



Aalto University
School of Science



Semantic Web Infrastructures

CS-E4410 Semantic Web, 17.2.2021

Eero Hyvönen

Aalto University, Semantic Computing Research Group (SeCo) <http://seco.cs.aalto.fi>

University of Helsinki, HELDIG

<http://heldig.fi>

eero.hyvonen@aalto.fi

Learning Objectives



Understand why infrastructures are needed in applications

Learn what kind of infrastructures are available

Learn about work on Finnish Semantic Web infrastructures

Contents

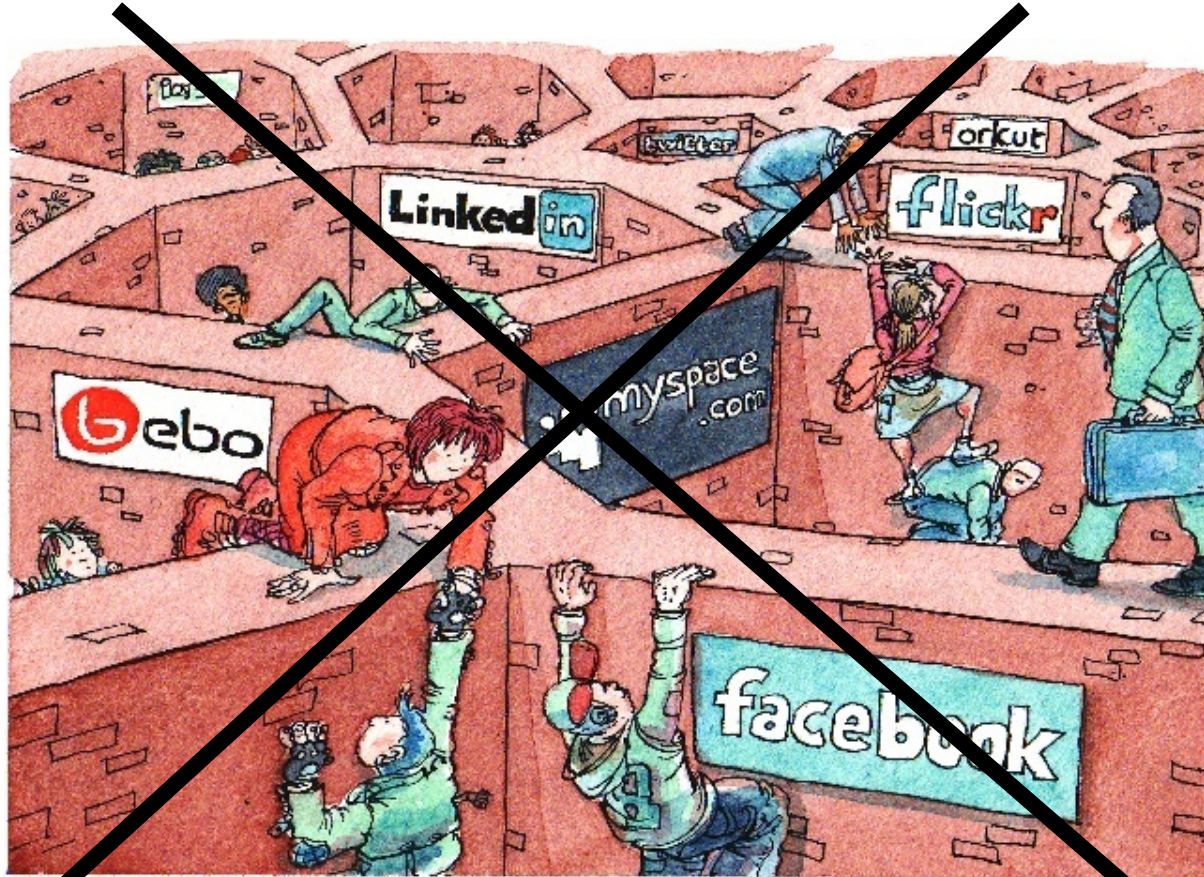


- **Why infrastructures are needed?**
- **Infrastructure types**
 - Ontologies & ontology services
 - Metadata schemas
 - Data & Linked Data services
 - Software tools for developers
- **Case: Finnish Linked Open Data Infrastructure for Digital Humanities**



Infrastructures for the Semantic Web

Problem: Interoperability of Data



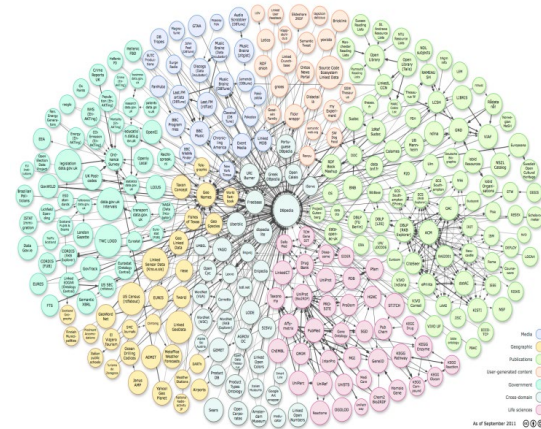
Solution: Content Infrastructure



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



Semantic Content Infra:
Ontologies, data, metadata



Infrastructure Types



Shared **ontologies** as services

- Creating a library service of mutually interoperable vocabularies/ontologies
- Developing the ontologies in collaboration

Shared **metadata schemas**

- Representing different information types, e.g., museum items, people, places, events

Shared **Linked Open Data** & services

- Reducing multiple work
- Enriching each others data

Shared **software and tools**

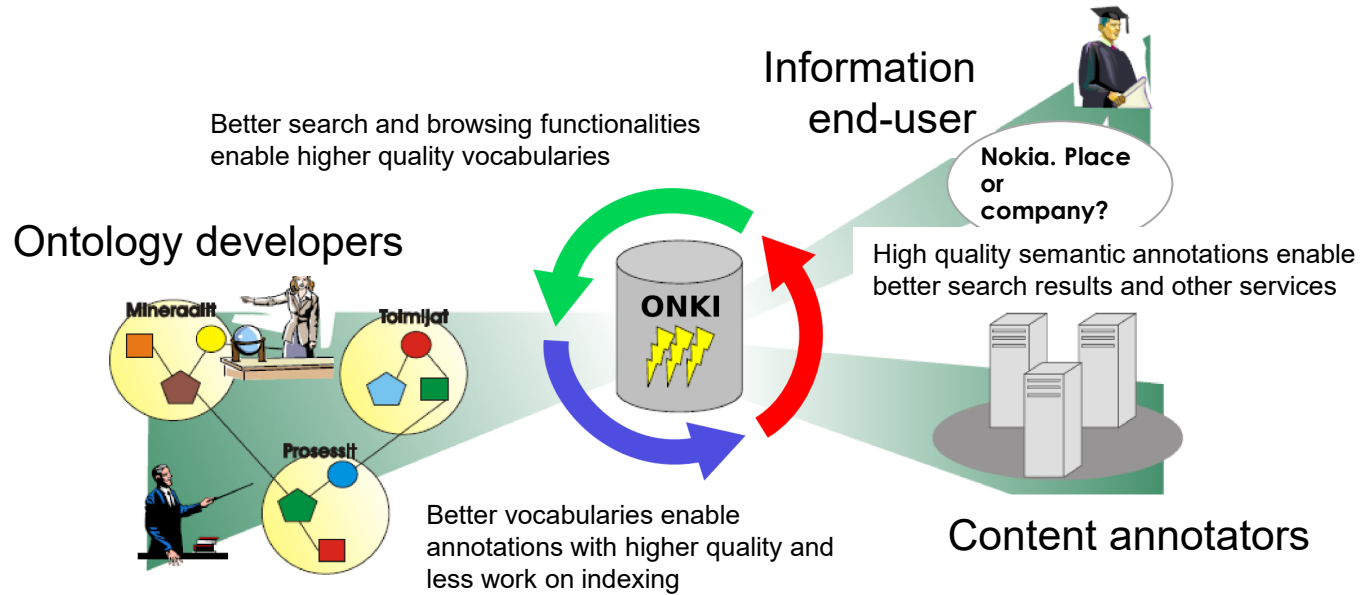
- Re-using existing results
- Not re-inventing the wheel again and again



Shared Ontologies & Ontology Services

Ontology library services: ONKI concept

– Users & interest groups



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

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

Examples

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

KOKO: Linked Open Ontology cloud



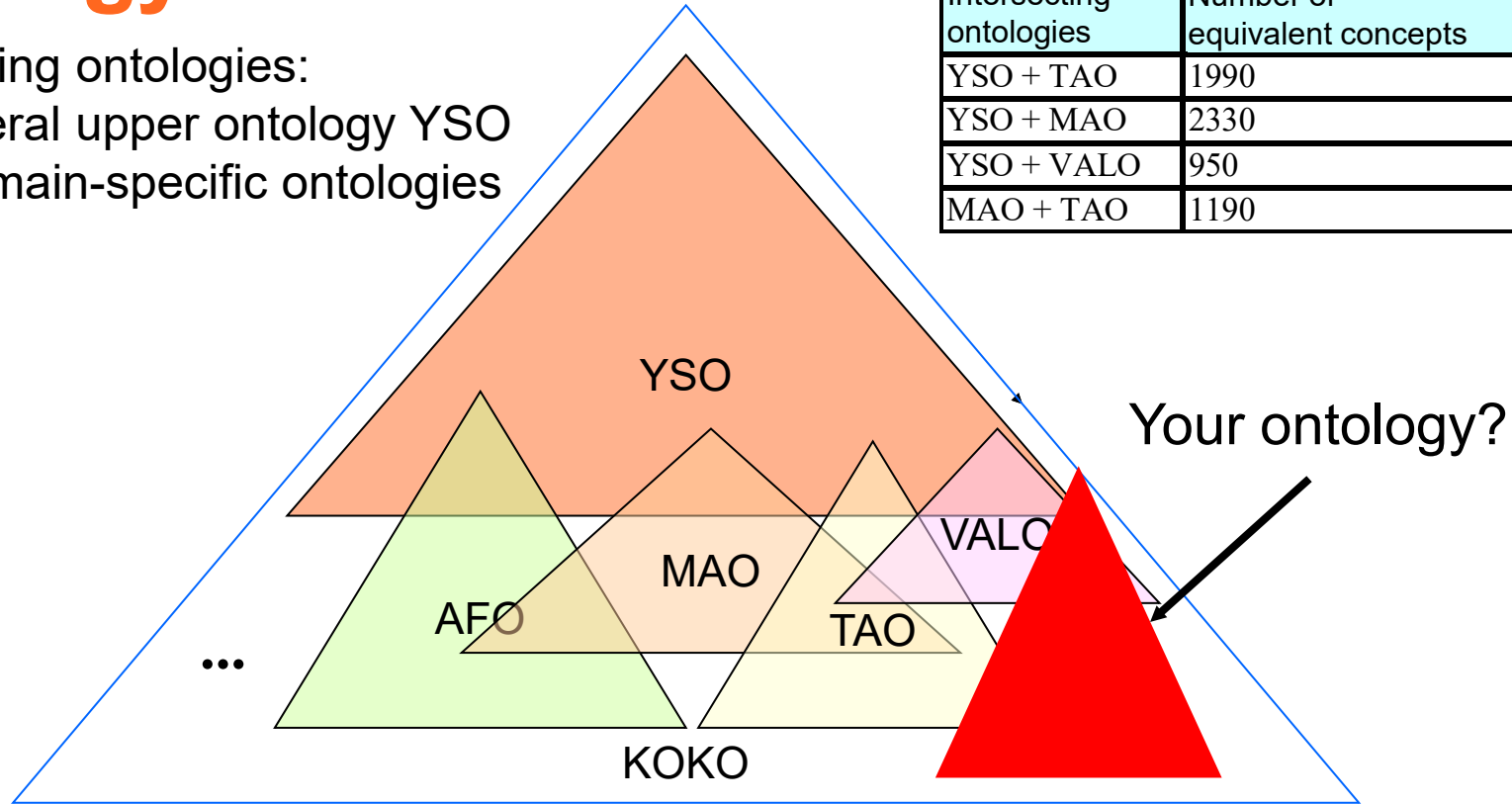
	Name	Ontology domain	Underlying thesaurus	Size	Maintaining Organization
1	YSO	General domain	General Finnish Thesaurus, YSA, Allärs	23700	National Library, Åbo Academy
2	MUSO	Music	Thesaurus of Music, MUSA/CILLA	1000	National Library
3	MAO	Museum domain	Thesaurus of Museum Domain, MASA	6800	National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



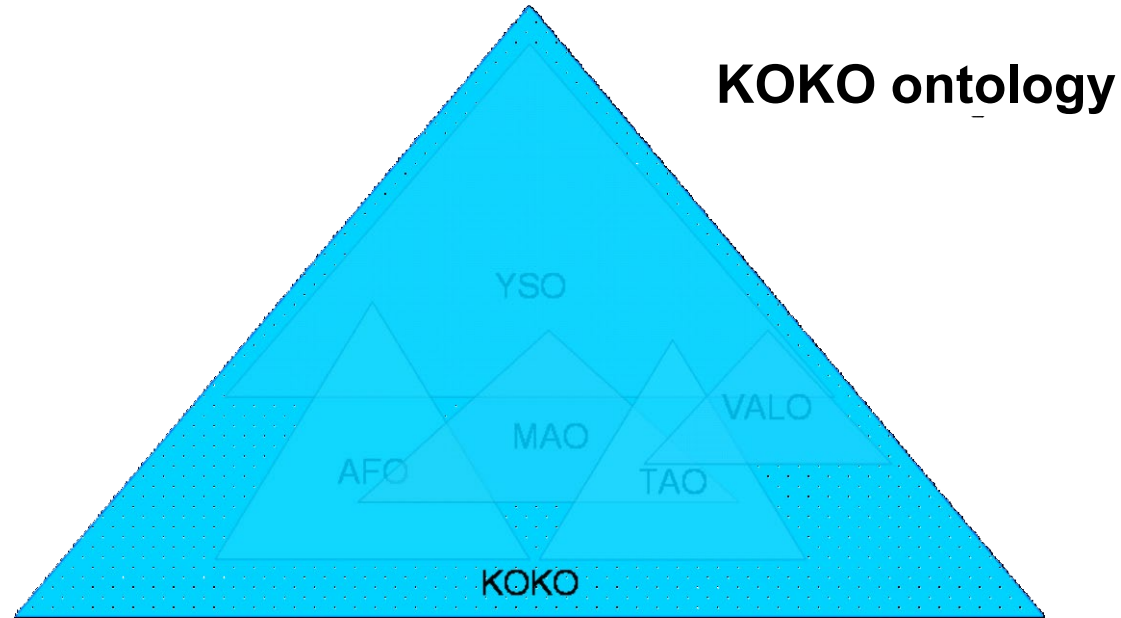
Aligning ontologies:
General upper ontology YSO
+ domain-specific ontologies

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



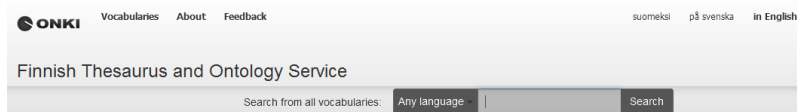
[Hyvönen et al., ESWC 2009]

KOKO from the “end-user” viewpoint



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

Permanent free national service funded by Ministry of Education and Culture and Ministry of Finance
2019: 32 million API calls



Available vocabularies and ontologies

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



Available vocabularies and ontologies

- GENERAL**
 - Allars - Allmän tesauros på svenska
 - Finnish Corporate Names
 - KOKO Ontology
 - Metatietosanasto
 - Pondus categories
 - YSA - General Finnish thesaurus
 - YSO - General Finnish ontology
- SOCIETY**
 - JUHO - Julkishallinnon ontologia
 - JUPO - Finnish Ontology for Public Administration Services
 - KEKO - Ontology for Education for Sustainable Development
 - Korkeakoulujen tutkimustiedonkeruussa käytettävä tieteenalauskuitus
 - LIITO - Liiketoimintaontologia
 - MERO - Merenkulkualan ontologia
 - PUHO - Puolustushallinnon ontologia
 - TSR ontology
- GEOGRAPHY AND GEINFORMATION**
 - PTO - Finnish Geospatial Domain Ontology
- SCIENCE AND MEDICINE**
 - AFO - Natural resource and environment ontology
 - Kassu - Finnish Names of Plants
 - Medical Subject Headings
 - TERO - Finnish Ontology of Health and Welfare
- ART AND CULTURE**
 - Iconclass
 - KULO - Kulttuurien tutkimuksen ontologia

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
- 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

- | | |
|--------------------------|---------|
| organiset objektit | Finnish |
| <i>organinen rakenne</i> | |
| organiska objekt | Swedish |
| <i>organisk struktur</i> | |

URI

<http://www.yso.fi/onto/ys/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

1. ONKI Concept Search Widget in default state

ontology selector search field language selector open ONKI Browser

yso Open ONKI Ontology Browser

yso
yso
yso
paikka
toimo
maa
ic
hpmulti
mesh
genre
audience
medium
yso-old
ks-paikka
ks-toiminta
(all)

2. Autocompletion search result

yso ship en Open ONKI Ontology Browser

General Upper Finnish Ontology
ship finds
ship traffic
ship travel
shipbuilding
shipbuilding industry
shipmasters
shipping administration
shinoinen business

3. Concept collector for selected concepts

concept collector

yso:ship travel [change] [X] Open ONKI Ontology Browser

yso ship en Open ONKI Ontology Browser

Major components of an ontology infrastructure



- Ontologies
- Ontology Library Services

Actor Ontologies: Resolving Identities



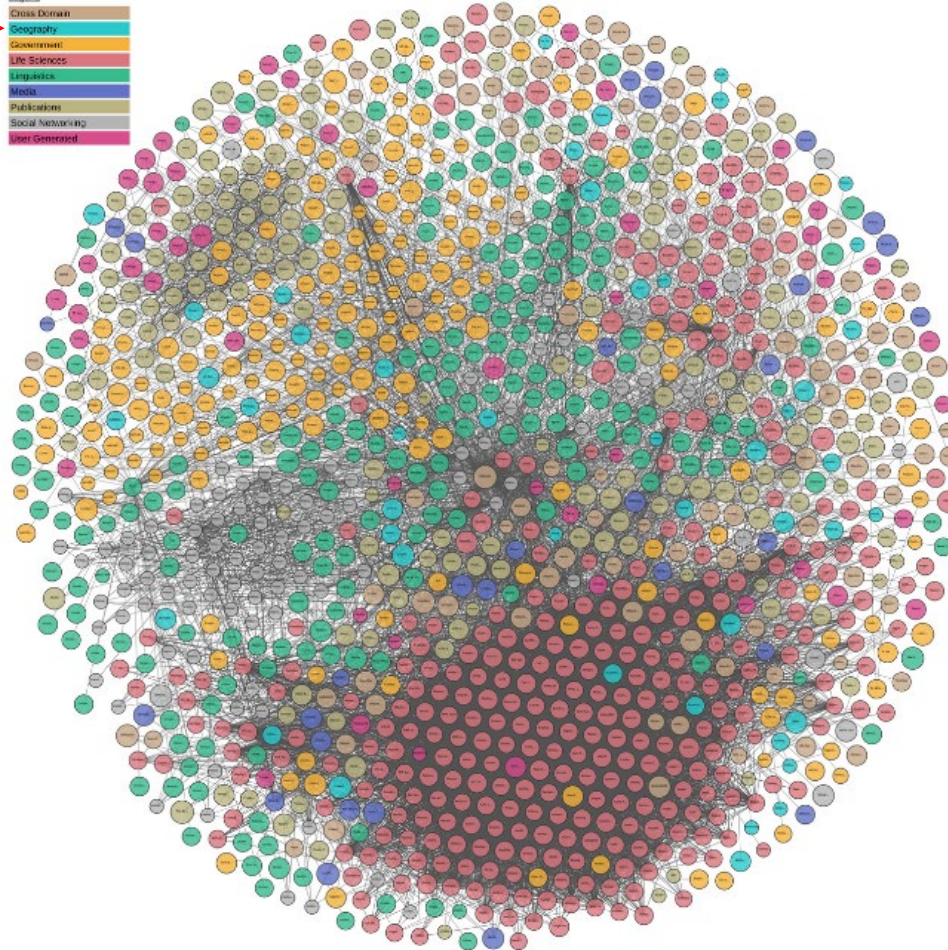
URI: http://dbpedia.org/resource/Pyotr_Ilyich_Tchaikovsky



Pjotr Tšaikovski (fi)
Пётр Ильич Чайко́вский (ru)
Pyotr Ilyich Tchaikovsky (en)
Pjotr Tjajkovskij (sv)
Pjotr Tsjajkovskij (no)
Pjotr Iljitsch Tschaikowski (de)
Piotr Ilitch Tchaïkovski (fr)
Piotr Ilich Chaikovski (es)
Pètr Il'ic 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>



The screenshot displays the Hipla.fi web application interface. At the top, the header includes the logo and the text "Finnish Ontology Service of Historical Places and Maps", along with navigation links for "About" and "Project home".

On the left side, there is a sidebar for selecting source datasets, with the following options:

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

Below the list are buttons for "+ Add a new place" and "View all places on current map view".

A search bar is present with the text "Search places" and "Maps" tabs. The search input field contains the text "musta".

The main map area shows a satellite view of Lappeenranta, Finland, with numerous blue location pins. A search results sidebar on the left lists items under the heading "Karelian map names (1922-44)":

- MUstasaari (Hypsographic feature, Johannes)
- Musta Riienlampi (Body of water, Uukuniemi)
- Musta-oja (Man-made feature)
- Mustajoensuu (Body of water, Kesälähti)
- Mustajoki (Body of water)
- Mustajoki (Body of water)
- Mustajoki (Village)
- Mustajoki (Body of water)
- Mustaiki (Body of water)

The map interface includes standard navigation controls like "Map" and "Satellite" tabs, a search icon, and a person icon for user profile. The bottom of the map shows "Map data ©2019 Google" and "Terms of Use".

NameSampo: <http://nimisampo.fi>



Nimisampo TIETOJA

Valitse lähdeaineistot

- Kotimaisten kielten keskuksen Nimiarkisto (NA) i
- Maanmittauslaitoksen paikannimirekisteri (PNR) i
- Maanmittauslaitoksen Karjalan karttanimet (KK) i
- The Getty Thesaurus of Geographic Names (TGN) i

Hae paikannimellä
mustalampi

1508 hakutulosta

Nimi

Hae...
 Mustalampi (1508)

TAULUKKO KLUSTEROITU KARTTA KARTTA **LÄMPÖKARTTA** TILASTOT LATAUS

Map Satellite

Sweden
Norway
Finland
Estonia

Oslo
Stockholm
Tampere
Helsinki
Tallinn

Map data ©2019 GeoBasis-DE/BKG (©2009), Google Terms of Use

A Aalto University School of Science

UNIVERSITY OF HELSINKI

HELDIG Helsinki Centre for Digital Humanities

SeCo

Institute for the Languages of Finland KOTIMAISTEN KIELTEN KESKUS

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

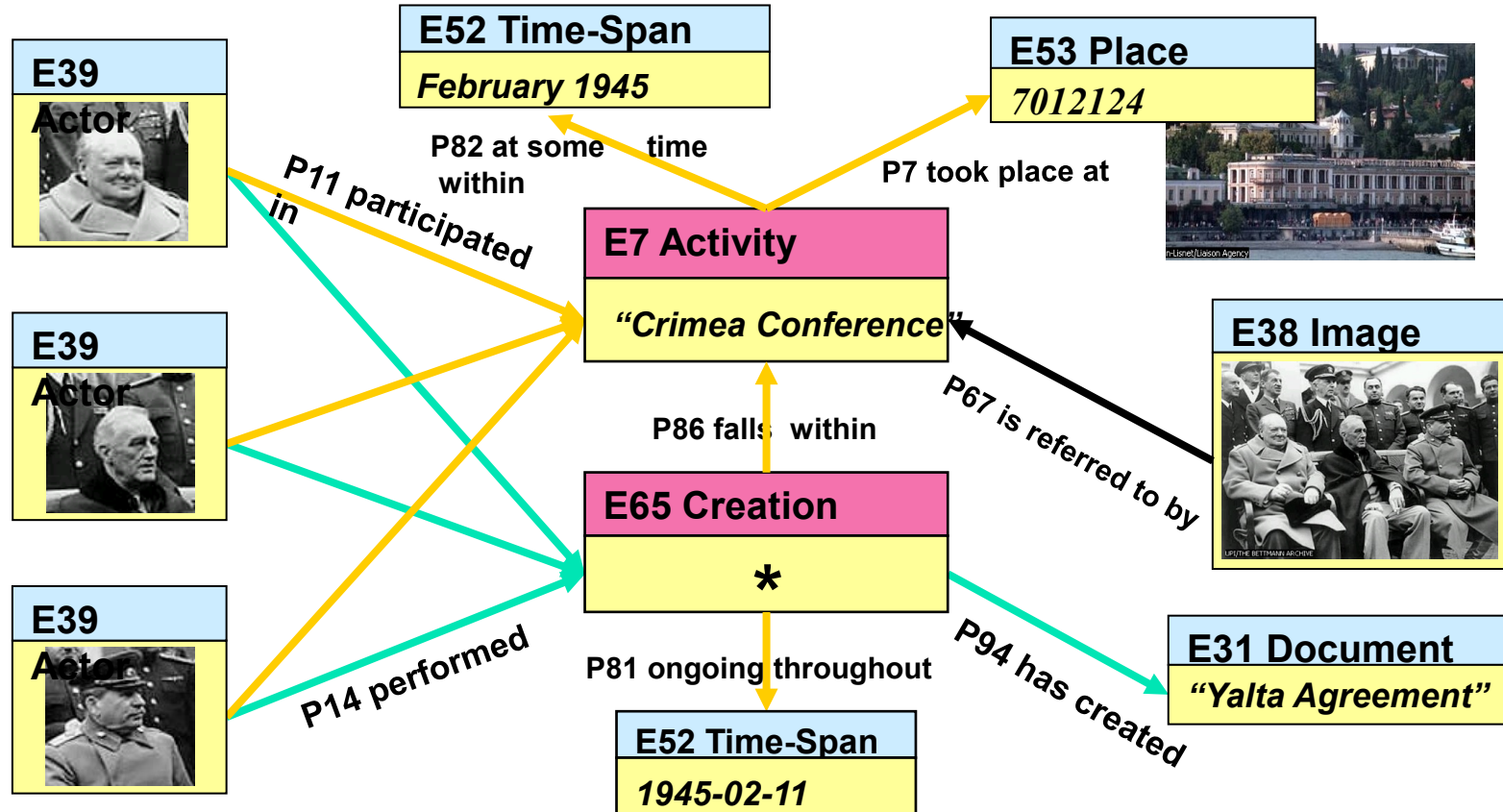
- Mapping/refining schemas using subproperties
- “Dumb down principle” is used
- <https://dublincore.org/>

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]



Shared Linked Open Data & Services



An example of a Linked Data Service



MENU

CORONA MEASURES

DNB FOR USERS

DNB PROFESSIONAL



Deutsch

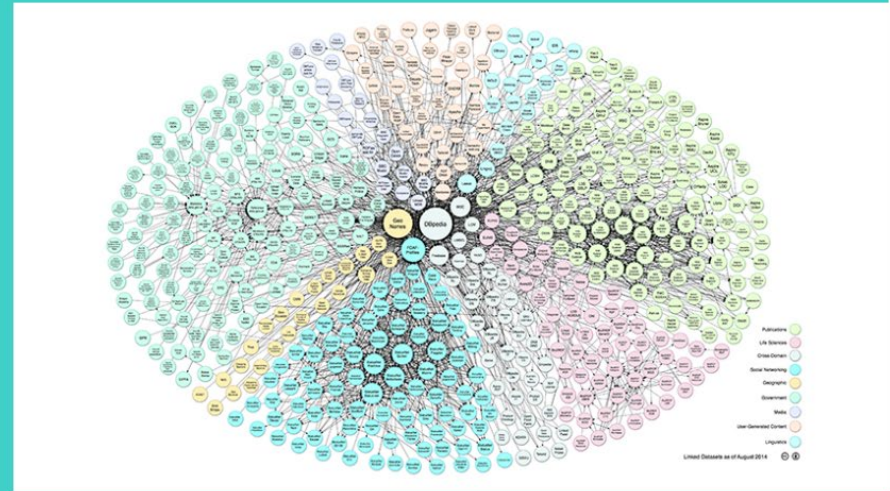
Sign language

Simple language

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

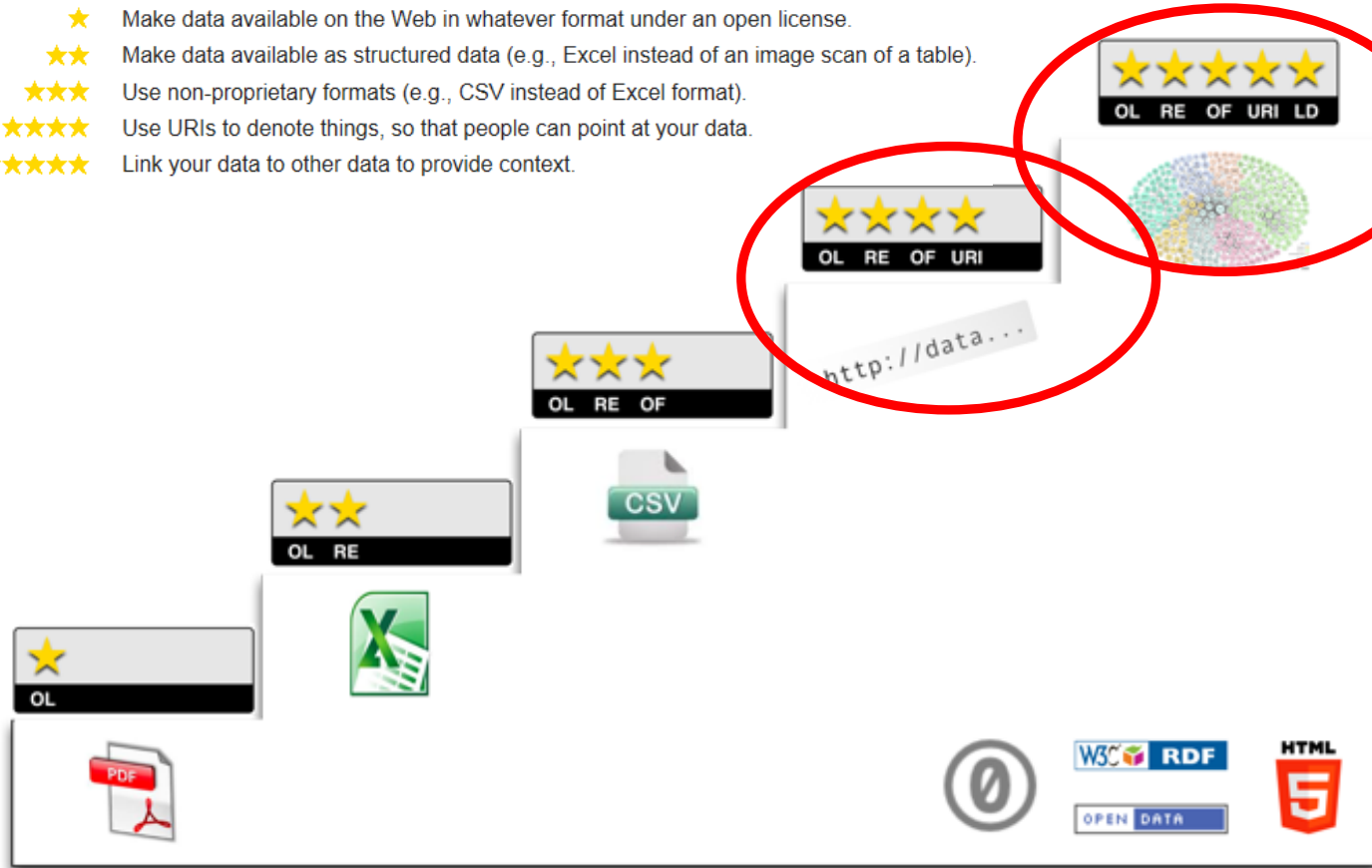


How to publish Linked Data?

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>

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

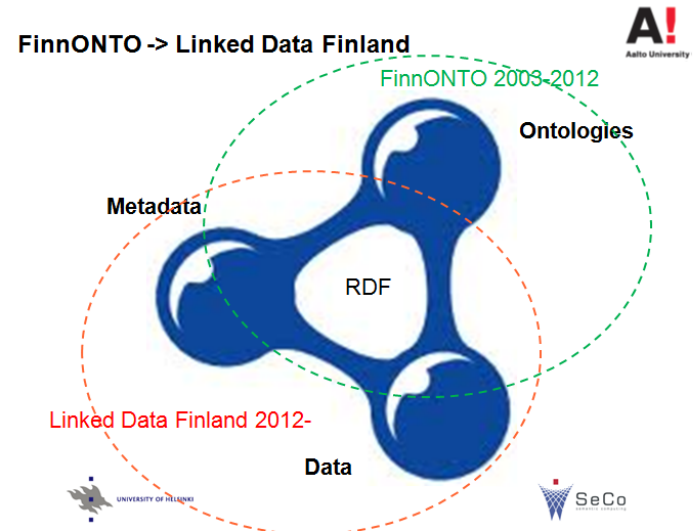
- Linked datasets
- Schemas

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:

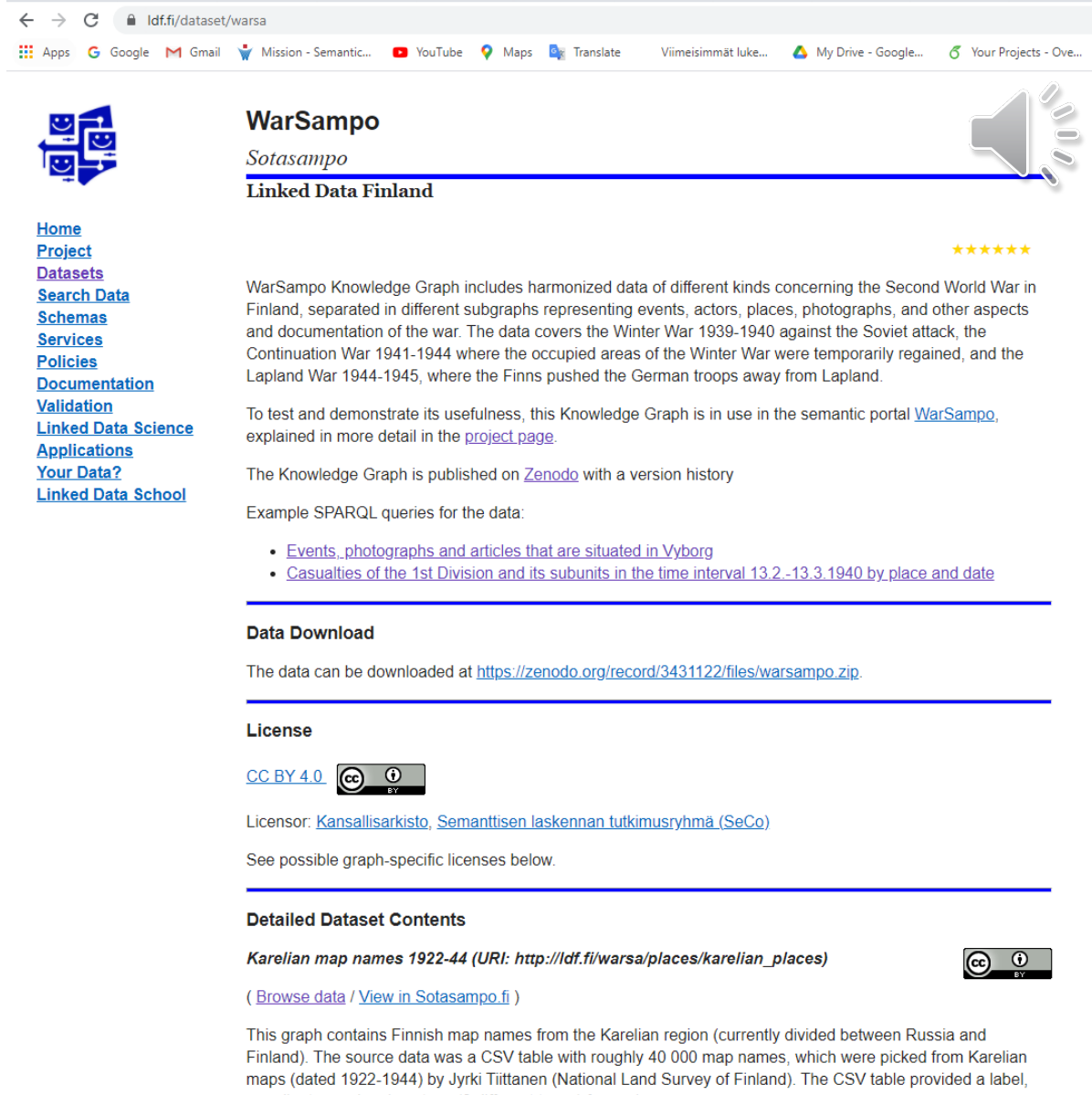
- ★★★★★★ Provide your data with a schema and documentation so that people can *understand and re-use* your data easily.
- ★★★★★★★ Validate your data and denote its provenance so that people can *trust the quality* of your data.

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>



The screenshot shows the WarSampo dataset page on the ldf.fi website. The browser's address bar displays 'ldf.fi/dataset/warsa'. The page features a navigation menu on the left with links to Home, Project, Datasets, Search Data, Schemas, Services, Policies, Documentation, Validation, Linked Data Science, Applications, Your Data?, and Linked Data School. The main content area is titled 'WarSampo' and 'Sotasampo', with a 'Linked Data Finland' badge and a five-star rating. The text describes the WarSampo Knowledge Graph, which includes harmonized data from the Second World War in Finland, covering the Winter War (1939-1940), the Continuation War (1941-1944), and the Lapland War (1944-1945). It mentions that the graph is used in the semantic portal WarSampo and is published on Zenodo. Example SPARQL queries are provided, such as 'Events, photographs and articles that are situated in Vyborg' and 'Casualties of the 1st Division and its subunits in the time interval 13.2.-13.3.1940 by place and date'. The 'Data Download' section provides a link to the data on Zenodo. The 'License' section shows the CC BY 4.0 license and identifies the licensor as Kansallisarkisto, Semanttisen laskennan tutkimusryhmä (SeCo). The 'Detailed Dataset Contents' section highlights the 'Karelian map names 1922-44' dataset, providing a URI and a link to browse the data. A Creative Commons BY license icon is also present at the bottom right of the page.

ldf.fi/dataset/warsa

Apps Google Gmail Mission - Semantic... YouTube Maps Translate Viimeisimmät luke... My Drive - Google... Your Projects - Ove...



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](#) 

Licensor: [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,



Services



- Customary 5-star Linked Data Services
 - *Viewing and browsing RDF*
 - *SPARQL endpoint services (using Fuseki)*
- Documentation
- Validation
- Visualization
- Data curation
 - *Automatic annotation, RDF editing, data linking*
- Sharing policies
 - *URI minting*
 - *Licensing*
- Your data?
 - *Open service for publishing useful Linked Data*



Software Tools for the Semantic Web



Aalto University
School of Science

Department of
Computer Science





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/>



Case: Finnish Linked Open Data Infrastructure for Digital Humanities

Case on Video: <https://vimeo.com/460086143>



**Building a National Level
Linked Open Data Infrastructure for
Digital Humanities in Finland**

Prof. Eero Hyvönen, Director
Helsinki Centre for Digital Humanities (HELDIG)
University of Helsinki and Aalto University
Semantic Computing Research Group (SeCo)
<http://seco.cs.aalto.fi/u/eahyvone/>

00:00

vimeo

Summary



Semantic Web infrastructures are needed

- for data interoperability
- for reusing data, schemas, ontologies, and software

Infrastructures include

- Ontologies & ontology services
- Shared metadata models
- Linked data services
- Shared software and tools

In Finland a national level solution is being developed