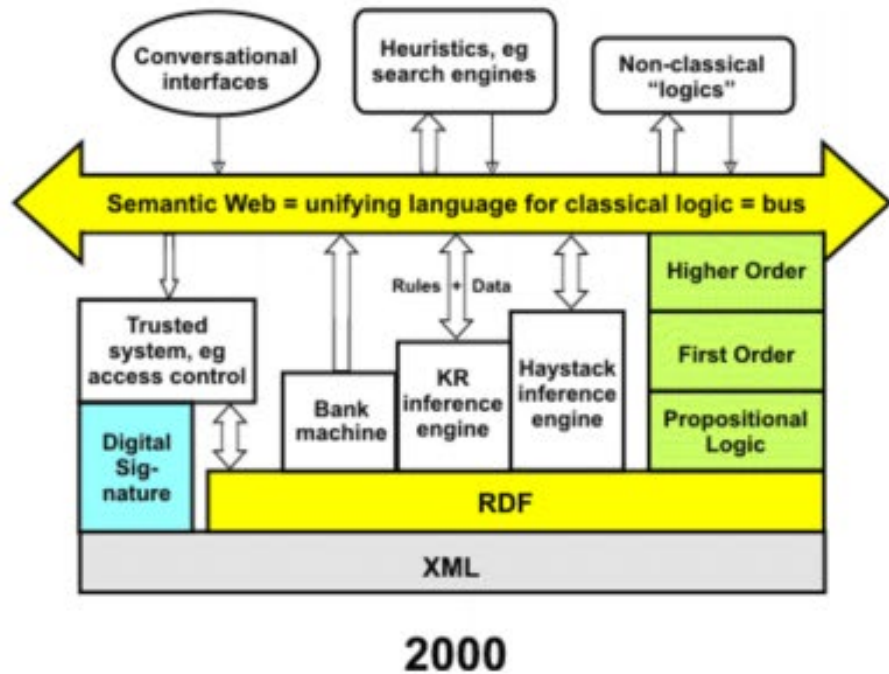


Figure 4. Evolution of the Semantic Web from 2000 to 2005

John Sowa: <http://www.jfsowa.com/ikl/>

Kingsley Uyi Idehen: <https://medium.com/openlink-software-blog/semantic-web-layer-cake-tweak-explained-6ba5c6ac3fab>

Key Challenges Addressed



- W3C SW standards are based on First Order Predicate Logic
- Logic is a nice application domain agnostic model for
 - Knowledge representation and
 - Reasoning

- **Domain-specific models** are needed, too
 - Based on W3C SW standards
- **National level models** are needed using W3C SW standards
 - To support national languages, terminologies, data models, practices, ...

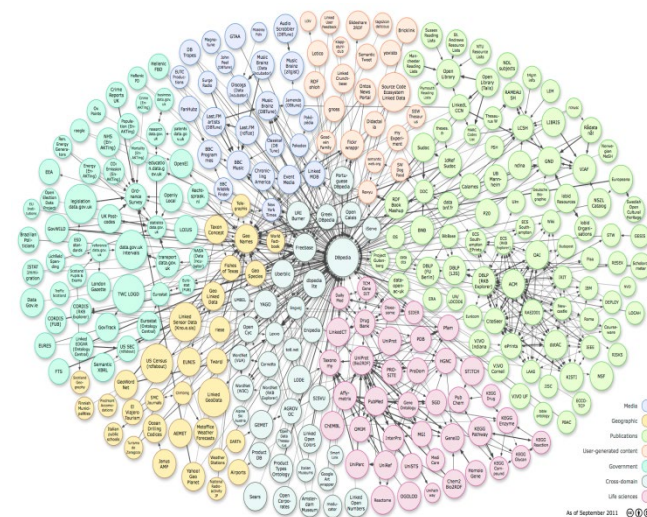
Result: Content Infrastructure



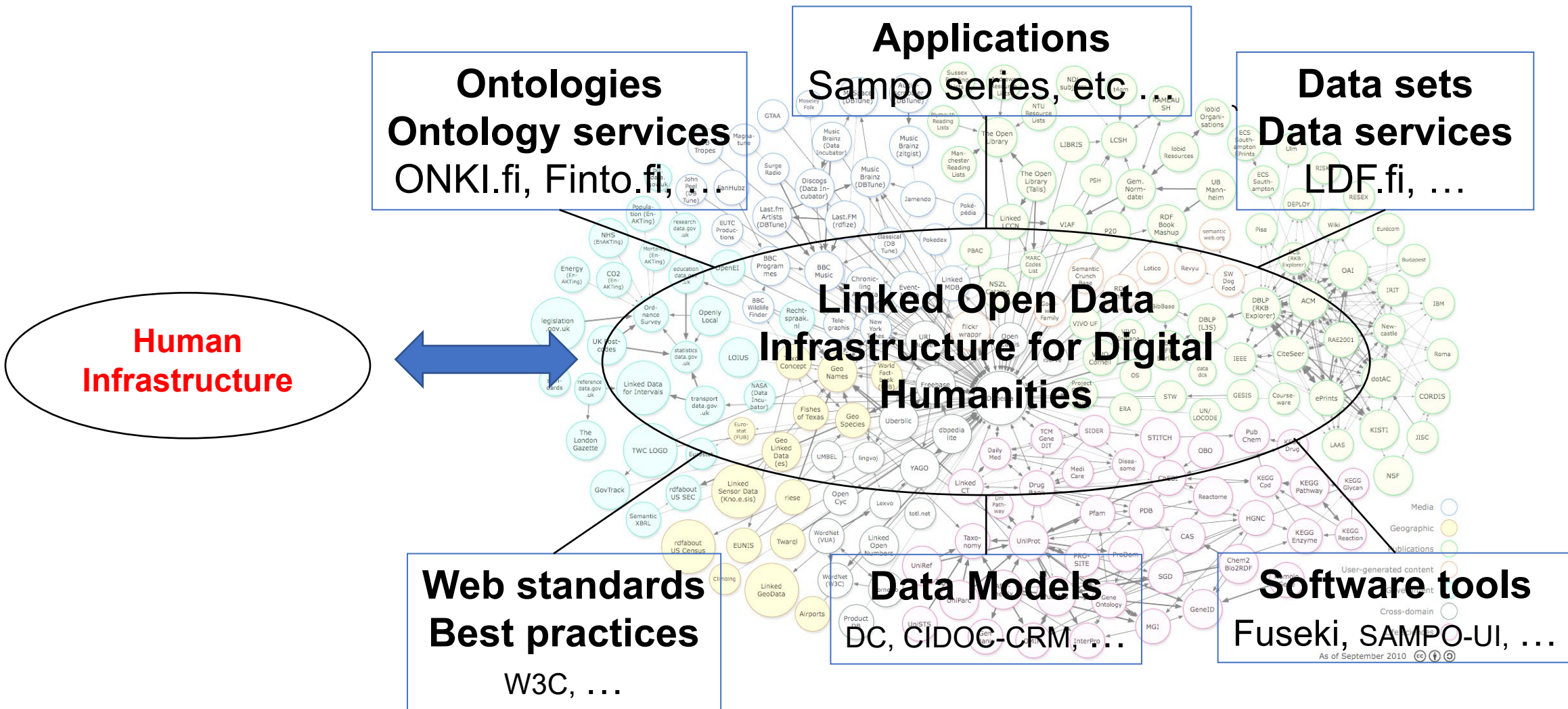
Traditional Infras:
(rail)roads, electricity, ...



Semantic Content Infra:
Ontologies, data, metadata, ...



Linked Open Data Finland: Elements of a National Infrastructure





2. HOW TO BUILD A NATIONAL LOD INFRASTRUCTURE?

Lessons learned in Finland

Starting Point Challenge 2002-2004



- Lessons learned when developing *MuseumFinland – Finnish Museums on the Semantic web*:
 - Semantic Web is good for publishing heterogeneous distributed data
 - Ontologies were not available for SW applications
 - However, there were several thesauri in use
 - Lots of data had been indexed using them
- Developing large cross-domain thesauri is a challenge
 - Domain specific expert groups are needed

=> **National FinnONTO Ontology Initiative in Finland 2003-2012**

FinnONTO Solution Approach 2003-2012



- Paradigm change: **thesauri -> ontologies**
 - Shared ontologies harmonize, interlink, and enrich data automatically
- **Linked ontologies**: Align ontologies for cross-domain applications
- Support **distributed domain-specific ontology development**
 - By different expert groups
- Support re-use by **centralized ontology services**
 - APIs for legacy systems to use
- Create a **sustainable infrastructure** for maintaining ontologies
- Gradually move to **URI-based data indexing on a national level**

Why infrastructure?



**“Intellectuals solve problems
- geniuses prevent them”**

Albert Einstein

Major Domain Ontology Types



- General concept ontologies
- Actor ontologies
- Place ontologies
- Time and period ontologies
- Event ontologies
- Domain nomenclatures and terminologies
 - E.g., medical terms
- Domain "ontology" refers thesaurus or gazetteer like KOSs whose resources are used is element values of metadata descriptions

General Concept Ontologies



Traditional keyword thesauri

- General terms like "wagon", "city", "war", "chair", ...
- Correspond to classes of individuals
- (However, many keyword thesauri contain individuals, too)

Examples

- Art and Architecture Thesaurus (AAT) (culture)
- Library of Congress Subject Headings (LCSH) (library)
- UNSPSC (products and services)
- ...

KOKO: From Thesauri to Ontologies

- Linked Open Ontology Cloud



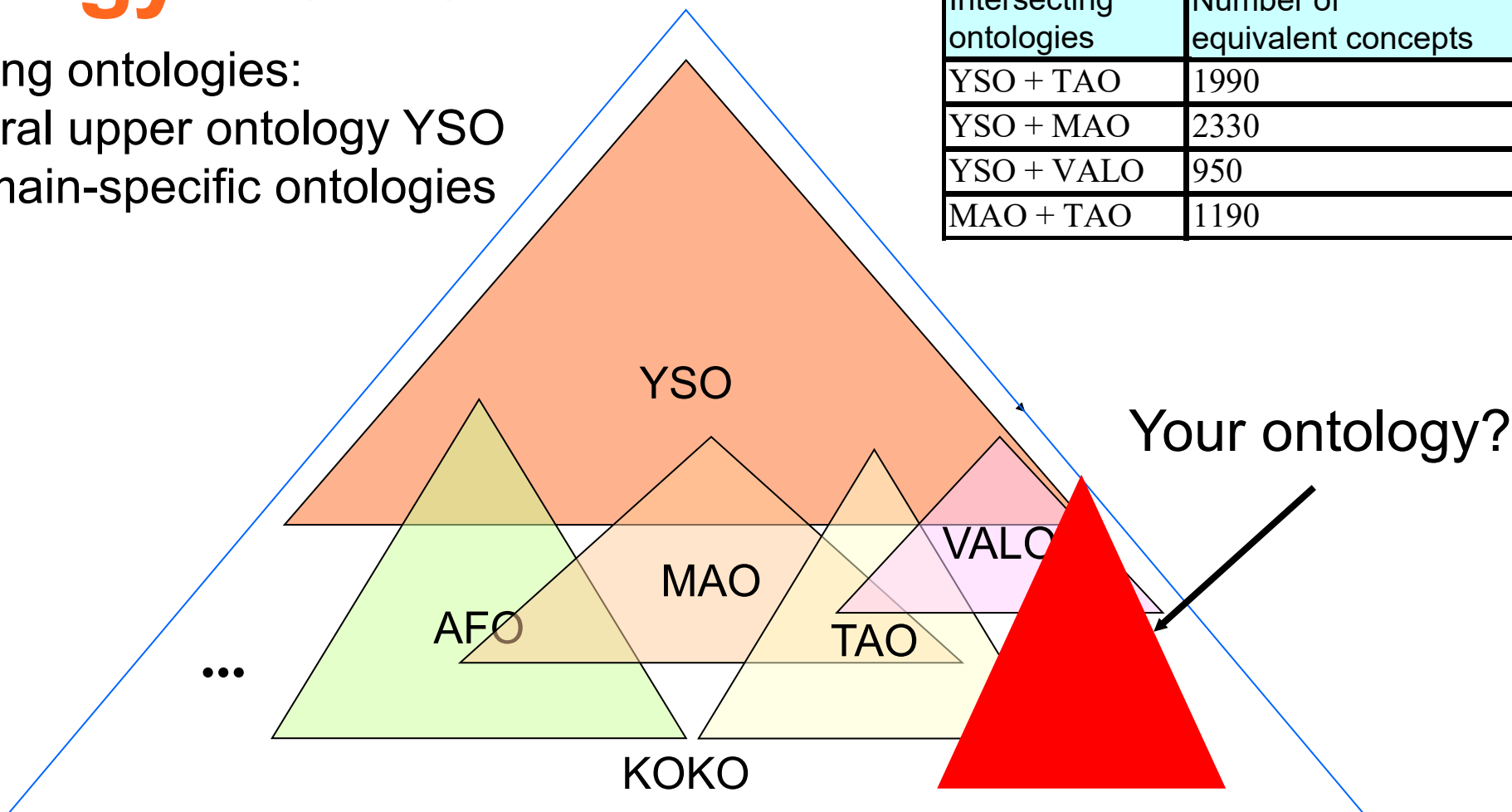
	Name	Ontology domain	Underlying thesaurus	Size	Maintaining Organization
1	YSO	General domain	General Finnish Thesaurus, YSA, Allärs	23700	National Library, Åbo Academy
2	MUSO	Music	Thesaurus of Music, MUSA/CILLA	1000	National Library
3	MAO	Museum domain	Thesaurus of Museum Domain, MASA	6800	National 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, MeSH, 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

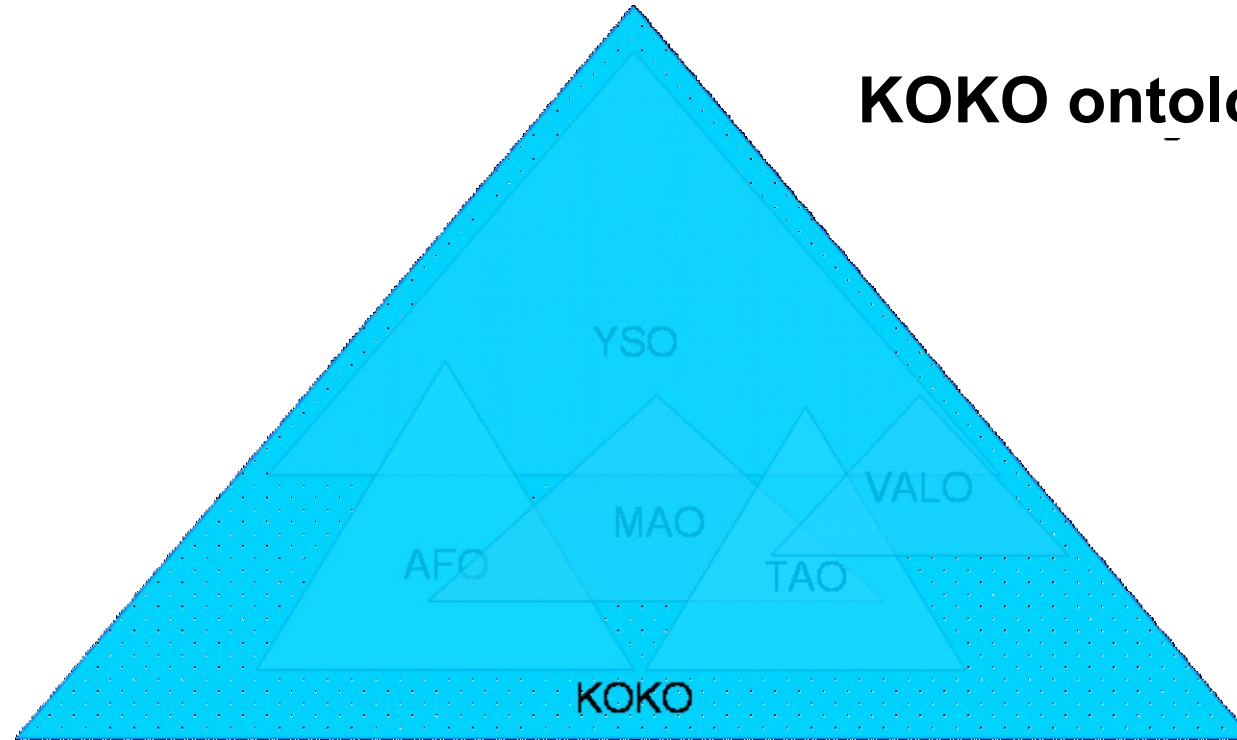


[Hyvönen et al., ESWC 2009]

KOKO from the “end-user” viewpoint



KOKO ontology



Thesaurus -> Ontology Transformation Method



- **Transformation into light-weight ontologies based on RDF Schema**
 - SKOS was developed only later on
- **First transform thesaurus into a subClassOf ontology and then edit it manually using Protégé :**
 1. **Disambiguate vague terms and re-position new split concepts**
 - E.g., “child” as age group \neq “child” as a family relation
 2. **Disambiguate BT relations into subClassOf and partOf**
 3. **Complement fragmentary BT clusters into full subClassOf hierarchies**

Example: University Terminology

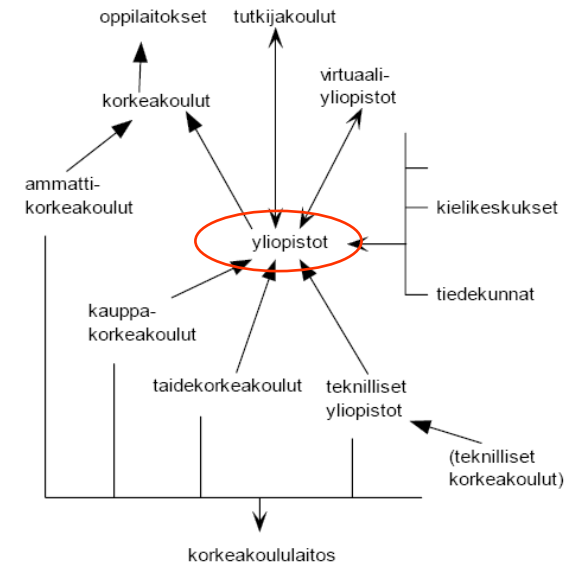
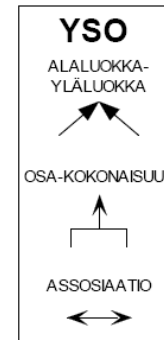
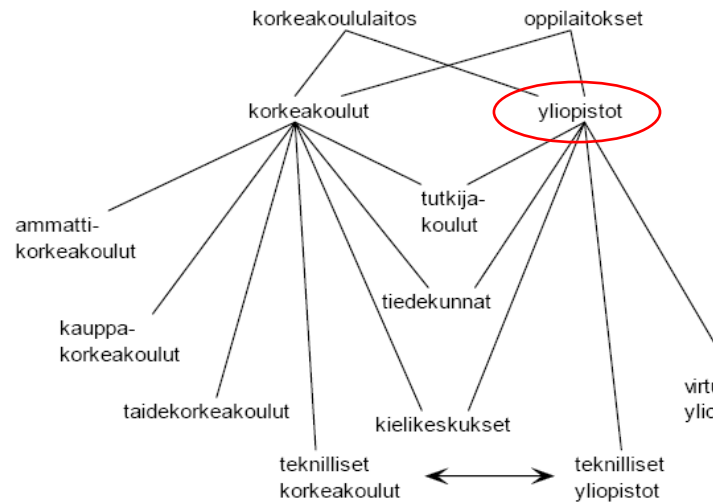
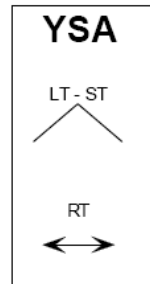
YSA Thesaurus -> YSO Ontology



YSA Thesaurus



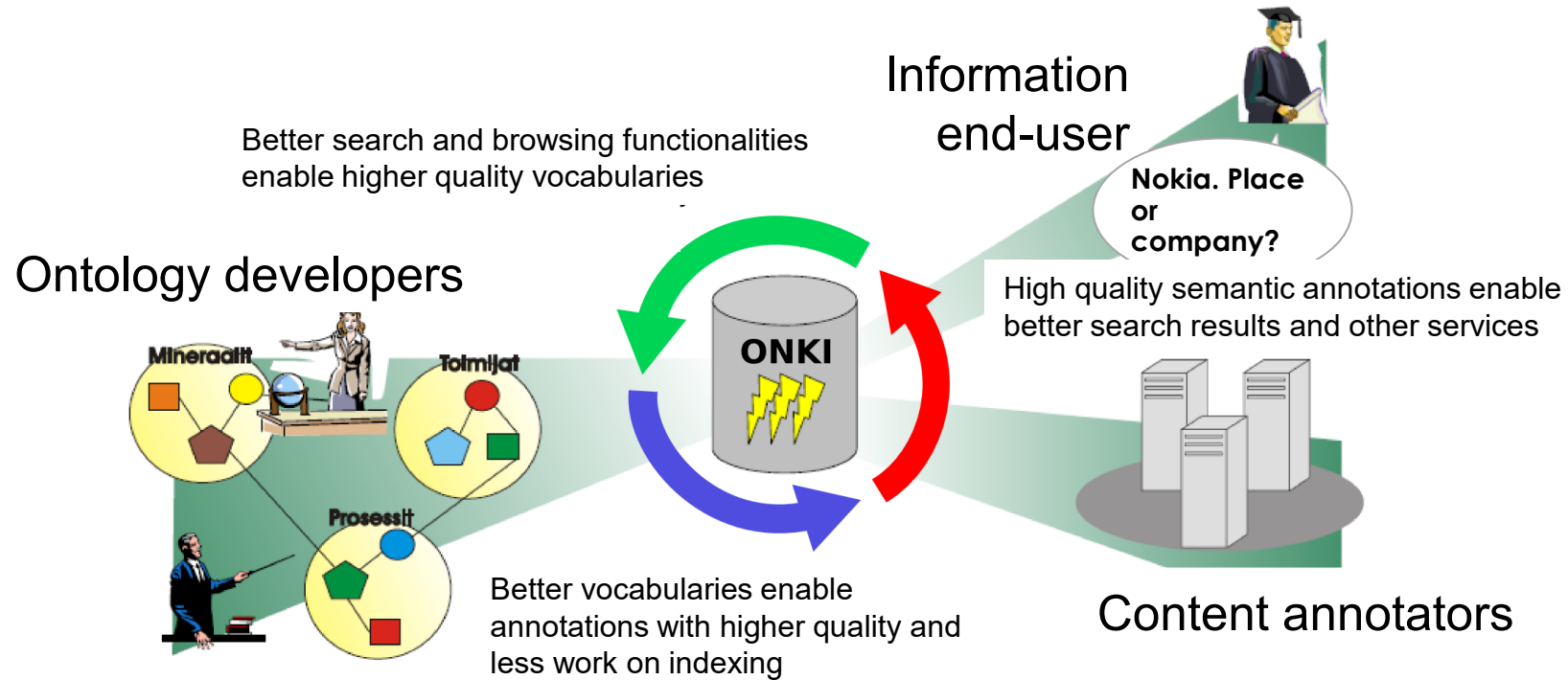
YSO Ontology



Semantic relations in use:
BT/NT and RT

Semantic relations in use:
subClassOf, partOf, and relatedConcept

Ontology library services: ONKI.fi concept



Supporters of the national semantic web infrastructure
Companies, government, EU, ...

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

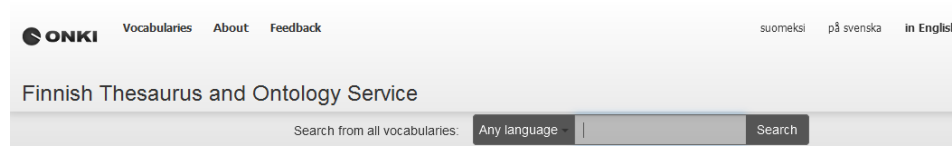


Permanent free national service funded by Finnish ministries

2019: 32 million API calls

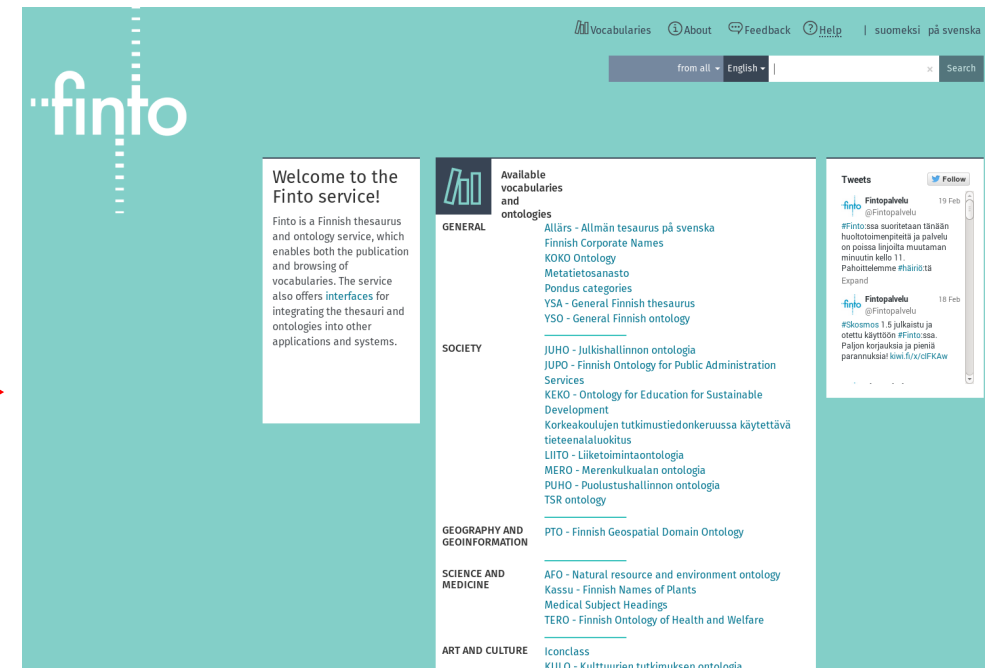
2020: 2,1 million page visits

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
Health	Ontology for Museum Domain
	Library of Congress Subject Headings - LCSH
	Finnish Ontology of Health and Welfare
Culture	Kaunokki Ontology
	The Finnish Ontology of Photography VALO
Public Administration	Kulttuurien tutkimuksen ontologia - KULO
	Finnish Music Thesaurus - MUSA/CILLA
	Kielitieteen ontologia - KTO
	Kirjallisuudentutkimuksen ontologia - KITO
	Puolustushallinnon ontologia - PUHO
Business	Julkishallinnon ontologia - JUHO
	Kunnat 2011
	Julkisten palveluiden ontologia JUPO
	Schools Online Thesaurus (ScOT)
Science	Merenkulkualan ontologia - MERO
	Kassu - Kasvien suomenkieliset nimet



YSO - General Finnish ontology

Content language English Search

A-Z Hierarchy Groups New

- events and action
- objects
 - abstract objects
 - physical objects
 - inanimate objects
 - matter
 - organic objects
 - abscesses
 - axons
 - body
 - capsid
 - carcasses
 - cell nucleus
 - cell walls
 - cells
 - cellular automata
 - chloroplasts
 - chromosomes
 - clones
 - galls (botany)
 - genes
 - malformations
 - membranes
 - microsatellites
 - mitochondria
 - organelles
 - organisms
 - parts of plants
 - parts of the body
 - pigment
 - polyps
 - receptors
 - scars
 - shell and peel
 - synapses
 - telomeres
 - tissues (organic objects)
 - physical whole
 - place
 - systems
 - properties

objects > physical objects > organic objects

PREFERRED TERM

organic objects

TYPE

Hierarchical concept

BROADER CONCEPT

physical objects

NARROWER CONCEPTS

abscesses
axons
body
capsid
carcasses
cell nucleus
cell walls
cells
cellular automata
cell walls
chloroplasts
chromosomes
clones
galls (botany)
genes
malformations
membranes
microsatellites
mitochondria
organelles
organisms
parts of plants
parts of the body
pigment
polyps
receptors
scars
shell and peel
synapses
telomeres
tissues (organic objects)

IN OTHER LANGUAGES

orgaaniset objektit	Finnish
orgaaninen rakenne	
organiska objekt	Swedish
organisk struktur	

URI

<http://www.yso.fi/onto/yso/p174>

Download this concept:

RDF/XML Turtle JSON-LD

Last modified 11/14/19

EXACTLY MATCHING
CONCEPTS

organic objects

KOKO Ontology

Images indexed with the term in Finna 0

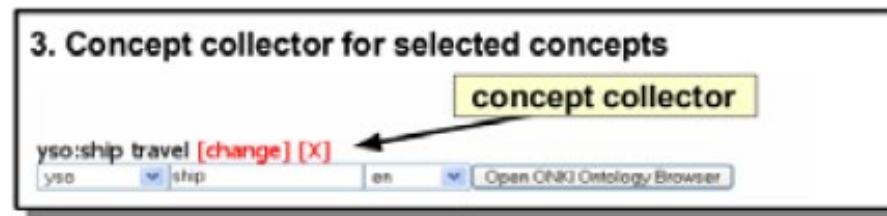
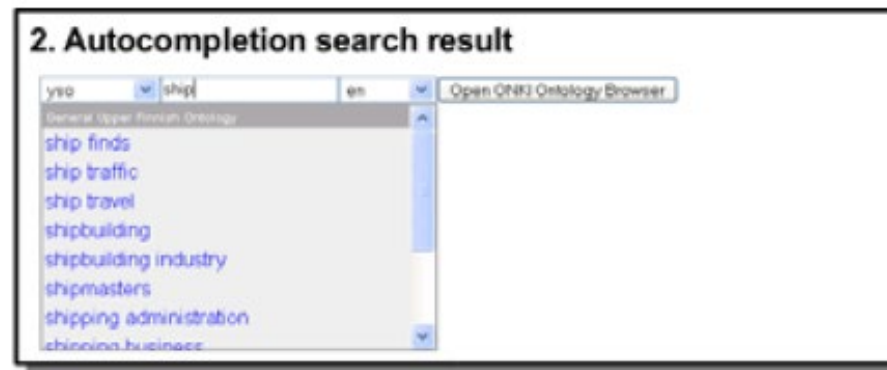
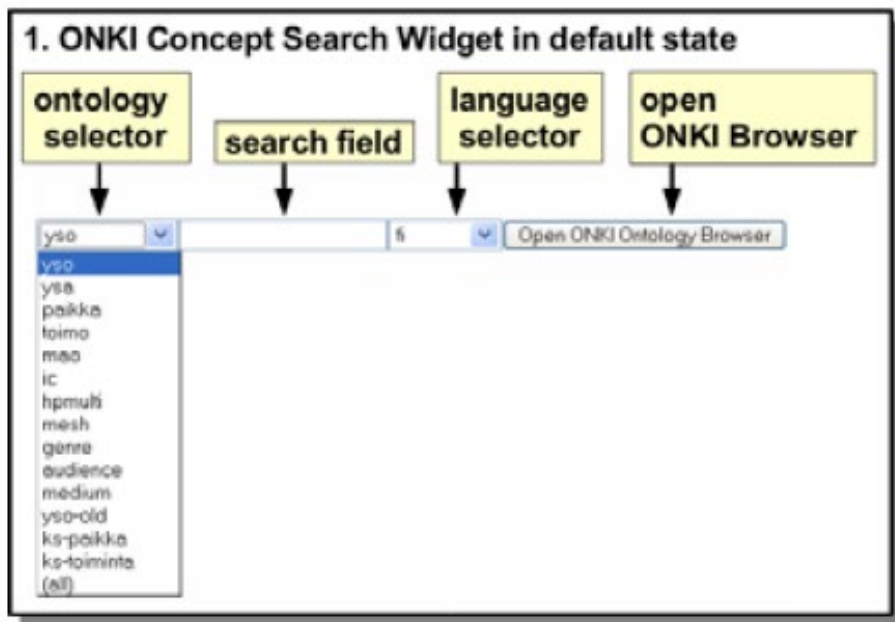
Image



ONKI Widget for Mashups



- Ontology services are automatically available after publishing a vocabulary or ontology with ONKI
- Simple AJAX-based widget for creating mash-ups





How to Deal with Overlapping Concepts in KOKO Ontology Cloud?

- YSO concepts are widely used in the domain ontologies
 - Up to 60% in some cases
 - YSO is good for a general upper ontology
- Domain ontologies also share concepts with each other
 - Much less than with YSO but up to 40% in some cases
- Lesson learned
 - Lots of redundant thesaurus work has been done in the intersecting areas
 - Collaborative re-organization of ontology work is needed, but is difficult
 - There is clearly need for merging some ontologies
 - How to deal with intersecting concepts?
 - How to manage distributed work of domain specific expert groups?
 - How propagate changes through the Linked Ontology Cloud?
 - Stakeholders work on these challenges in the Finto network / National Library

Actor Ontologies: Resolving Identities



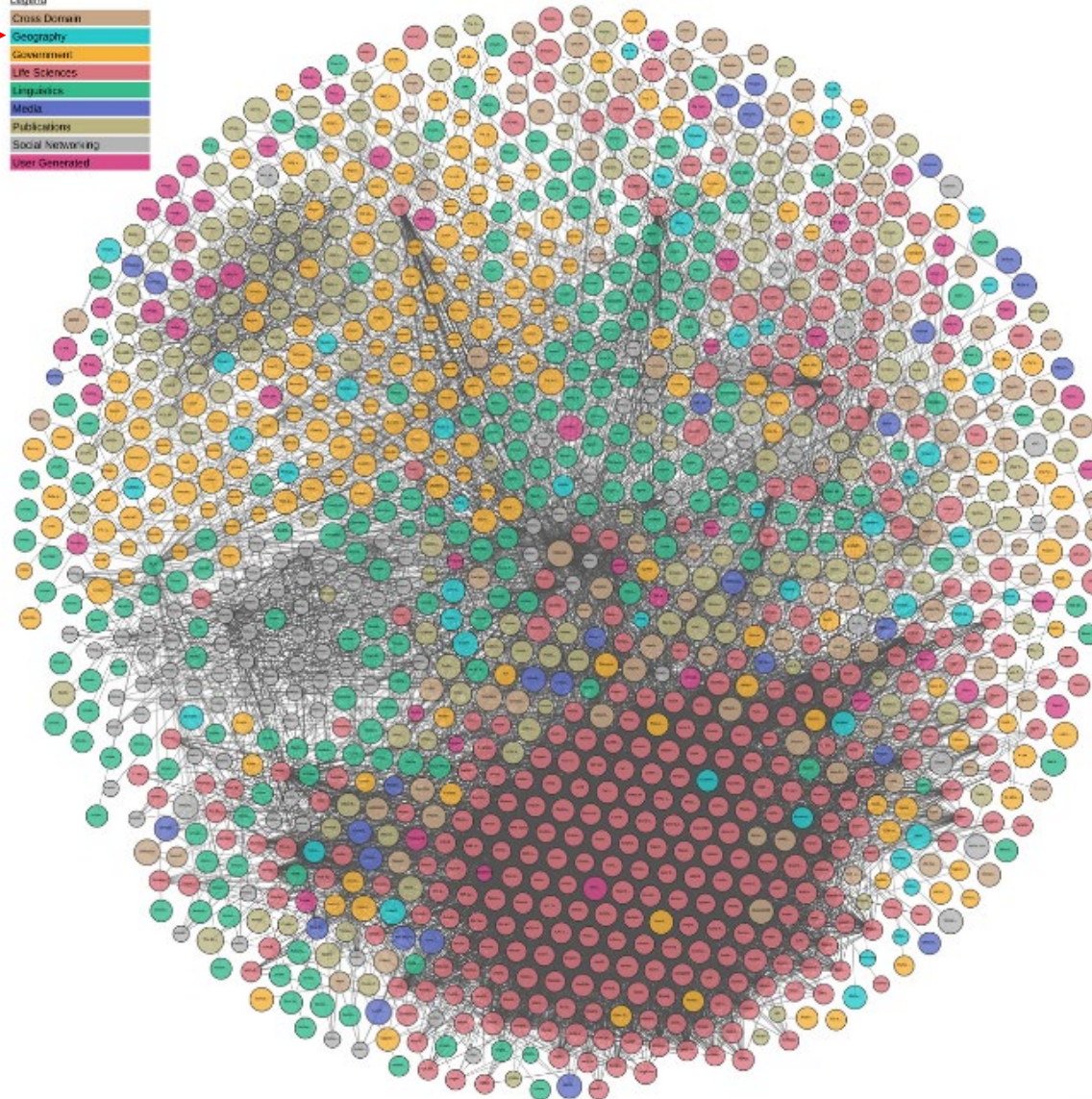
URI: http://dbpedia.org/resource/Pyotr_Ilyich_Tchaikovsky



Pjotr Tšaikovski (fi)
Пётр Ильич Чайковский (ru)
Pyotr Ilyich Tchaikovsky (en)
Pjotr Tjajkovskij (sv)
Pjotr Tsajkovskij (no)
Pjotr Iljitsch Tschaikowski (de)
Piotr Ilitch Tchaïkovski (fr)
Piotr Ilich Chaikowski (es)
Pëtr Il'ič Čajkovskij (it)
Pjotr Iljitsj Tsjaikovski (nl)
Piotr Ilitch Tchaikovsky (pt)
Piotr Czajkowski (pl)
Piotr Ilici Ceaikovski (ro)
Pjotr Iljics Csajkovszkij (hu)

Geography: A Key Element in the Linked Open Data Cloud

<https://lod-cloud.net/>



Semantic Web

LODstats.aksw.org:
10 000 datasets
150 000 000 000 triples

Finnish Ontology Service of Historical Places and Maps:

<http://hipla.fi>



hipla.fi Finnish Ontology Service of Historical Places and Maps About Project home

Select source dataset(s)

- ☐ Finnish municipalities (1939-44)
- ☒ Karelian map names (1922-44)
- ☒ Finnish Geographic Names (contemp.)
- ☐ SAPO (1865-2010)
- ☒ Getty TGN
- ☒ Kotus
- ☐ Suggested places

[+ Add a new place](#)

[View all places on current map view](#)

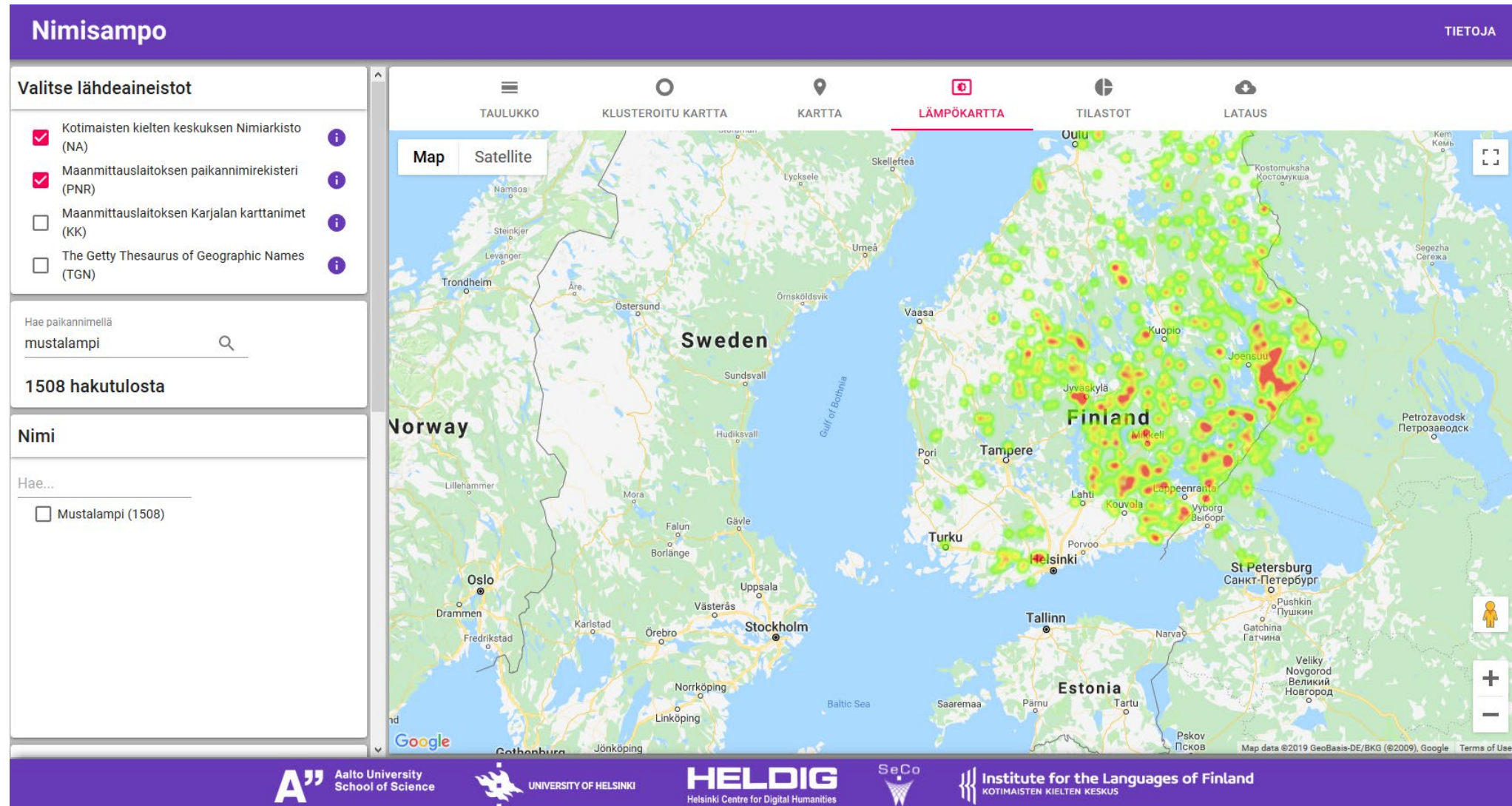
Search places [Maps](#)

musta*

Karelian map names (1922-44)

- MUSTasaari (Hypsographic feature, Johannes)
- Musta Riienlampi (Body of water, Ukuniemi)
- Musta-oja (Man-made feature)
- Mustajoensuu (Body of water, Kesälahti)
- Mustajoki (Body of water)
- Mustajoki (Body of water)
- Mustajoki (Village)
- Mustajoki (Body of water)
- Mustajoki (Body of water)

NameSampo: <http://nimisampo.fi>



There are 1508 places with name “Mustalampi” (= black small lake) in Finland!

Time Ontologies



- Modeling linear and cyclic time
- Time periods are different in different countries
 - *E.g., Bronze Age in Egypt and Nordic Countries*
- Modeling uncertainty in time

Event Ontologies



Events are "semantic glue" that link together:

- Places **where** events occur
- Times **when** events occur
- Actors **who** participate in events in roles
- Other related events



Shared Metadata Schemas

Two Main Approaches



Dublin Core approach

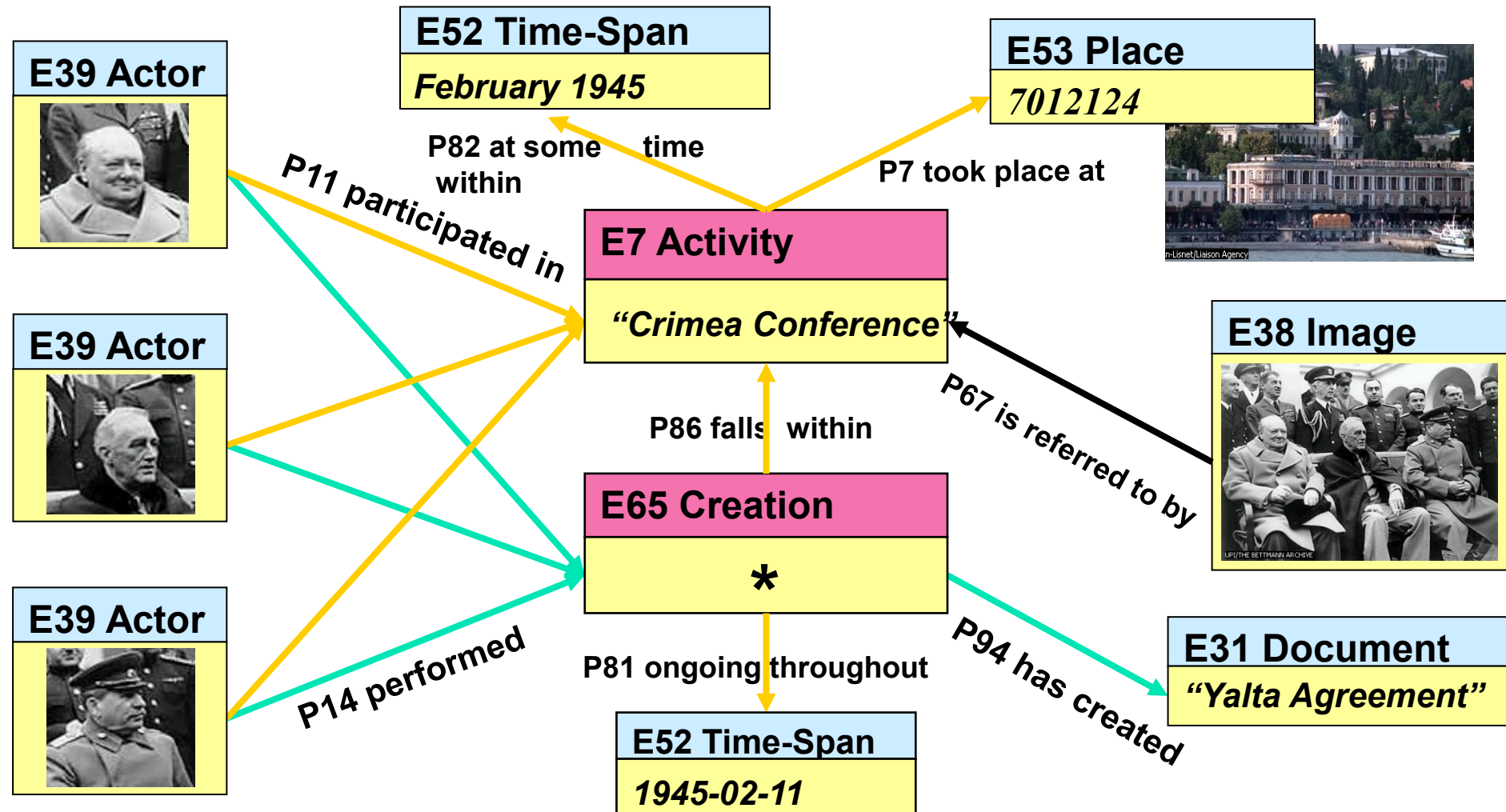
- Mapping/refining schemas using subproperties
- “Dumb down principle” is used
- <https://dublincore.org/>

Using foundational ontology models

- Different schemas are mapped onto a shared ontology
- CIDOC CRM is a prominent standard of this
 - <http://www.cidoc-crm.org/>

CIDOC CRM:

Using events as the foundation for knowledge representation



[Slide by: Stephen Stead]

Lessons Learned: Ontologies



- **Ontology development**
 - “Little semantics goes a long way” (Jim Hendler)
 - Little can mean a lot of work in big ontologies
 - Just transforming thesauri into SKOS format is useful
- **Distributed domain-specific ontology development**
 - Needed but creates new linking challenges
- **Centralized national ontology services**
 - Very useful in a country like Finland
- **Focus on sustainability processes**
 - This is never ending work
- **Project management & funding**
 - Make baby steps: demonstrate use case after each step
 - Create large collaboration networks including companies, too



Linked Data Service Infrastructure

Linked Data Finland 2012-

How to publish Linked Data?

5-star Linked Data model



- ★ Make data available on the Web in whatever format under an open license.
- ★★ 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.



(Tim Berners-Lee)
<http://5stardata.info>



An example of a Linked Data Service



MENU

CORONA MEASURES

DNB FOR USERS

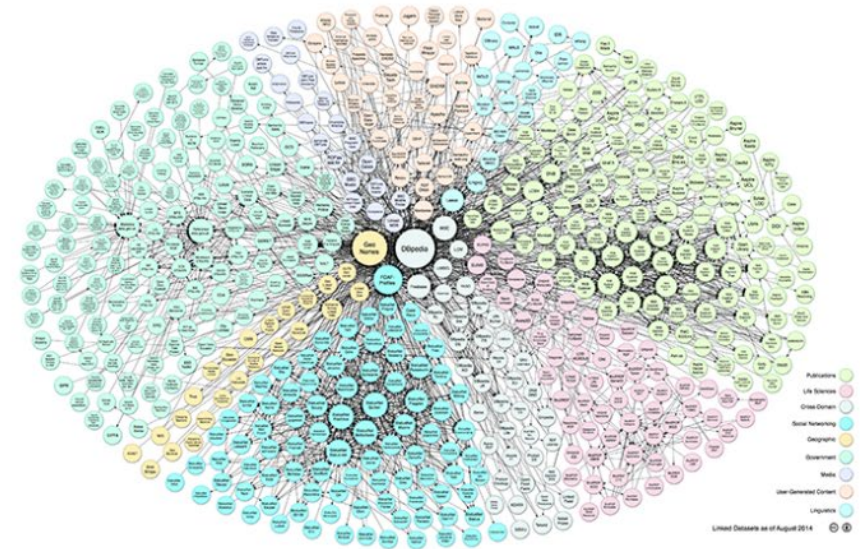
DNB PROFESSIONAL



Home > DNB Professional > Metadata Services > Linked Data Service

LINKED DATA SERVICE

- ▼ Overview
- ▼ Integrated Authority File (GND)
- ▼ Bibliographic data
- ▼ Test data
- ▼ Subscription Terms and Terms of Use
- ▼ Further development and service information
- ▼ Frequently asked questions (FAQ)
- ▼ Documentation
- ▼ Download
- ▼ Contact



Case: 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

Why LDF.fi?



Living Laboratory for publishing Linked Open Data

- Same idea as in **ontology services**
- But for **data** and **schemas**

Data Services for

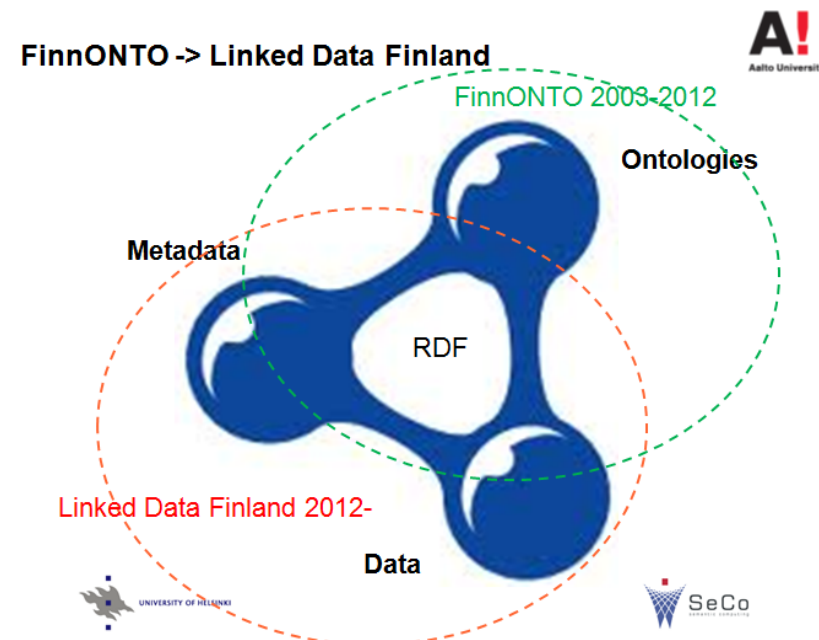
- Linked datasets
- Schemas

Links to

- Related services
- Related applications

Learning Center

- For publishing and using Linked Data



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

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



Example dataset: WarSampo Linked Data & SPARQL endpoint

<https://www.ldf.fi/dataset/warsa>



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

WarSampo

Sotasampo

Linked Data Finland



★★★★★

WarSampo Knowledge Graph includes harmonized data of different kinds concerning the Second World War in Finland, separated in different subgraphs representing events, actors, places, photographs, and other aspects and documentation of the war. The data covers the Winter War 1939-1940 against the Soviet attack, the Continuation War 1941-1944 where the occupied areas of the Winter War were temporarily regained, and the Lapland War 1944-1945, where the Finns pushed the German troops away from Lapland.

To test and demonstrate its usefulness, this Knowledge Graph is in use in the semantic portal [WarSampo](#), explained in more detail in the [project page](#).

The Knowledge Graph is published on [Zenodo](#) with a version history

Example SPARQL queries for the data:

- [Events, photographs and articles that are situated in Vyborg](#)
- [Casualties of the 1st Division and its subunits in the time interval 13.2.-13.3.1940 by place and date](#)

Data Download

The data can be downloaded at <https://zenodo.org/record/3431122/files/warsampo.zip>.

License

[CC BY 4.0](#)



Licensors: [Kansallisarkisto](#), [Semanttisen laskennan tutkimusryhmä \(SeCo\)](#)

See possible graph-specific licenses below.

Detailed Dataset Contents

Karelian map names 1922-44 (URI: http://ldf.fi/warsa/places/karelian_places)

([Browse data](#) / [View in Sotasampo.fi](#))



This graph contains Finnish map names from the Karelian region (currently divided between Russia and Finland). The source data was a CSV table with roughly 40 000 map names, which were picked from Karelian maps (dated 1922-1944) by Jyrki Tiittanen (National Land Survey of Finland). The CSV table provided a label,



Software Tools for the Semantic Web



Languages & standards of W3C and others

- *Data exchange language:* *RDF*
- *Vocabulary/schema languages:* *SKOS, OWL*
- *Data/ontology query language:* *SPARQL*
- *Rules for reasoning:* *RIF, SWRL, ...*
- *Metadata and ontology models* *DC, CIDOC CRM, ...*

Triple stores for data services

- *Fuseki, Sesame, Redland, Virtuoso, ...*
- *<http://en.wikipedia.org/wiki/Triplestore>*

Development tools

- **Ontology editors**
 - *Protégé* <https://protege.stanford.edu/>
 - *TopBraid Composer* <https://www.topquadrant.com/topbraid-composer-install/>
- **Software development tools**
 - *Java: Apache Jena* <https://jena.apache.org/>
 - *Python: RDFLib* <https://pypi.org/project/rdflib/>

Lessons Learned: Data Services



- **LDF.fi platforms makes data publishing very easy**
 - Services are generated automatically from Service Description metadata
- **Reusing services and support functions is cost-efficient**
 - Unnecessary re-implementations can be avoided
- **LDF.fi is important for Human Infrastructure building**
 - In hackathons
 - In educational university courses
- **Basic services can be run with little maintenance**
 - However, technical expertise is needed
- **Sustainability is needed**
 - Part of national research infrastructure roadmap of the Academy of Finland
 - Servers provided “for free” for universities by CSC / Ministry of Education and Culture
- **Data maintenance issues remain a challenge**
 - How to keep the data & services up-to-date



3. HOW TO USE SW INFRA FOR APPLICATIONS?

Sampo Model and Sampo Series of Systems

Video “The Semantic Web and AI for Digital Humanities”: <https://vimeo.com/470313703>

video

Applications: Cultural Heritage

"Samos" 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 users in 2020]
5. **WW1LOD** — World War I Linked Open Data (2014)
6. **WarSampo** — Finnish World War 2 on the Semantic Web (2015-19) [742 000 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) [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)
13. **AcademySampo** — Finnish Academic People 1640 – 1899 (2021)
14. **FindSampo** — Archaeological Finds on the Semantic Web (2021)
15. **LetterSampo, LawSampo, ParliamentSampo**, ... underway



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



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



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



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



Sampo Model Principles



Table 1. Sampo Model Principles P1–P6

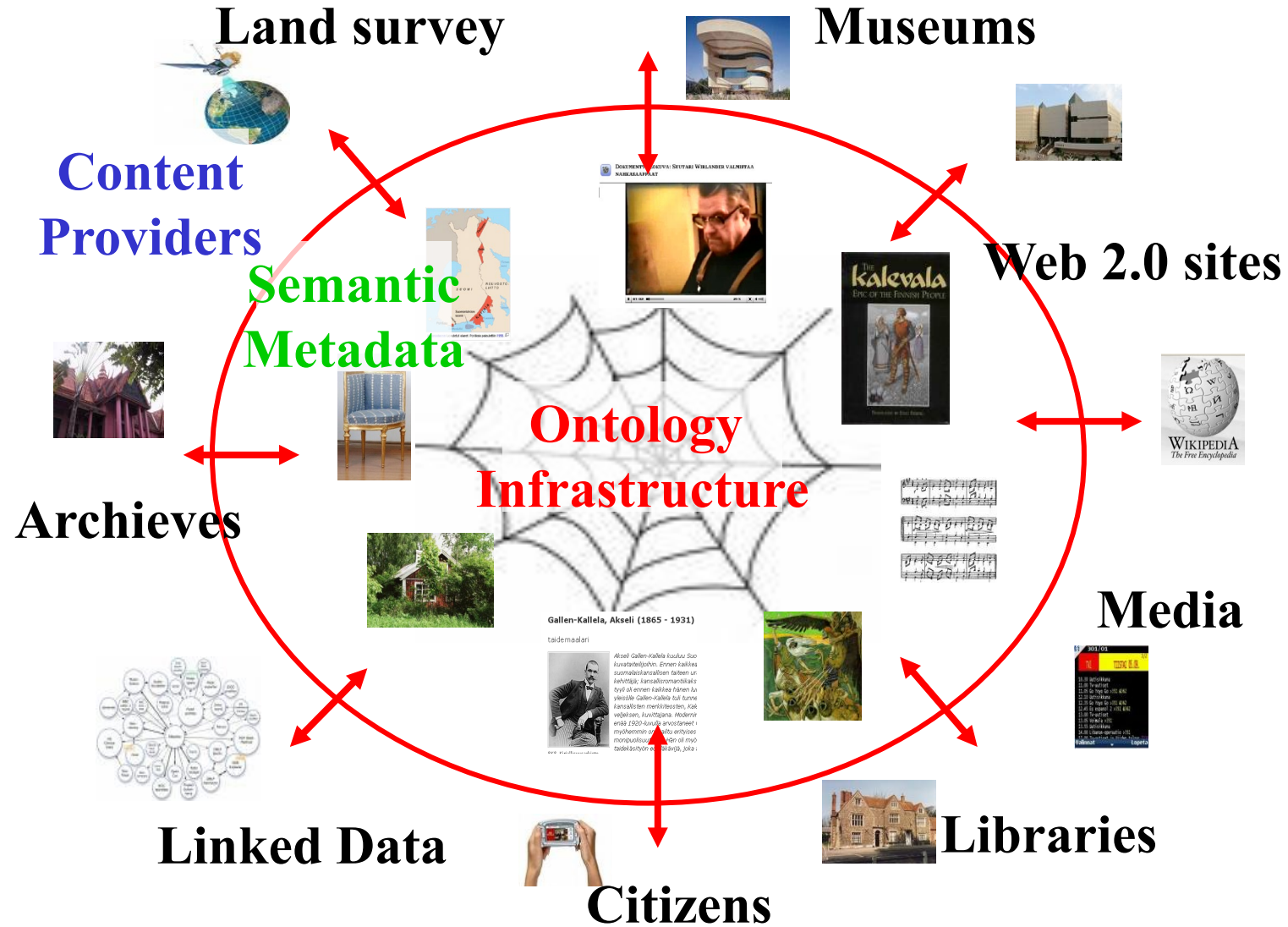
P1. Support collaborative data creation and publishing
P2. Use a shared open ontology infrastructure
P3. Support data analysis and knowledge discovery in addition to data exploration
P4. Provide multiple perspectives to the same data
P5. Standardize portal usage by a simple filter-analyze two-step cycle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



<https://seco.cs.aalto.fi/publications/2021/hyvonen-sampo-model-2021.pdf>

P1. Support collaborative data creation and publishing





P2. Use a shared open ontology infrastructure



Elements of National Ontology Infrastructure for Digital Humanities



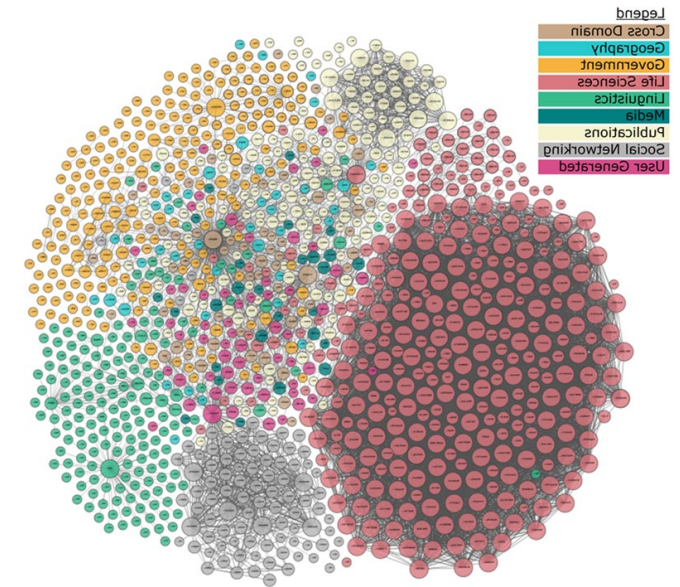
- **Domain ontologies**

- Historical Places and Maps
- Historical Persons
- Historical Times
- Historical Events
- Historical Keyword Concepts
- ...

- **Shared metadata models**

- Dublin Core, CIDOC CRM, ...

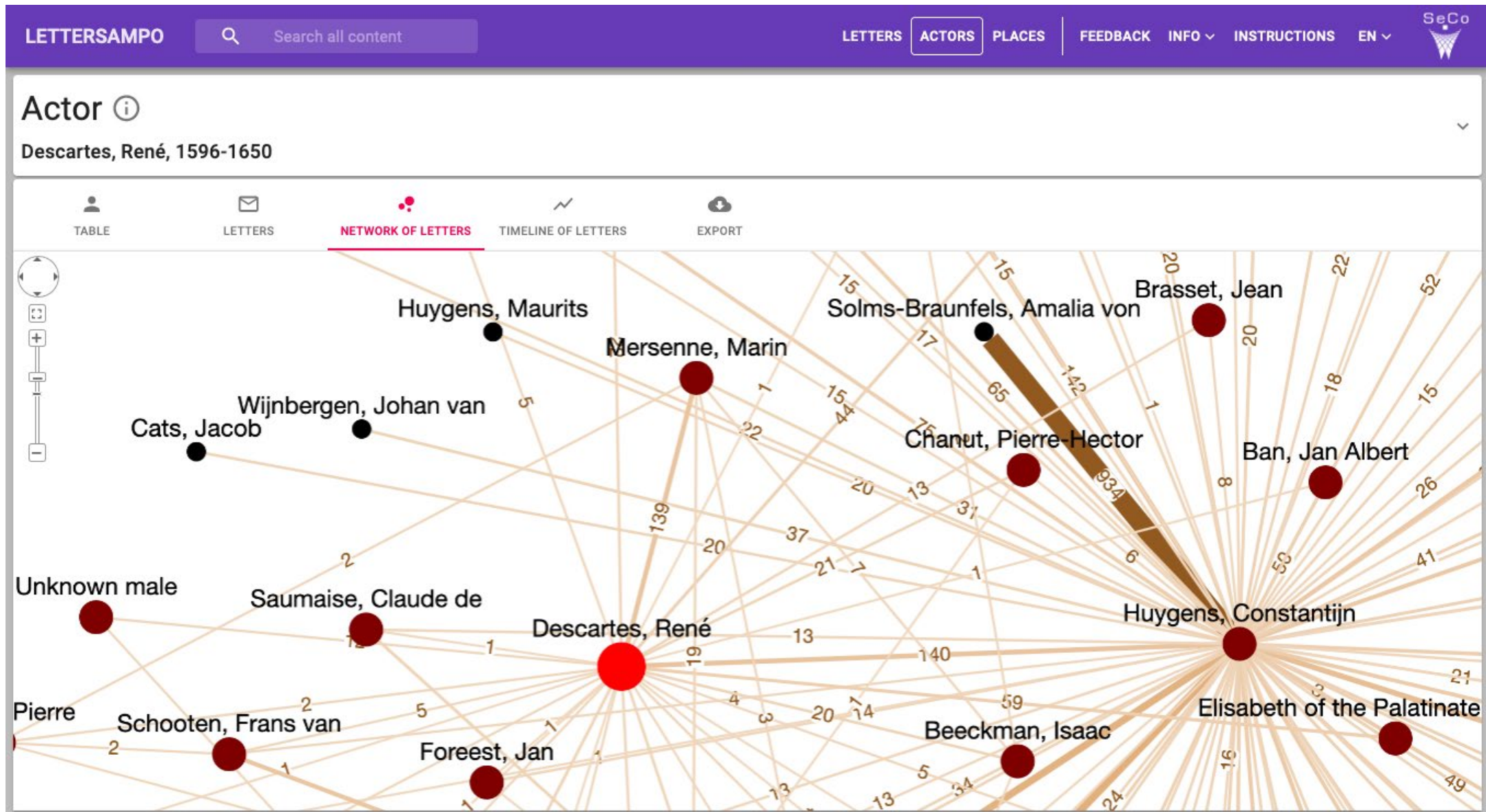
- **Ontology and linked data services online**



P3. Support data analysis and knowledge discovery in addition to data exploration



LetterSampo: Network analyses of Early Modern Correspondences: Case Rene Descartes



BiographySampo: Correlations of Vocational Groups between Parents and Children using Google Colab



M. Tamper et al. / Analyzing Biography Collections Historiographically as Linked Data: Case National Biography of Finland

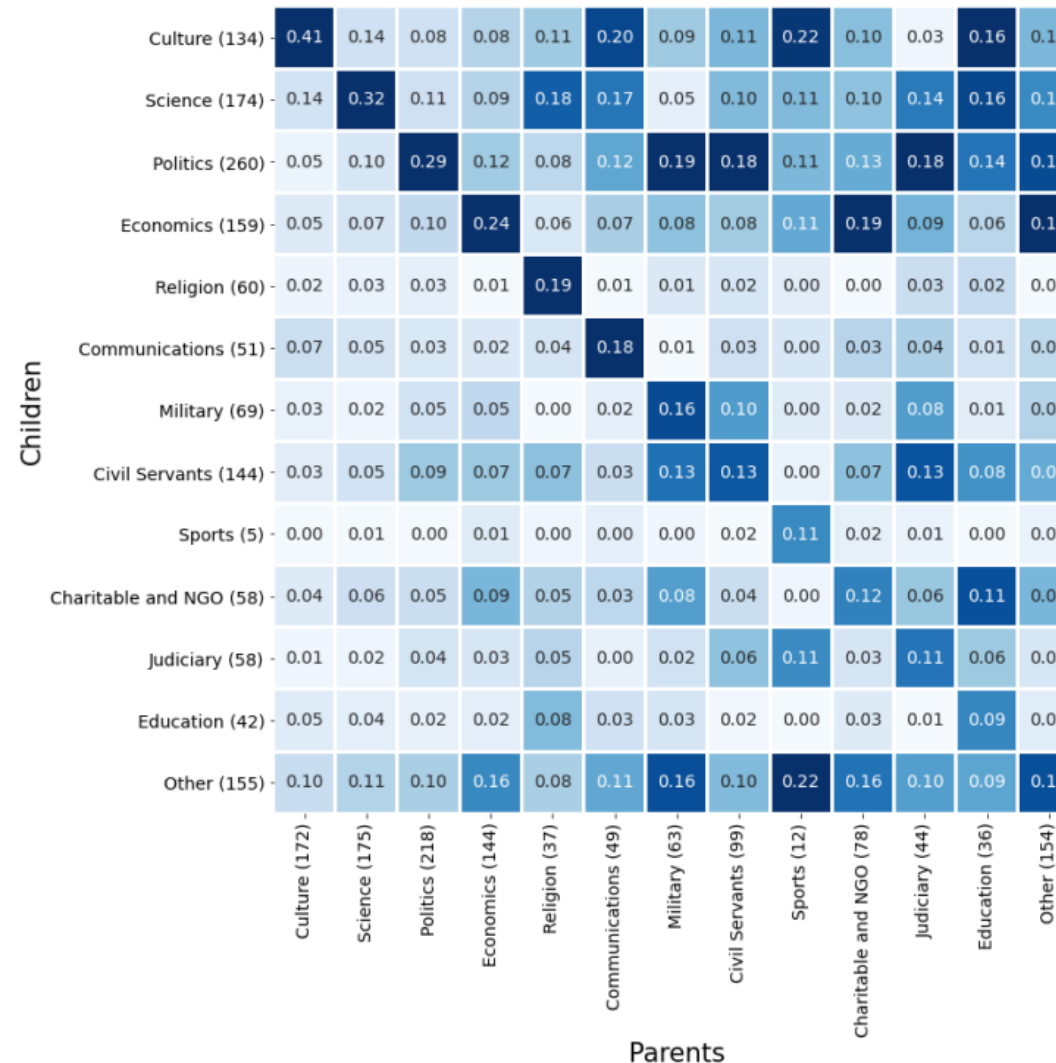



FIG. 13.: Correlations between the vocational groups of parents and children

P4. Provide multiple perspectives to the same data



WarSampo – Finnish World War II on the Semantic Web: Nine perspectives to war history




Perspectives ▾ Suomeksi Feedback Info ▾ Instructions

WarSampo

Finnish World War II on the Semantic Web

The WarSampo Portal enables both historians and laymen to study the war history and destinies of their family members in the war from different interlinked perspectives ([instructions](#)).

 [Join the WarSampo Facebook group](#)

Select a perspective to search and browse the WarSampo data

Events

Events of the Winter and Continuation War visualized using a timeline and a map with related linked data

Persons

Data about persons with related links from various sources

Army Units

Events and other related data about army units visualized using i.a. maps

Places

Search and browse places and maps covering the war zone area in Finland and discover additional data such as events and photographs linked to places

Kansa taisteli magazine articles

Faceted semantic search for Kansa taisteli magazine articles containing mostly memoirs of soldiers related to WW2

Casualties

A table-like view of war casualty records that can be filtered using faceted semantic search, enriched with links to other WarSampo datasets

Photographs

Browse the content of the Finnish Wartime Photograph Archive with faceted search

War Cemeteries

War cemeteries of Finland with photographs

Prisoners of War

Finnish prisoners of war in the Soviet Union

P5. Standardize portal usage by a simple filter-analyze two-step cycle



Active filters

Hierarchical checkbox facet

Perspective tabs

Active filters:

Production place: Egypt (ancient) X Production date: -500 to 500 X

Narrow down by:

Label (i) Author (i) Work (i)

Production place (i)

Search...

☐ Unknown [377]

☐ World [1353]

☐ Africa [1273]

☐ Egypt [308]

☒ Egypt (ancient) [964]

☐ North Africa [1]

☐ Asia [10]

☐ Europe [76]

Production date (i)

-500 500

Min year -500 Max year 500

APPLY

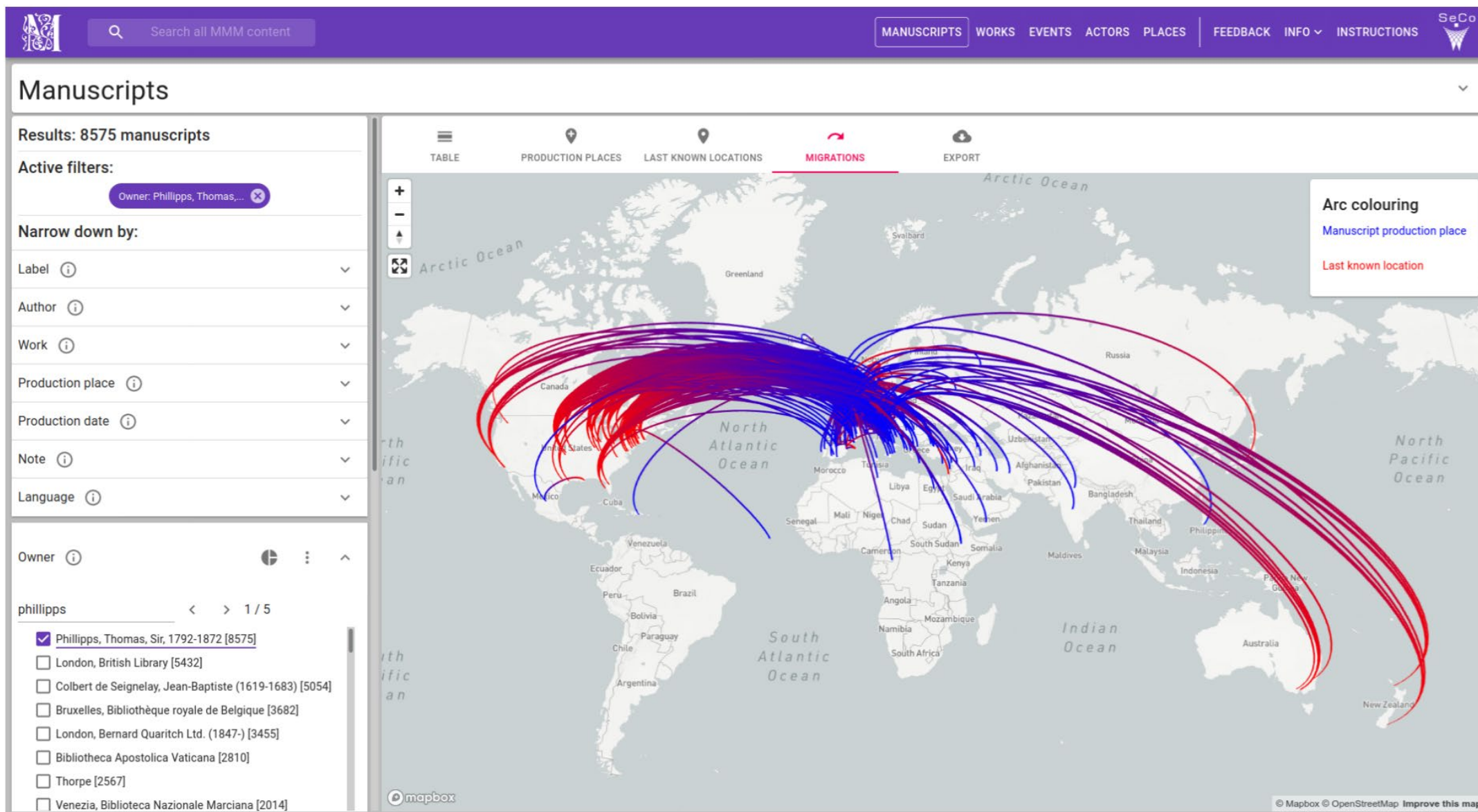
Table:

Label (i)	Author (i)	Work (i)	Expression (i)	Production place (i)	Production date (i)
MS_Gr_class.d.97 (P)	Achilles Tatius	Achilles Tatius: Leucippe et clitophon [Greek]	Expression of 'MS_Gr_class.d.97 (P)'	Al Bahnasā ...	0200 - 0
MS_Gr_class.g.51 (P)	Aeschines	Aeschines: In ctesipontem [Greek]	Expression of 'MS_Gr_class.g.51 (P)'	Al Bahnasā ...	0200 - 0
MS_Gr_class.e.72 (P)	Aesop ...	Aesop: Fables [Greek] ...	Expression of 'Fabulae' ...	Egypt (ancient)	0500 - 0
MS_Gr_class.d.80 (P)	Anaximenes of Lampsacus	Anaximenes of Lampsacus: Rhetorica ad alexandrum [Greek]	Expression of 'MS_Gr_class.d.80 (P)'	Egypt (ancient)	-030 - 0
MS_Gr_class.d.75 (P) - Part 2	Anubio	Anubio: Astrological epigrams [Greek]	Expression of 'MS_Gr_class.d.75 (P) - Part 2'	Al Bahnasā ...	0290 - 0
MS_Gr_class.e.44 (P)	Apollonius Rhodius	Apollonius Rhodius: Lexicon homericum [Greek]	Expression of 'MS_Gr_class.e.44 (P)'	Egypt (ancient)	0100 - 0
MS_Gr_class.f.112 (P)	Apollonius Rhodius	Apollonius Rhodius: Argonautica [Greek]	Expression of 'MS_Gr_class.f.112 (P)'	Egypt (ancient)	0300 - 0
MS_Gr_class.g.44 (P) - Part 2	Aristophanes	Aristophanes: Plutus [Greek]	Expression of 'MS_Gr_class.g.44 (P) - Part 2'	Egypt (ancient)	0100 - 0
MS_Gr_class.e.87 (P)	Aristophanes	Aristophanes: Lysistrata [Greek]	Expression of 'MS_Gr_class.e.87 (P)'	Egypt (ancient) ...	0300 - 0
MS_Gr_class.f.72 (P)	Aristophanes	Aristophanes: Knights [Greek]	Expression of 'MS_Gr_class.f.72 (P)'	Egypt (ancient) ...	0350 - 0
MS_Gr_class.d.76 (P)	Aristotle	Aristotle: Protreptikos [Greek]	Expression of 'MS_Gr_class.d.76 (P)'	Al Bahnasā ...	0100 - 0
MS_Gr_class.c.77 (P)	Callimachus	Callimachus: Aitia [Greek]	Expression of 'MS_Gr_class.c.77 (P)'	Al Bahnasā ...	0001 - 0
MS_Gr_class.g.60 (P) - Part 2	Callimachus	Callimachus: Iambus [Greek]	Expression of 'MS_Gr_class.g.60 (P) - Part 2'	Al Bahnasā ...	0100 - 0
MS_Gr_class.c.72 (P)/1-7	Callimachus	Callimachus: Aitia [Greek] ...	Expression of 'Aitia' ...	Al Bahnasā ...	0300 - 0

Link to entity landing page

Figure 6. A selection of SAMPO-UI components for building a faceted search perspective of a semantic portal.

Mapping Manuscripts Migrations: Case Thomas Phillips (1792-1872)



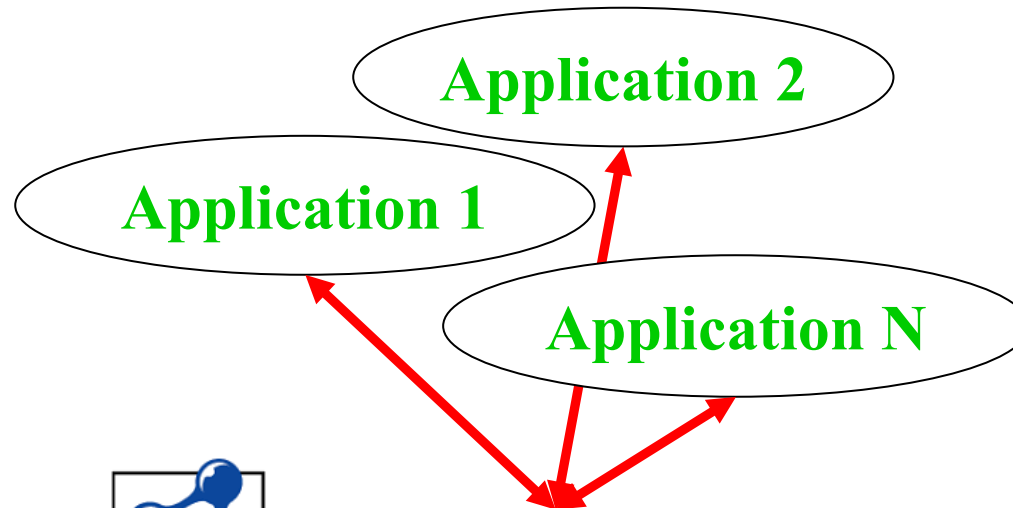
P6. Make clear distinction between the LOD service and the user interface (UI)



Linked Data Publishing Model



Client Side
(Browser)



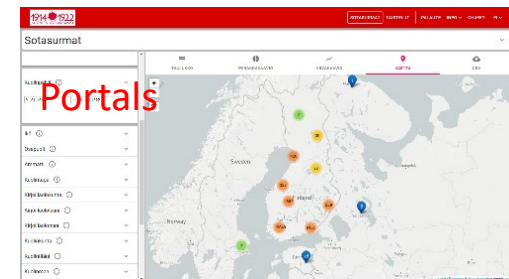
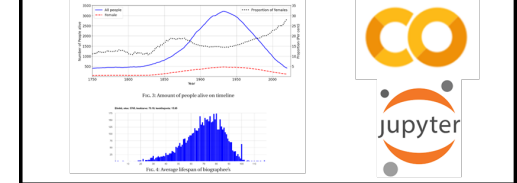
SPARQL End Point



Server Side
<https://ldf.fi>

Linked Data Finland Service
WWW Standard Model

Google Colab, Jupyter, YASGUI, etc.



Portals

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 [link 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:



An Example of A Sampo Portal and Data Service

**LetterSampo – Historical Letters on the Semantic Web
(2021)**

Example Video of a Sampo System: <https://vimeo.com/461293952>

"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

Lessons Learned: Sampo Applications



- Domain agnostic SW standards and practices can be applied on “all” domains
- National SW infrastructure is the key for cost-efficient application development
- It is possible to create popular systems in use with novel features
- Reusable software tools such as Sampo-UI quite essential
- Sustainability of applications remains a challenge
 - Technologies are new to data owners
- Paradigm change in web publishing
 - new avenues to the future are open ...



SUMMARY



Re-usable, shared LOD infrastructure is the key for successful Semantic Web Applications



But the Lunch is not Free

- More collaboration is needed -> complicates work
- Integration of semantic systems with legacy systems
- Manual annotations are costly and may not scale up
- Automatic annotation and linking lowers data quality



Need more source criticism and data literacy!

- What the data actually is and mean?
- Big data quality issues: completeness, skewness, errors



More Information – Questions?

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



Semantic Computing Research Group (SeCo)

[Making computers and the web more intelligent and interoperable](#)

- » [print this page](#)
- » [Follow us on Twitter](#)
- » [Be our friend on Facebook](#)

HOME PAGE

MISSION
NEWS
PROJECTS
EVENTS
PEOPLE
PARTNERS
CAREERS
TEACHING
CONTACT

PUBLICATIONS
APPLICATIONS
ONTOLOGIES
LINKED DATA
INFRA FOR DH
SERVICES
TOOLS
SOFTWARE
DOWNLOAD

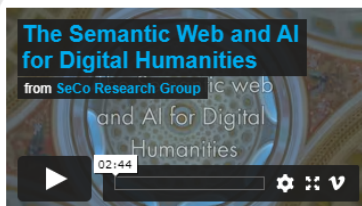
Wikipedia: [Semantic web suomeksi](#)

Google this site

SeCo Welcomes You!

Semantic Computing Research Group (SeCo) researches machine-processable semantics related to, e.g., the Semantic Web. We are located at the [Aalto University, School of Science, Department of Computer Science](#) and [HELDIG – Helsinki Centre for Digital Humanities](#), University of Helsinki, Faculty of Arts. Our group consists of researchers both from the Aalto University, School of Science, Department of Computer Science and [University of Helsinki, Department of Computer Science](#), and HELDIG.

Our research is [focused](#) on semantic technologies, such as the Semantic Web and intelligent web services. In addition to research and [publications](#), we also create prototype applications that demonstrate the new possibilities of semantic technologies, such as **semantic portals for end users**, **semantic infrastructural data and ontology services**, and **ontologies and tools** for creating semantic applications.



450 publications about the infras and their applications:

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

Videos:

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

Application Domains

Our work is highly cross-disciplinary, including application domains, such as Digital Humanities, Health, Learning, Government, Commerce, Geography, and Biology.

Selected SeCo Applications and Demos for End Users



[Mapping Manuscript Migrations](#)
Online since 2020. [Try it!](#)



[WarVictimSampo 1914-1922](#)
Online since 2019. [Try it!](#)



[BiographySampo - Finnish Life Stories on the Semantic Web](#)
Online since 2018. [Try it!](#)



[U.S. Congress Prosopographer](#)
Online since 2018. [Try it!](#)



[Norrsi Alumni on the Semantic Web](#)
Online since 2017. [Try it!](#)



[Semantic Finlex - Finnish Law](#)



[WarSampo - Second World War](#)



[Hipla.fi - Finnish Ontology](#)



[Semantic National Biography](#)



[BirdWatch - Mobile Semantic](#)

Latest News

[Five new SeCo papers accepted for the ISWC 2021](#)
The 20th International Semantic Web Conference (ISWC 2021), the...

2.8.2021 6:53 by eahyvonen

[FindSampo ? Finnish Archaeological Finds on the Semantic Web](#)

FindSampo – Finnish Archaeological Finds on the Semantic

[starting in](#)
/ intangible

Eero Hyvönen: [Relevance Feedback Search Based on Automatic Annotation and Classification of Texts](#)

Eero Hyvönen: [How to Create a National Cross-domain Ontology and Linked Data Infrastructure and Use It on the Semantic Web](#)

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](#)

Petri Leskinen and Eero Hyvönen: [Reconciling and Using Historical Person Registers as Linked Open Data in the AcademySampo Knowledge Graph](#)

SeCo on Twitter

Tweets by @secoresearch

SeCo Research Group Retweeted



ManuscriptMigrations
@MSMigrations

"Harmonizing and publishing heterogeneous premodern manuscript metadata as linked Open