

# Linked Open Data Infrastructure for Digital Humanities in Finland

## LODI4DH

**Eero Hyvönen and Jouni Tuominen**

Aalto University and University of Helsinki  
HELDIG – Helsinki Centre for Digital Humanities

<http://heldig.fi>

Semantic Computing Research Group (SeCo)

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

# Contents

- **Linked Data and Semantic Web**
- **The Vision of LODI4DH**
- **LODI4DH Initiative Outline**

# Linked Data and Semantic web

# Semantic Web Activity at W3C Started 2001

# SCIENTIFIC AMERICAN™

SEARCH  

- [Log In or Register](#)
- [Log In to SA Digital](#)

[Energy & Sustainability](#) ▾ [Evolution](#) ▾ [Health](#) ▾ [Mind & Brain](#) ▾ [Space](#) ▾ [Technology](#) ▾ [More Science](#)

[Home](#) » [Scientific American Magazine](#) » [May 2001](#)

[Feature Articles](#) |



## The Semantic Web

A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities

By [Tim Berners-Lee](#), [James Hendler](#) and [Ora Lassila](#) | [May 17, 2001](#) | [10](#)

[Share](#) [Email](#) [Print](#)



**Happy 20th Birthday, World Wide Web**

**CERN on March 13 celebrates the 20th anniversary of a proposal entitled, "Information Management: A Proposal," by Tim Berners-Lee, which would become the blueprint for the**

**World Wide Web »**

March 12, 2009



UNIVERSITY OF HELSINKI



# 2001

## Department of Computer Science

### Department information

Homepage

News and events

Research

Studies

Admission

Computing facilities

Administration

Quality manual

Contact information



Helsingin yliopisto - Tietojenkäsittelytieteen laitos

[Koti](#) [Yhteystiedot](#) [Laitos lyhyesti](#) [Henkilöt](#) [Palvelut](#) [Opiskelu](#) [Tutkimus](#)  
[Uutiset ja tapahtumat](#)

## SEMANTIC WEB KICK-OFF IN FINLAND - ÄLYKÄS WWW SUOMESSA



Helsingin yliopisto, Porthania, sali P3

Perjantai 2.11.2001, klo 9:00-16:30



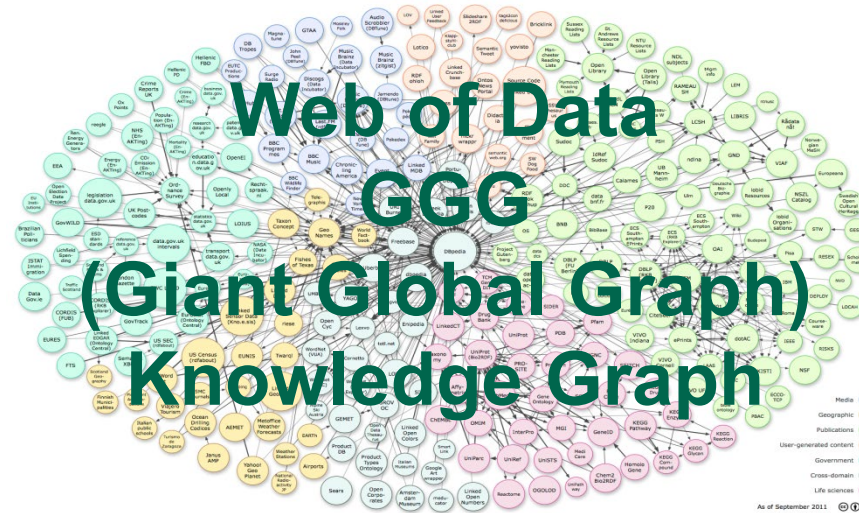
UNIVERSITY OF HELSINKI



SeCo  
semantic computing



## Web of Pages WWW



# WWW Document: Semantic portal HealthFinland

http://www.ttl.fi/NR/rdonlyres/142A177B-FE02-4A77-A434-3049C37EBC61/0/maijan\_lasit\_tulosta.pdf - Windows Internet Explorer

http://www.ttl.fi/NR/rdonlyres/142A177B-FE02-4A77-A434-3049C37EBC61/0/maijan\_lasit\_tulosta.pdf

Save a Copy Hand Tool Select Snapshot Tool Zoom In 84% Rotate Clockwise Check Spelling Undo Copy Search Web Yahoo! Download New Reader Now

Pages Attachments Comments

TTL - Maijan uudet näyttöpäätelasit <http://www.ttl.fi/ergonomia>

## Maijan uudet näyttöpäätelasit

Sopivien silmälasien hankinta tietotyöhön voi useinkin olla pulmallista. Seuraavassa näet, miten ikänäköinen Maija päätyi silmälasiratkaisuunsa

---

Maija sai lähinään huonontuessa aluksi lukulasit. Hän hankki kymmenen vuoden aikana kolmet lukulasit: uudet aina hiukan entistä vahvemmat.

Päätteitä lukeminen alkoi käydä ongelmalliseksi. Maija ei osannut kirjoittaa sokkona joten näppäimistölle näkeminen oli myös välttämätöntä. Niska ja hartiat jomottivat, silmiä kirveli ja vasemmassa silmässä oli elohiiri viikkokausia.

Lukulasit käytössä.  
Maija joutuu istumaan epämiellyttävän lähellä päätettä. Kynärpäiden voimakas koukistaminen sitoo käsiä ja vaikeuttaa näppäilyä.

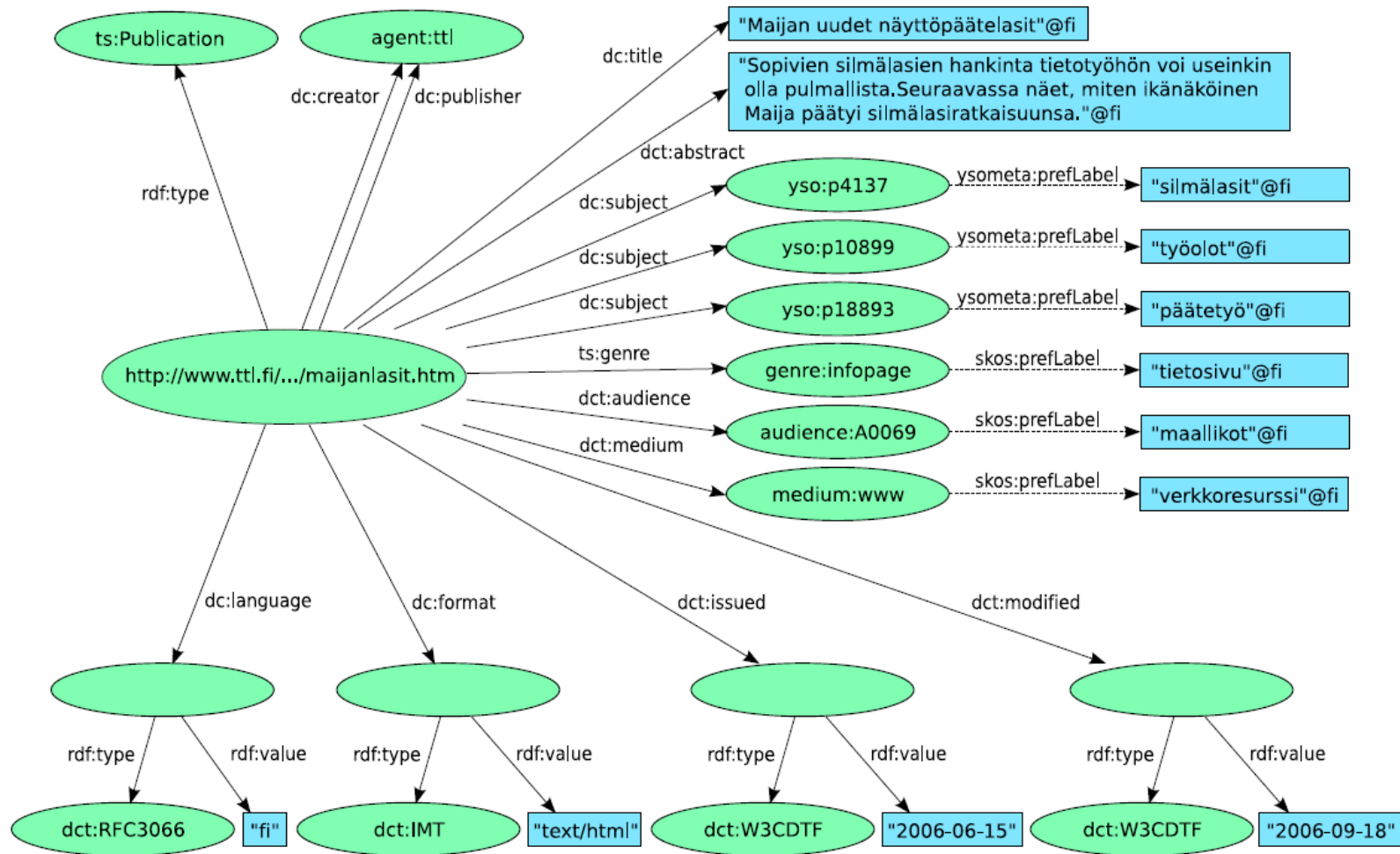


Lukulasit käytössä.  
Maija joutuu taivuttamaan niskaa voimakkaasti eteen nähäkseen näppäimet. Päätteen ja näppäimistön vuorottainen katselu aiheuttaa jatkuvaa niskan liikettä.



Hide Toolbars 1 of 4 Unknowm Zone

# Corresponding GGG graph (in RDF)







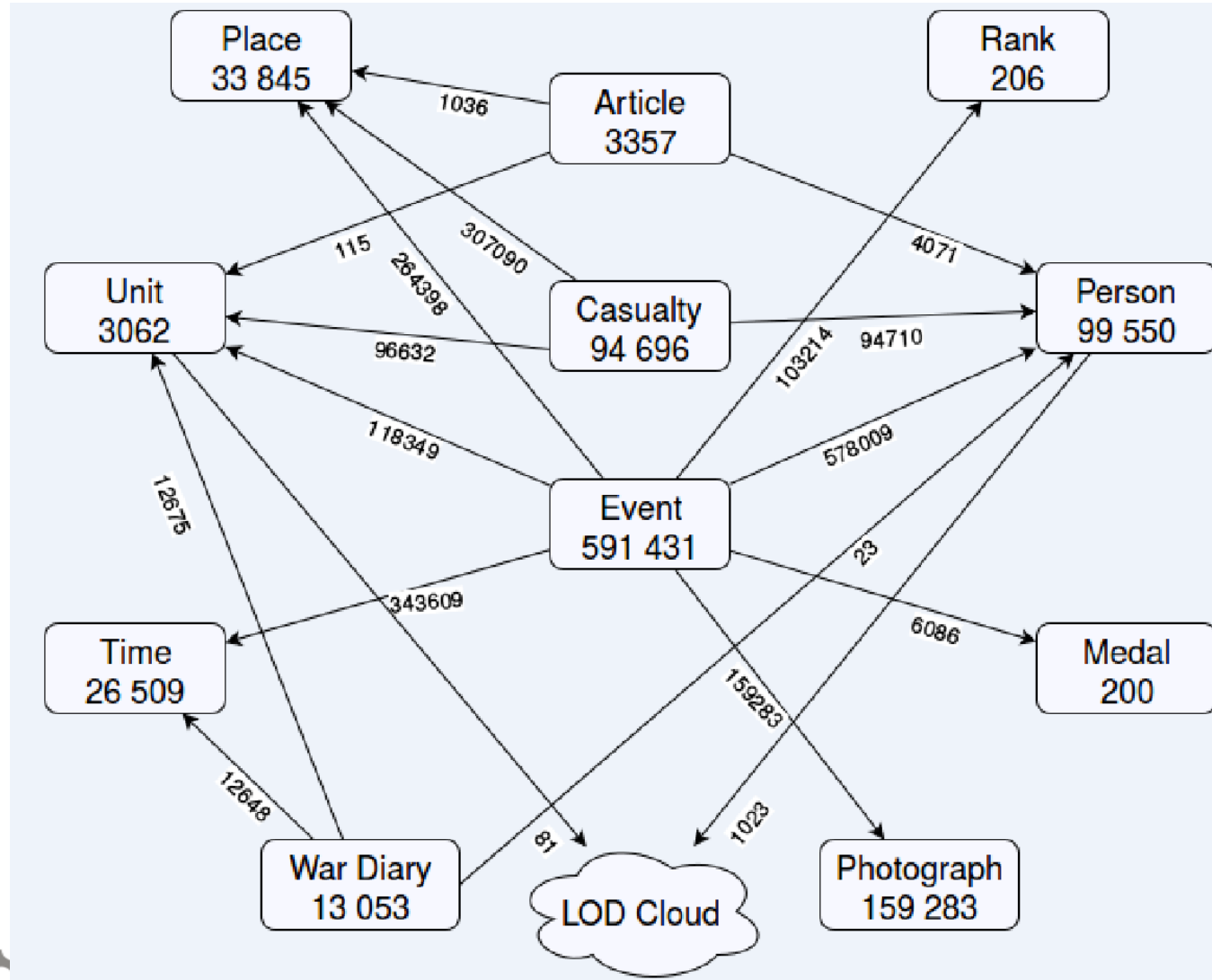
Example Dataset, Data Service, and Application:

**WarSampo** – Finnish WW2 on the Semantic Web

<http://sotasampo.fi>

(Hyvönen et al., ESWC 2016)

# WarSampo Linked Open Data Cloud



12M triples  
knowledge graph

[Leskinen et al.]

# Semantic Portal <http://sotasampo.fi/en> 8 Perspectives to War History



Perspectives - Suomeksi

Feedback Info - Instructions

**WarSampo**  
Finnish World War II on the Semantic Web

**In-use since Nov 2015**  
**130 000 users in 2017**



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

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

**2. Persons**  
Data about persons with related links from various sources

**3. Army Units**  
Events and other related data about army units visualized using a timeline

**4. Places**  
Search and browse places and maps covering the war zone area. Includes a timeline and other additional data such as events and photographs linked to places

**5. Deaths**  
Faceted semantic search and contextual reader for Kanssa taisteli magazine articles containing mostly names of soldiers related to WWII

**6. Memoirs**  
A table-like view of war casualty records that can be filtered using a tree view. The data is enriched with links to other WarSampo datasets

**7. Photos**  
Browse the content of the Finnish Medtime Photograph Archive with faceted search

**8. Cemeteries**  
War cemeteries of Finland with photographs

<https://vimeo.com/212249404>

Example Pages

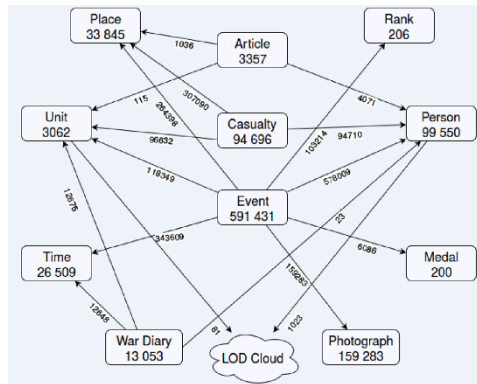




# Linked Open Data Cloud 2018

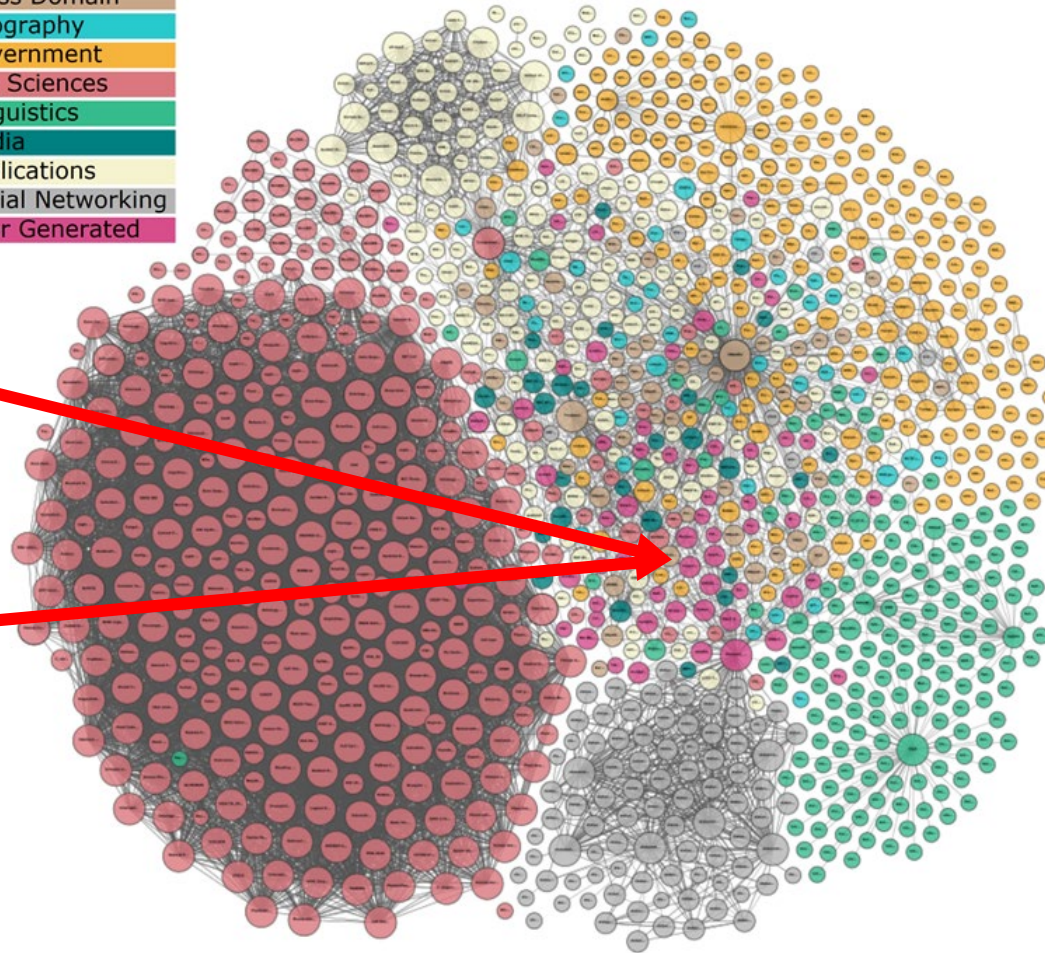
<http://linkeddata.org>

## WarSampo



### Legend

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



**LODStats 2018 (<http://stats.lod2.eu/>):  
10000 datasets, 150 billion triples**

# Knowledge Graphs: Big Boys Have Entered the Game

<http://schema.org>

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

The Google logo, featuring the word "Google" in its signature multi-colored font (blue, red, yellow, blue, green, red).The Microsoft logo, consisting of the four-pane Windows logo (red, green, blue, yellow) followed by the word "Microsoft" in a grey sans-serif font.The IBM logo, featuring the word "IBM" in a blue, horizontally-striped font.The eBay logo, featuring the word "eBay" in a multi-colored font (red, blue, yellow, green).The Facebook logo, consisting of a dark blue square with the word "facebook" in white lowercase letters.

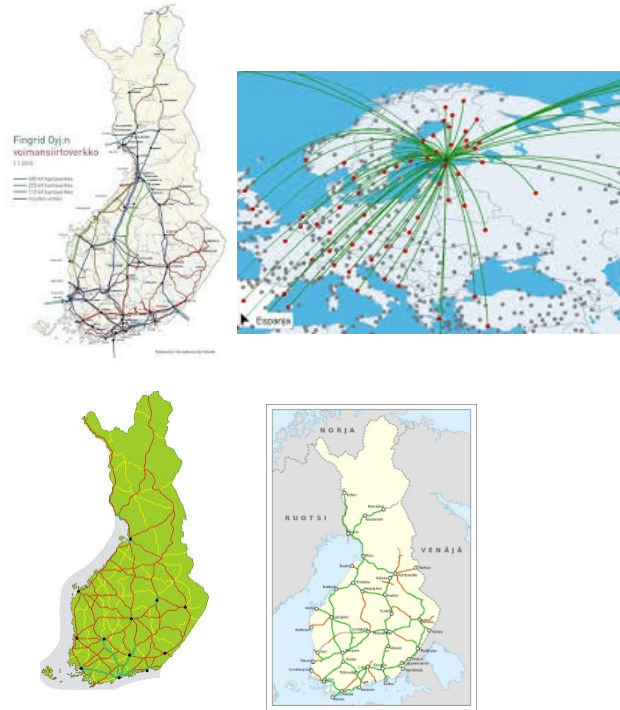


# The Vision of LODI4DH

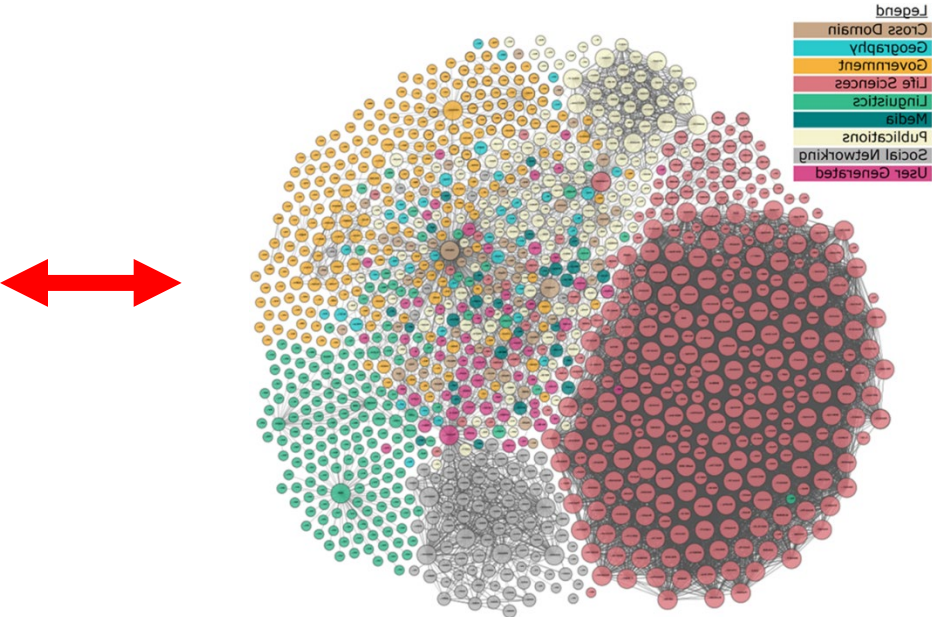
## Linked Open Data Infrastructure for Digital Humanities in Finland

# Linked Data Infrastructure

## Traditional Infras



## Linked Data Infra



# **Building Semantic Web Infrastructure in Finland with Applications (2003-)**

- **National Finnish Ontology Project 2003-2012**
- **Linked Data Finland 2012-2014**
- **Linked Open Science 2014-2017**
- **LODI4DH 2018-**

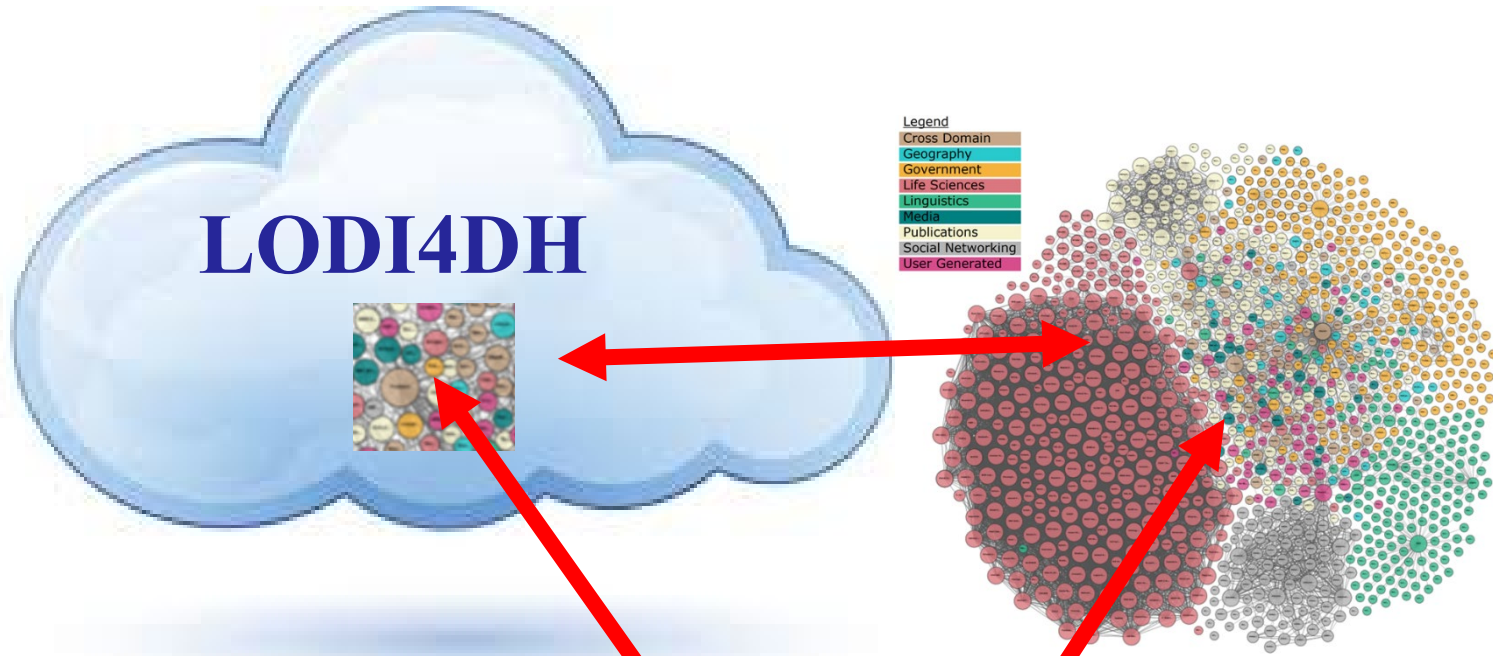




Word Wide Web Consortium

# Linked Open Data

<http://linkddata.org>



**DARIAH-EU** Digital Research Infrastructure for the Arts and Humanities

Eero Hyvönen

## Semanttinen web

Linkitetyn  
avoimen  
datan  
käsikirja



Gaudeamus



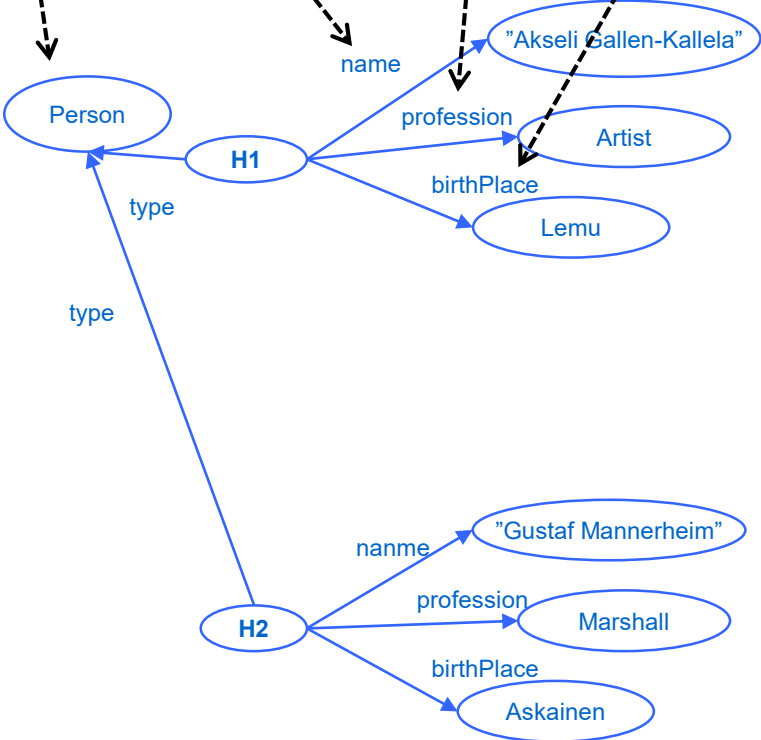
## How Does Linked Data Work in Practise?





# Biographical Registries Collect Data about Persons

henkilö	nimi	ammatti	syntymapaikka	...
H1	Akseli Gallen-Kallela	taiteilija	Lemu	
H2	Gustaf Mannerheim	marsalkka	Askainen	
...				

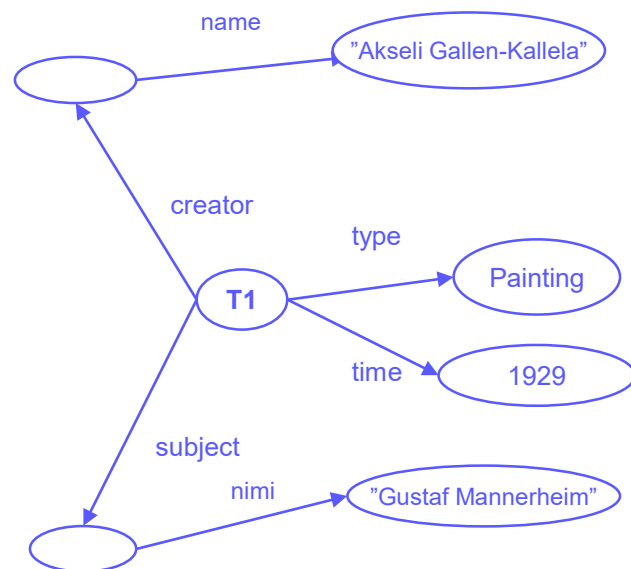


## Biography Center



# Art Museum Catalogs Paintings

teos	nimi	tekijä	aika	aihe	...
T1	Mannerheimin muotokuva	Akseli Gallen-Kallela	1929	Gustaf Mannerheim	
T2	Aino-triptyykki	Akseli Gallen-Kallela	1891	Aino, Kalevala	
...					

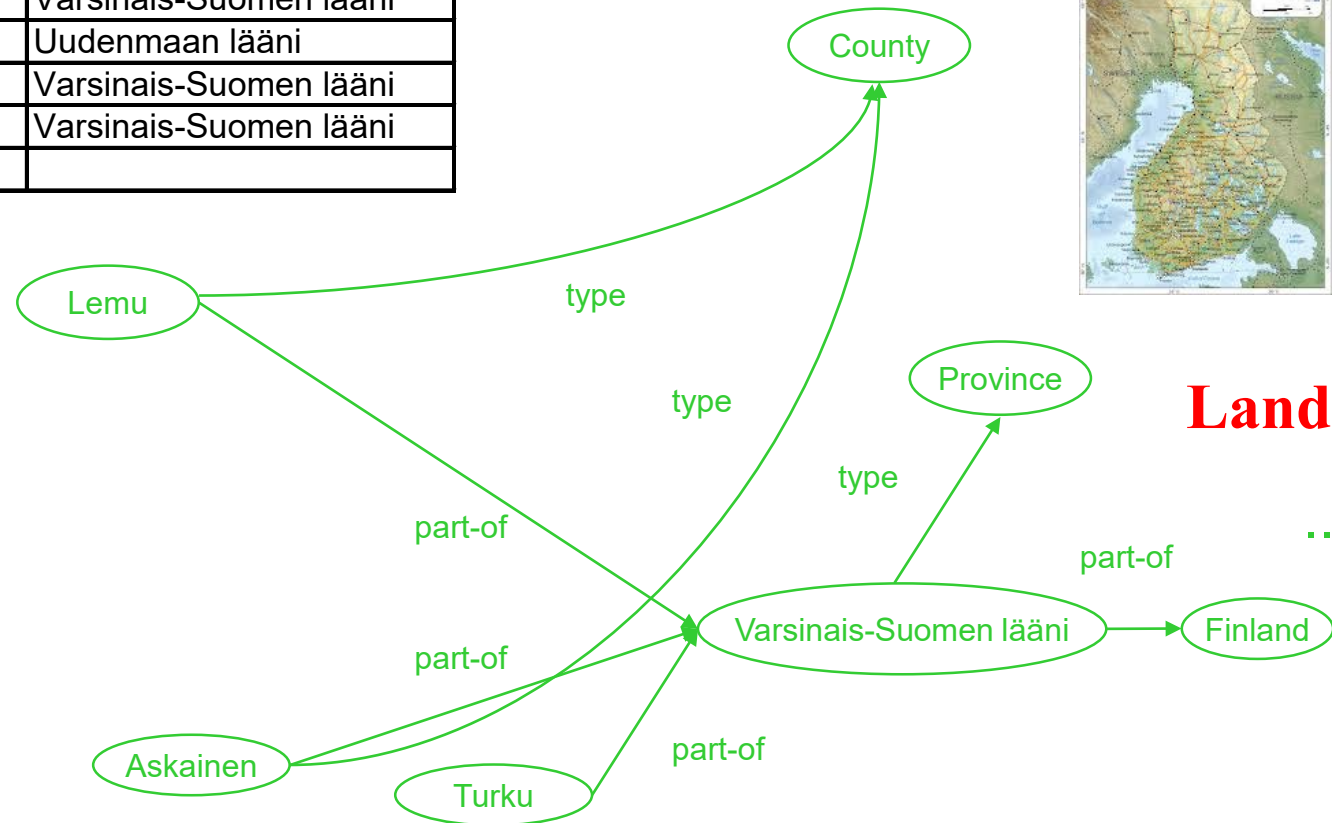


## Art Museum Collection



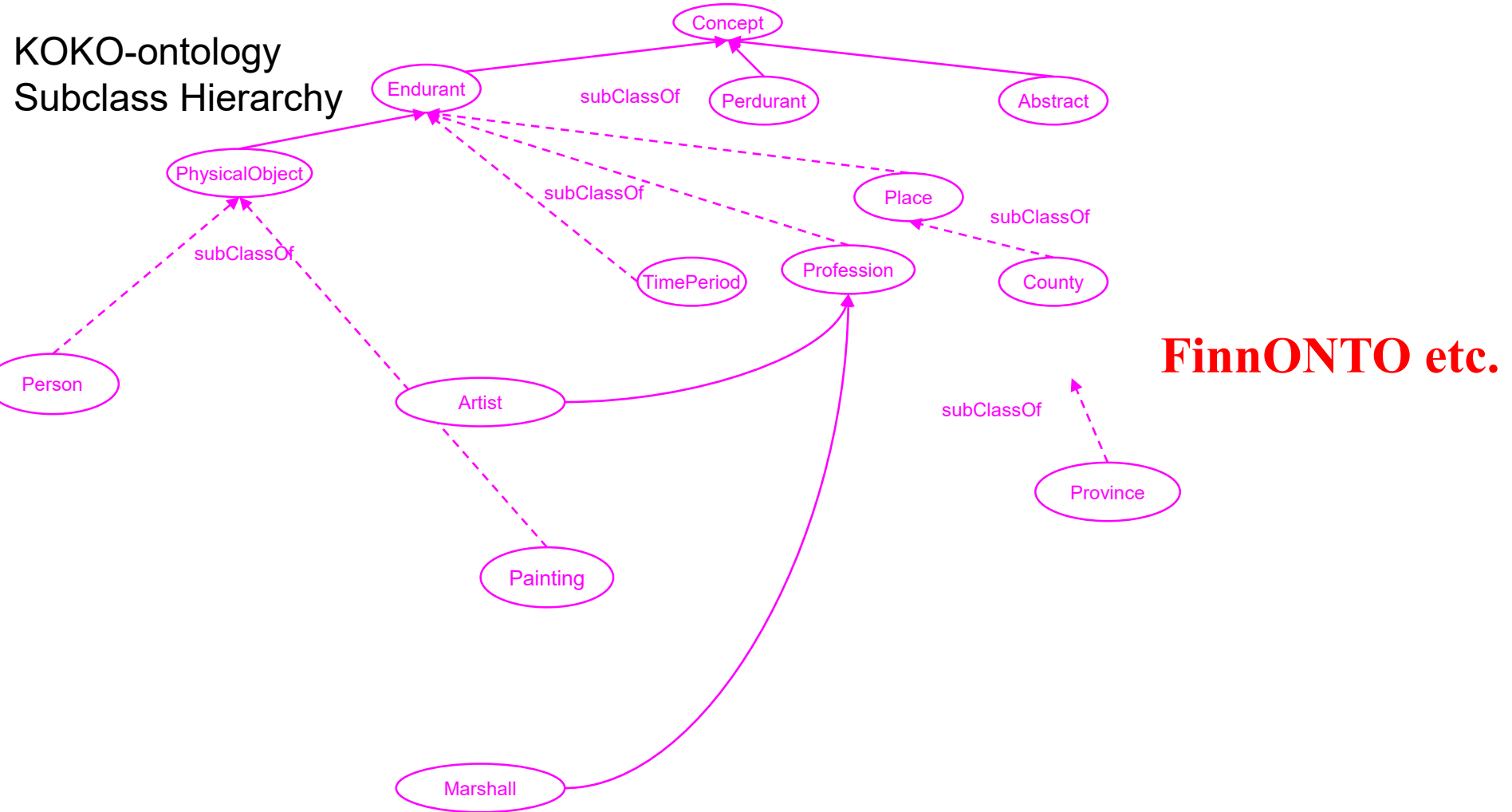
# Land Survey Organizations Know Places

kunta	lääni
Askainen	Varsinais-Suomen lääni
Helsinki	Uudenmaan lääni
Lemu	Varsinais-Suomen lääni
Turku	Varsinais-Suomen lääni
...	

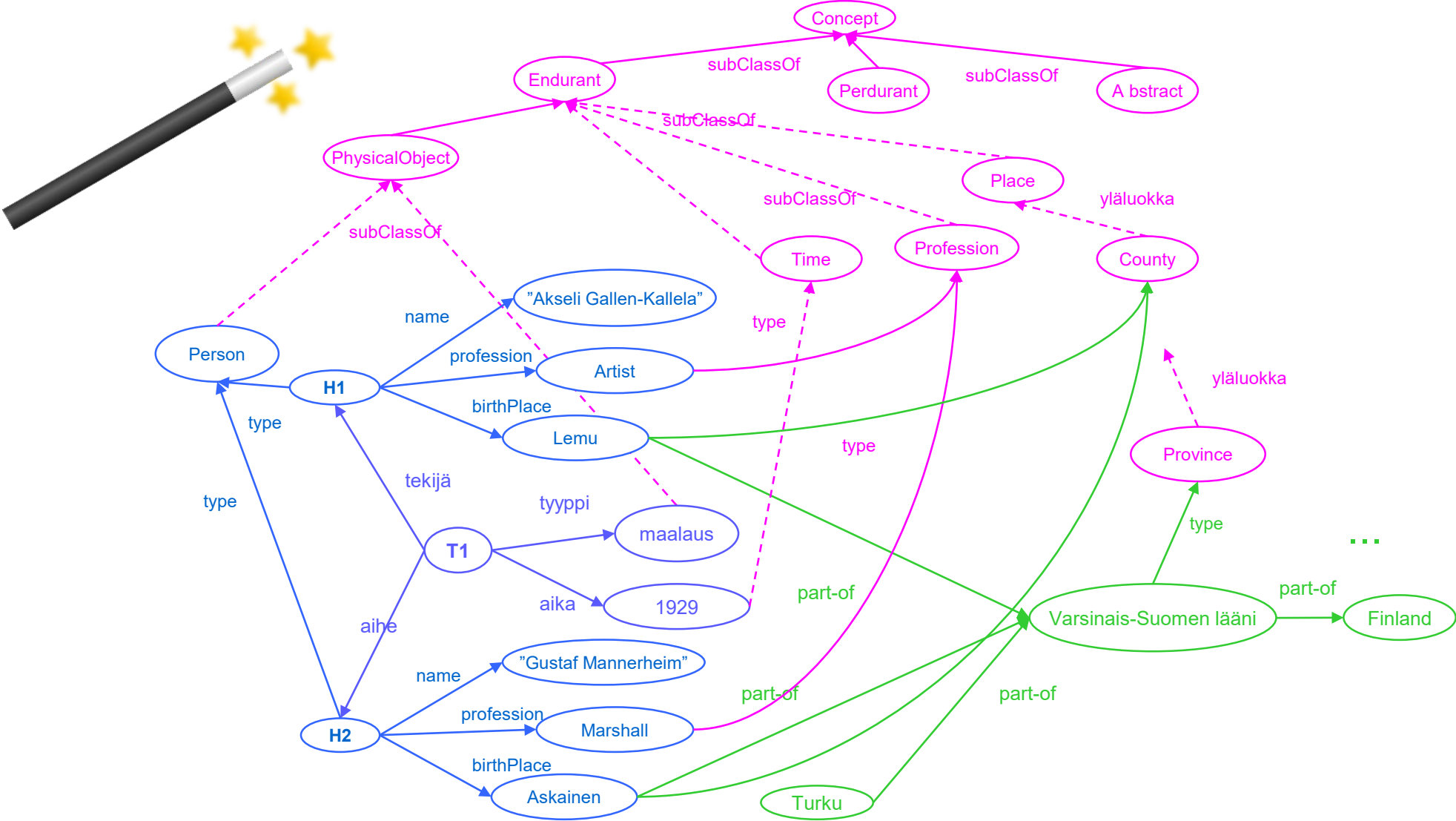


**Land Survey**

# Ontologies Developed by Semantic Web Researchers



# RDF Connects and Harmonizes Linked Data into a GGG





1 + 1 > 2

# AI

In Principle a Piece of Cake but ...

How to **align concepts** (URIs) used by different organizations?

How to **align metadata** models used by different organization?

**SHARED INFRA NEEDED!**



# Major Components Needed



# Key Components of a Linked Data Intrastructure

**1. Domain Ontologies**

**2. Metadata Models**

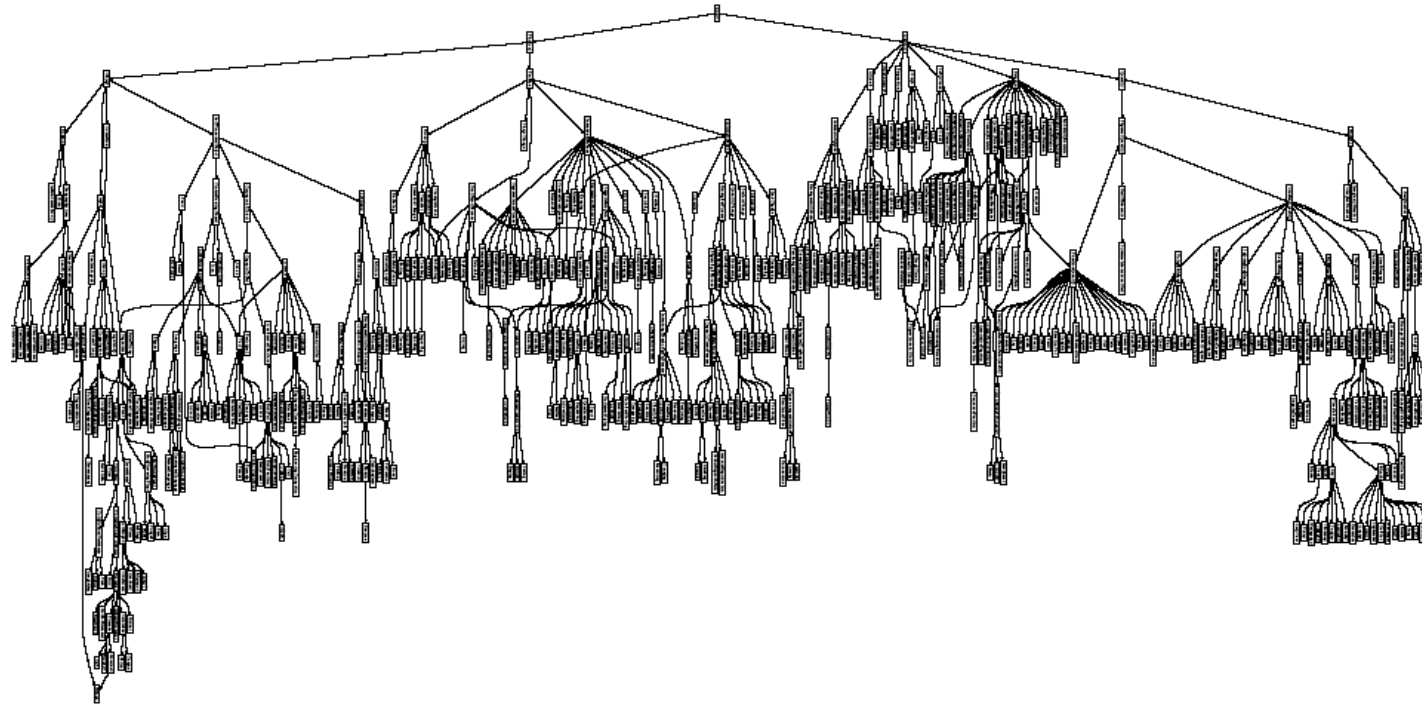


**RDF**

**3. Datasets**

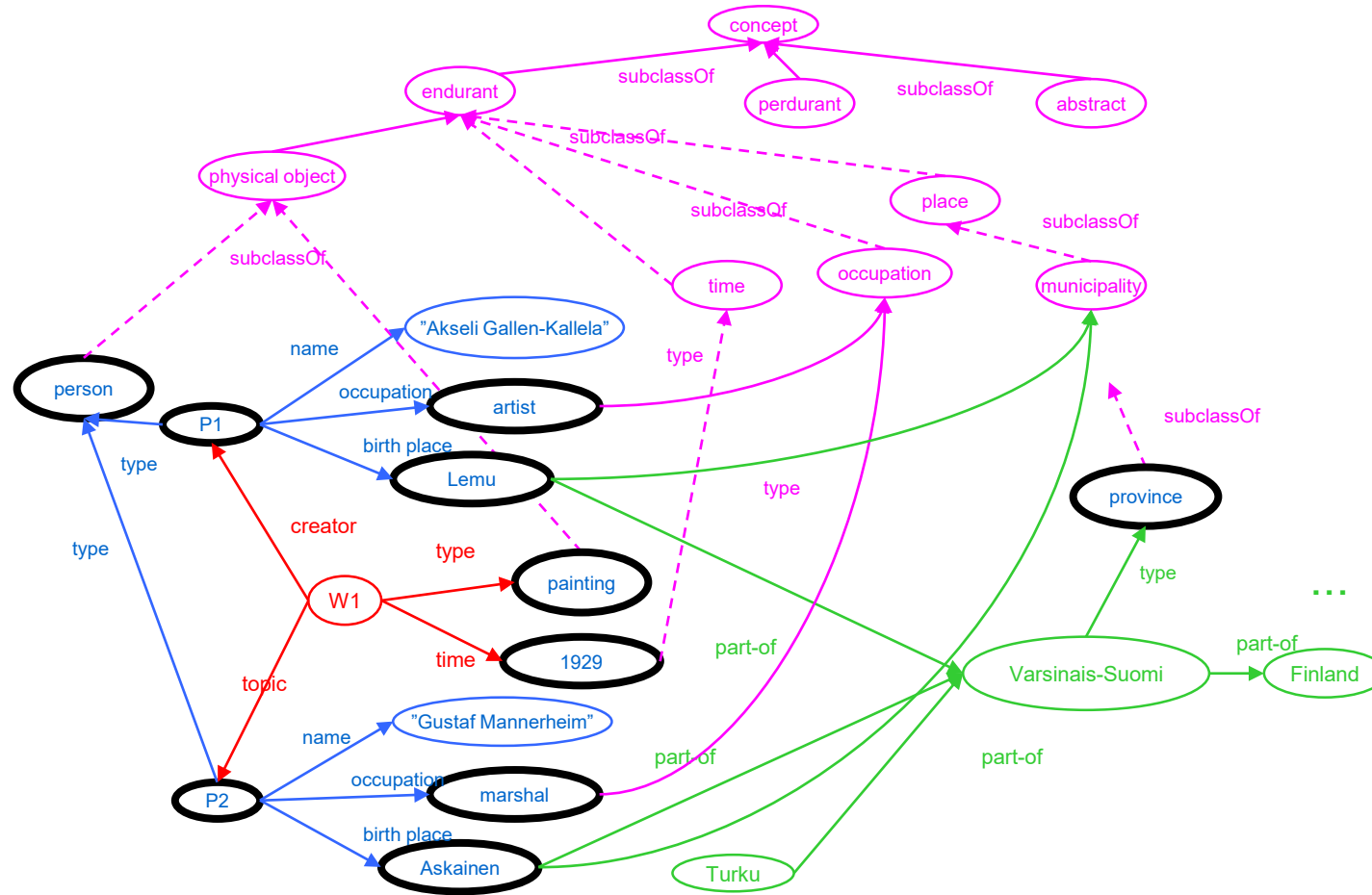


## Component 1: Domain Ontologies



# Concepts (IDs) and Data Model Glue Data Together

## Solution: Shared Ontology Infrastructure



# Ontology of Historical Persons

# Resolving Identities



URI: [http://dbpedia.org/resource/Pyotr\\_Ilyich\\_Tchaikovsky](http://dbpedia.org/resource/Pyotr_Ilyich_Tchaikovsky)



Pjotr Tšaikovski (fi)  
Пётр Ильич Чайковский (ru)  
Pyotr Ilyich Tchaikovsky (en)  
Pjotr Tjajkovskij (sv)  
Pjotr Tsjajkovskij (no)  
Pjotr Iljitsch Tschaikowski (de)  
Piotr Ilitch Tchaïkovski (fr)  
Piotr Ilich Chaikovski (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)

# Ontologies of Historical Places and Maps



# Finnish Ontology Service of Historical Places and Maps

<http://hipla.fi>

The screenshot displays the Hipla.fi web application interface. At the top, the logo "hipla.fi" is on the left, and the title "Finnish Ontology Service of Historical Places and Maps" is in the center. On the right, there are links for "About" and "Project home".

The left sidebar contains a "Select source dataset(s)" section with the following options:

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

Below this list are two buttons: "+ Add a new place" and "View all places on current map view".

A search bar is present with the text "Viipuj" entered. Below the search bar, there are two tabs: "Search places" and "Maps".

The main content area shows a map of the Lappeenranta region. The map is overlaid with a red outline representing a specific area of interest. Several place names are visible on the map, including "Lappeenranta", "Imatra", "Svetogorsk", "Lesogorsky", "Kamennogorsk", "Krasnyy Sokol", "Ozyorskoye", "Vasilyevo", "Romashki", "Barzhevo", "Losevo", "Michurinskoye", "Korobitsyno", "Volochevka", "Pervomayskoye", "Roshchino", "Tarasovskoye", "Polyanu", "Primorsk", "Pionerskoye", "Kamenka", "Glebychevo", "Vyazy", "Baltiysk", "Luzhanka", "Vainikkala", "Raippo", "Anttila", "Ylänmaa", "Hujakkala", "Muirikkala", "Tortyanovka", "Bolsheoy Bor", "Kuzhechnoye", "Sevastyanovo", "Yarkoye", "Priozersk", "Motornoye", "Otradnoye", "Prikladnoye", "Sosnovo", "Orekhovo", "Vaskelovo", "Korobitsyno", "Volochevka", "Pervomayskoye", "Roshchino", "Tarasovskoye", "Polyanu", "Primorsk", "Pionerskoye", "Kamenka", "Glebychevo", "Vyazy", "Baltiysk", "Luzhanka", "Vainikkala", "Raippo", "Anttila", "Ylänmaa", "Hujakkala", "Muirikkala", "Tortyanovka", "Bolsheoy Bor".

An inset map in the bottom right corner shows the location of the main map area within Finland and Estonia. The bottom of the page includes the Google logo and copyright information: "Map data ©2016 Google Terms of Use Report a map error".

# Ontology of Historical Events

”Semantic glue” of everything

## Example of an Event on HISTO Ontology of Finnish History

SAHA3 | historia - search

<http://agricola.utu.fi/rdf/phe863>

### historiallinen tapahtuma: YYA-sopimus raukesi

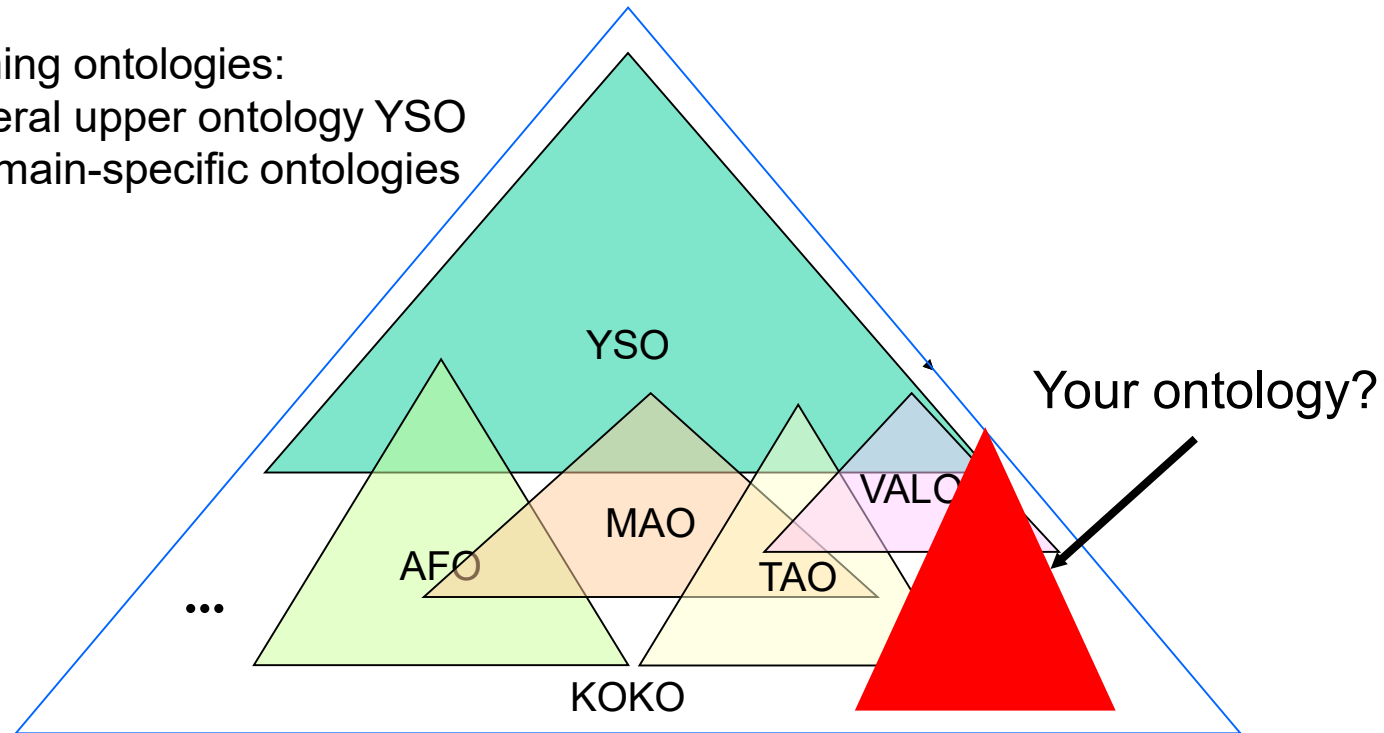
[\[edit\]](#)

nimi	(fi) YYA-sopimus raukesi
tapahtumatyyppi	<a href="#">kahdenkeskiset kansainväliset sopimukset</a>
kuuluu teemaan	<a href="#">poliittinen historia</a>
kuka	<a href="#">Aho, Esko</a> , <a href="#">Burbulis, Gennadi Eduardovitš (1945-)</a> , <a href="#">Suomi (1917-)</a> , <a href="#">Venäjä (1990-)</a>
toiminta	<a href="#">allekirjoittaminen</a> , <a href="#">irtisanominen</a> , <a href="#">poliittinen päätöksenteko</a>
tulos	<a href="#">Suomen tasavallan ja Venäjän federaation välinen sopimus suhteiden perusteista 63/1992</a>
tapahtuma-aika	<a href="#">1992-01-20</a>
tapahtumapaikka	<a href="#">Helsinki(1966-2008)</a>
asiasana	<a href="#">diplomatia</a> , <a href="#">kansainväliset sopimukset</a> , <a href="#">kansainväliset suhteet</a> , <a href="#">ulkopolitiikka</a> , <a href="#">YYA-sopimus</a>
liittyy	<a href="#">YYA-sopimus solmittiin Neuvostoliiton kanssa</a> , <a href="#">YYA-sopimusta jatkettiin kahdellakymmenellä vuodella</a>
liittyvä web-sivu	(fi) <a href="http://www.finlex.fi/fi/sopimukset/sopsteksti/1992/19920063">http://www.finlex.fi/fi/sopimukset/sopsteksti/1992/19920063</a>
syy	<a href="#">Neuvostoliitto lakkautetaan</a>
tyyppi	<a href="#">historiallinen tapahtuma</a>

# Ontologies of Historical Keyword Concepts

# Holistic Collaborative Finnish Ontology KOKO

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



[Hyvönen et al., ESWC 2009]

<http://finto.fi>





## Ontology of Historical Times and Periods



## Component 2: Metadata Models

### **Object and document centric**

- Traditional approach in museums
  - Dublin Core

### **Event-centric**

- Object = sum of the events related to it
- New standards for data harmonization and enriching
  - CIDOC CRM (museums) + IFLA LRM (libraries)

### **Process-centric?**

- Intangible cultural heritage

# Example

## Document/Object-centric model

**Type:** Image  
**Title:** Allied Leaders at Yalta  
**Date:** 1945  
**Publisher:** United Press International (UPI)  
**Source:** The Bettmann Archive  
**Copyright:** Corbis  
**References:** Churchill, Roosevelt, Stalin



*Photos, Persons*

*Metadata*



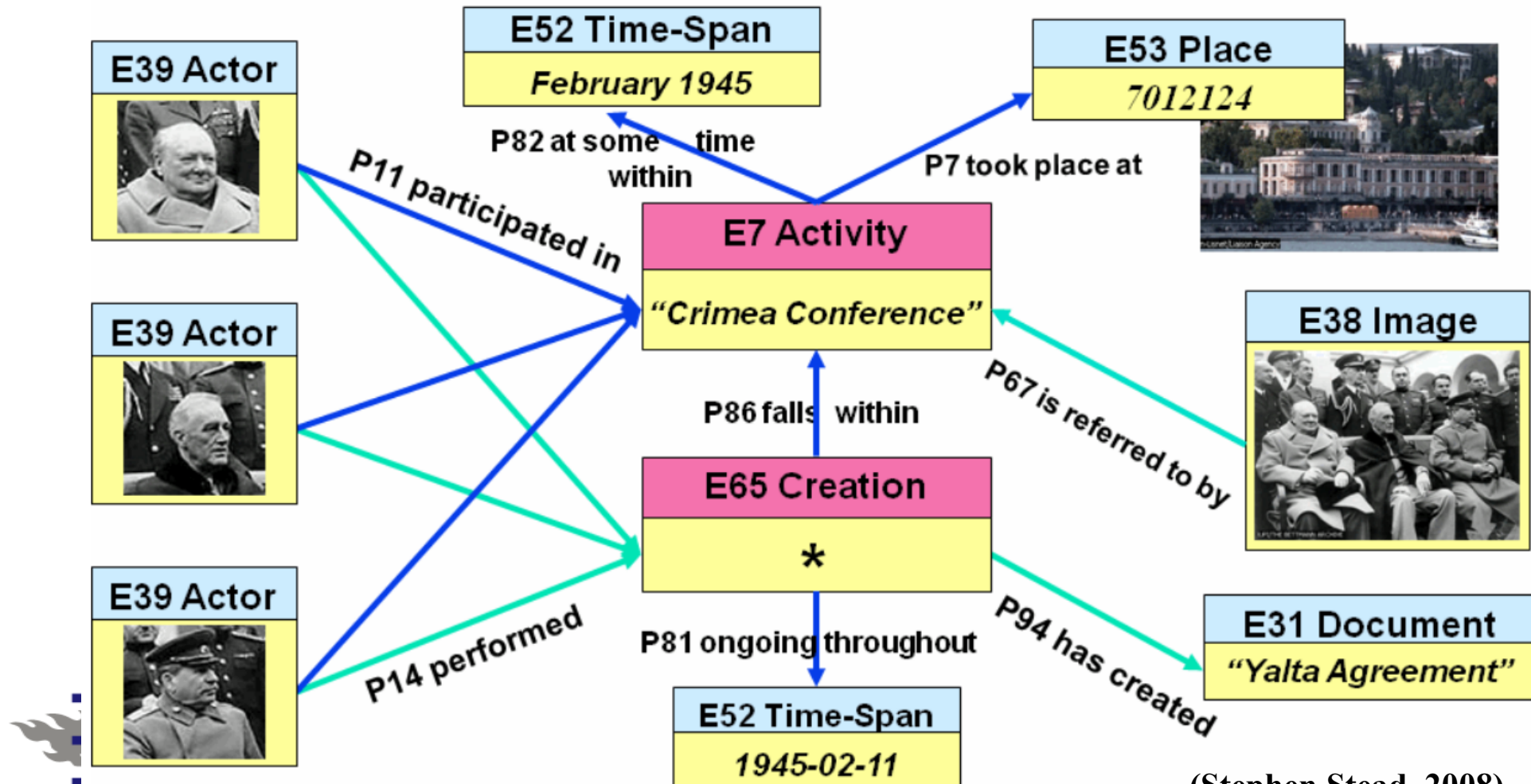
*About...*



41

[Slide by: Stephen Stead]

## Explicit Events, Object Identity, Symmetry



## Component 3: Datasets for the Finnish DH Cloud



# Examples

- Persons
  - Authority files, National Biography of Finland
- Places
  - Nimiarkisto (Kotus) -> NameSampo
  - PNR Gazetteer (Land Survey)
  - Map collections (National Archives)
- Events (Acricola and other timelines of Finnish history)
- Museums and Libraries
  - Finna, Fennica
  - Finnish National Gallery, art collections
  - BookSampo, CultureSampo
- Media
  - YLE and KAVI archives
- Legislation
  - Semantic Finlex



# LODI4DH Initiative Outline

## Work Packages

- **WP1 Core Ontologies for Data Linking**
  - WP1.1 Historical Places and Maps
  - WP1.2 Historical Persons
  - WP1.3 Historical Events
  - WP1.4 Keyword Concepts
- **WP2 Datasets for Research and Applications**
  - WP2.1 NameSampo Infrastructure (Nimiarkisto)
  - WP2.2 BiographySampo Infrastructure (Kansallisbiografia)
  - WP2.3 WarSampo Infrastructure
  - WP2.4 University of Helsinki Person Registry (1640-)
  - WP2.5 Semantic Finlex

- **WP3 Data Production and Analysis Services**
  - **WP3.1 Automatic Annotation Services**
  - **WP3.2 Automatic Anonymization Services**
  - **WP3.3 Data Analysis and Visualization Services**
- **WP4 Implementation and Support Services**
  - **WP4.1 Implementation**
  - **WP4.2 Online Learning Materials**
  - **WP4.3 Seminar Series and Data Clinic**
- **WP5 DARIAH Collaboration and International Visits**

- **Project Organization Core**

- 1. Aalto University, Department of Computer Science (coordinator)
  - » 1. School of Science (home)
  - » 2. School of Arts, Design and Architecture
  - » 3. School of Engineering
  - » 4. School of Electrical Engineering
  - » 5. School of Business
- 2. University of Helsinki, HELDIG – Helsinki Centre for Digital Humanities
  - » 1. Faculty of Arts (home)
  - » 2. Faculty of Social Sciences
  - » 3. Faculty of Science
  - » 4. Faculty of Law
  - » 5. Faculty of Theology
  - » 6. Faculty of Educational Sciences

## Research Collaborators



[Photo: Timo Väänänen]

# Funding

- Research projects at Aalto and HELDIG (running)
- University of Helsinki digi-strategy 2018- (pending)
- Academy of Finland FIRI Research Infra Roadmap Application (submitted)



# How to Deploy the Infrastructure for the Clients to Use?

# Linked Data Finland Living Lab: <http://ldf.fi>



---

[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.

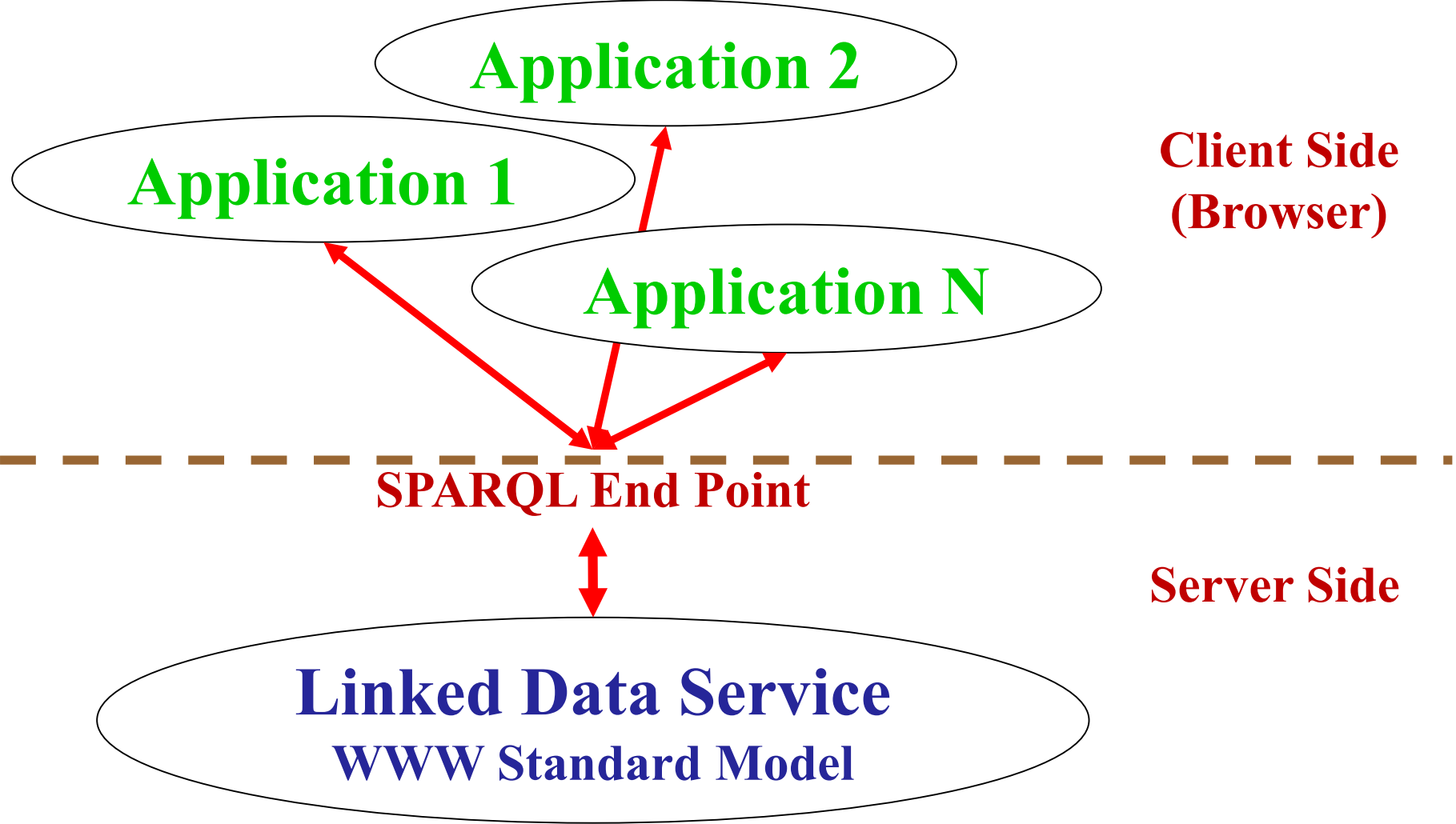
#### 7-star Linked Data Service

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:

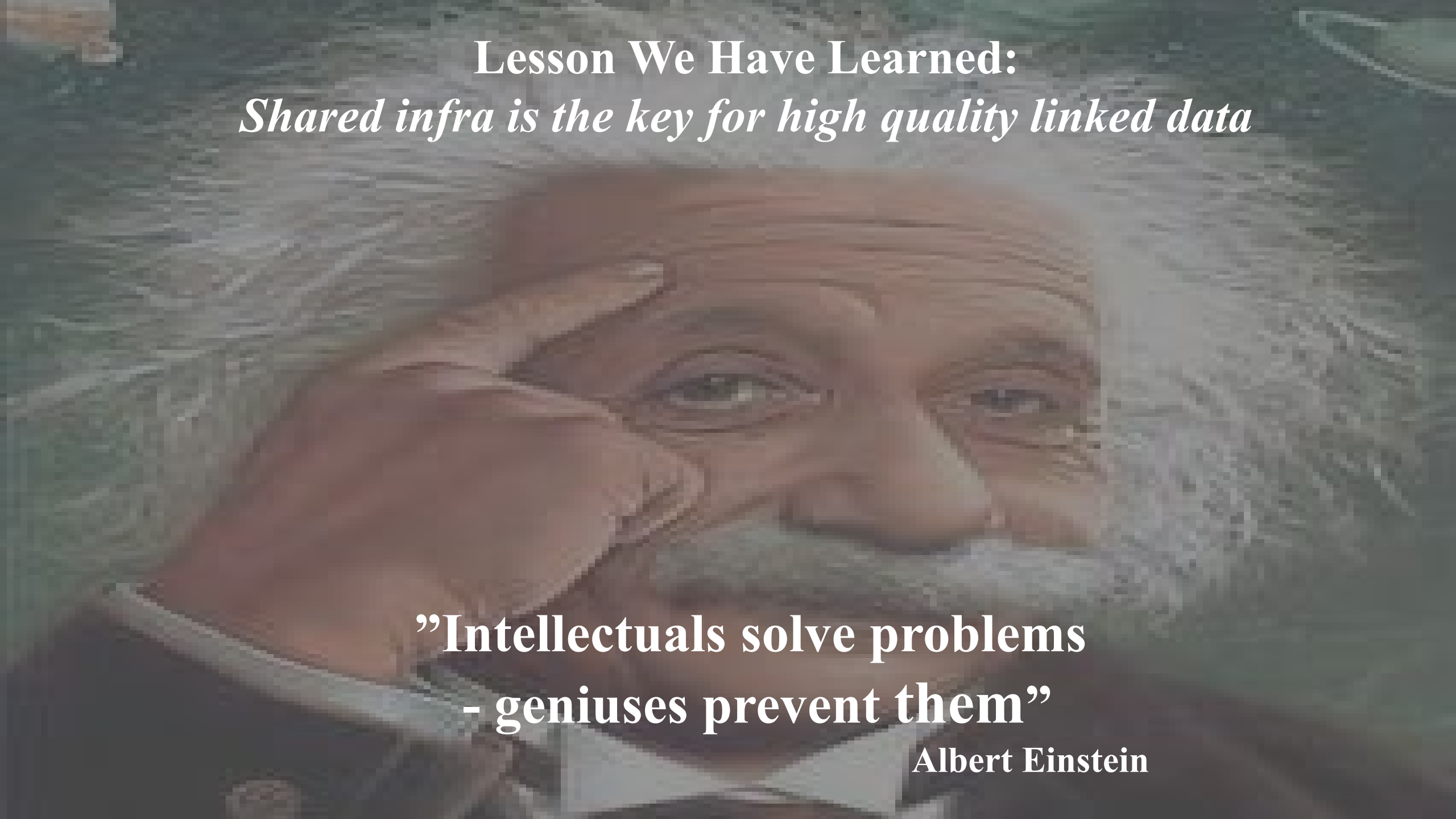
- ★★★★★★ Provide your data with a schema and documentation so that people can *understand and re-use* your data easily.
- ★★★★★★ Validate your data and denote its provenance so that people can *trust the quality* of your data.

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

# Developers View to Linked Data: Rich Internet Applications (RIA)



# Conclusions



**Lesson We Have Learned:**  
*Shared infra is the key for high quality linked data*

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

**Albert Einstein**

# More Info – Questions?

**LODI4DH**

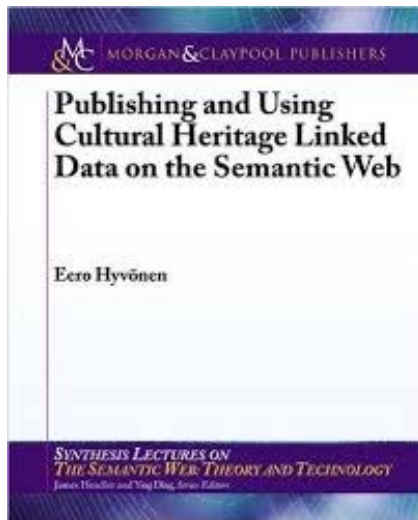
<https://seco.cs.aalto.fi/projects/lodi4dh/>

**Semantic Web & Linked Data**

<http://www.w3.org/standards/semanticweb/>

**Sampo Model & Applications**

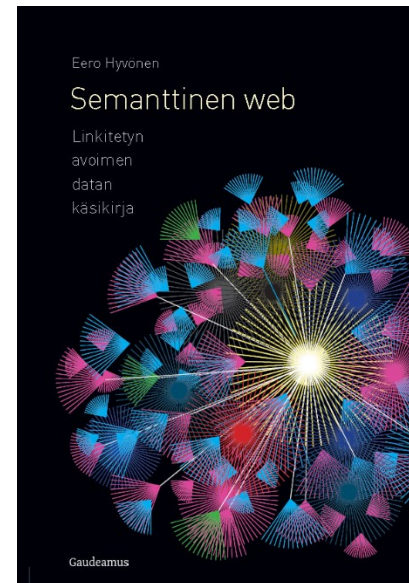
<http://seco.cs.aalto.fi/publications>



In English

2012

<https://www.amazon.com/Publishing-Cultural-Heritage-Synthesis-Technology/dp/1608459977>



In Finnish

2018

<https://www.gaudemus.fi/semanttinen-web/>