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

Aalto University

UNIVERSITY OF HELSINKI

SeCo semantic computing

Aalto University

Linked Data and Semantic web

UNIVERSITY OF HELSINKI

SeCo
semantic computing

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

Aalto University

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

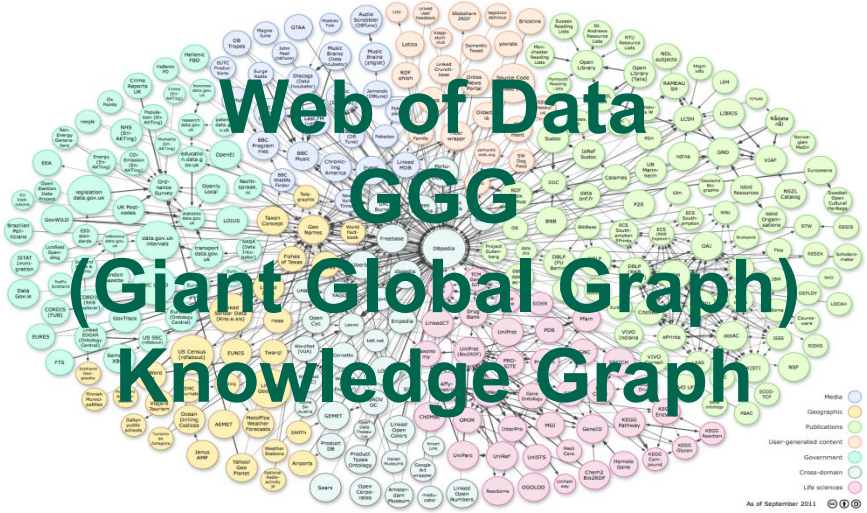
SEMANTIC WEB KICK-OFF IN FINLAND - ÄLYKÄS WWW SUOMESSA

Helsingin yliopisto, Porthania, sali P3

Perjantai 2.11.2001, klo 9:00-16:30

Word Wide Web Consortium

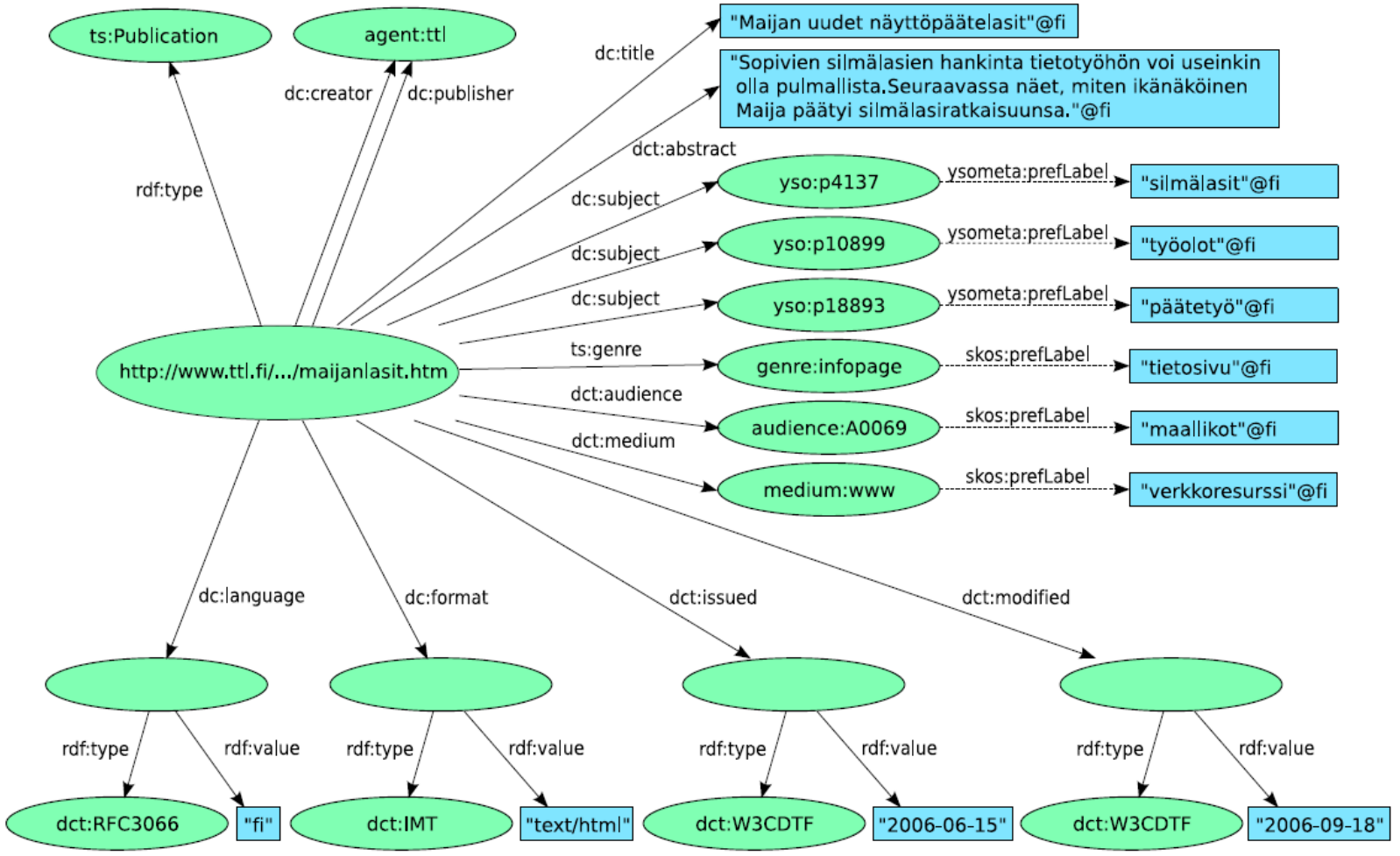
Web of Pages
WWW



Web of Data
GGG
(Giant Global Graph)
Knowledge Graph

WWW Document:
Semantic portal HealthFinland

Corresponding GGG graph (in RDF)



[Suominen et al.]

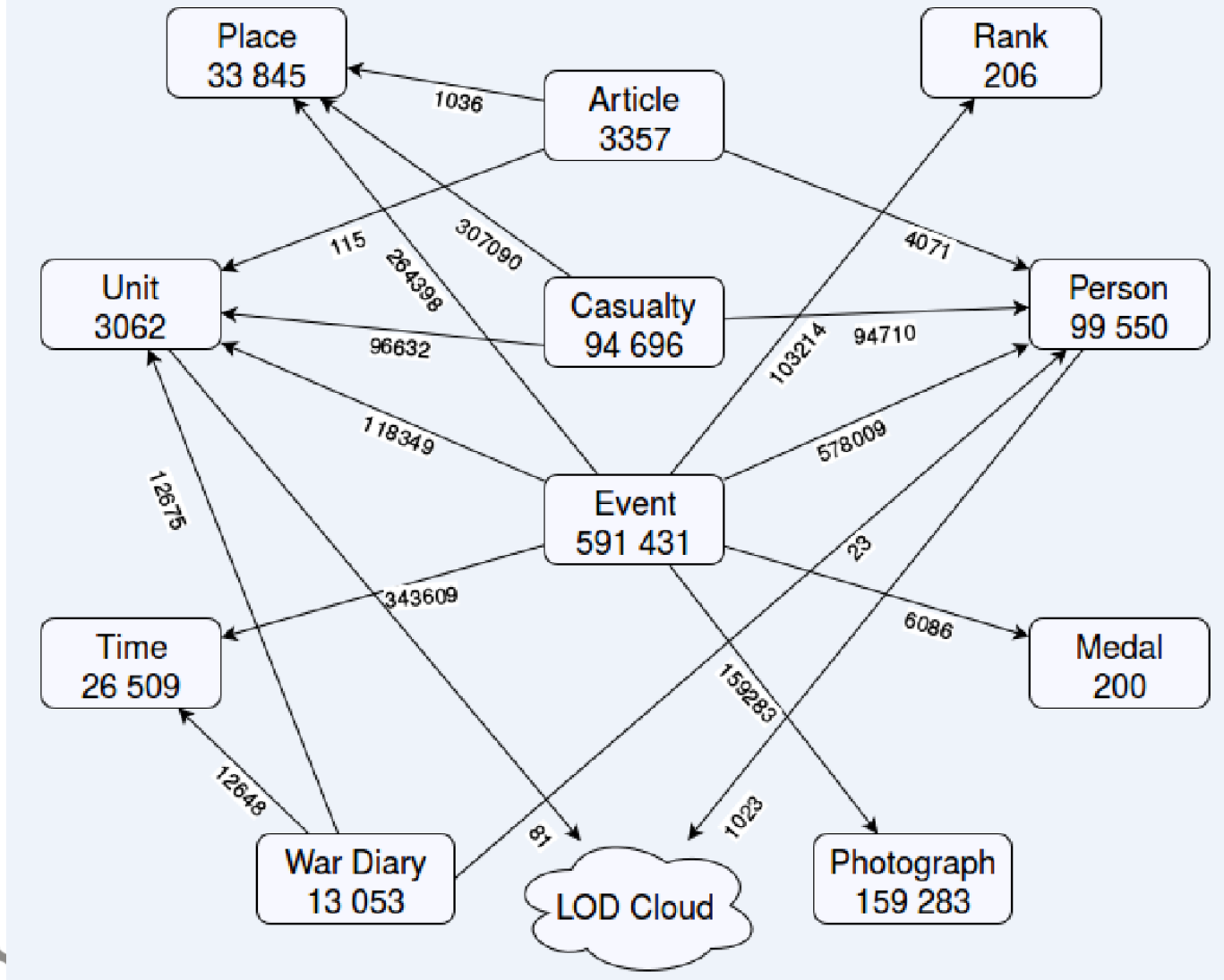
Example Dataset, Data Service, and Application:

WarSampo – Finnish WW2 on the Semantic Web

http://sotasampo.fi

(Hyvönen et al., ESWC 2016)

WarSampo Linked Open Data Cloud



12M triples knowledge graph

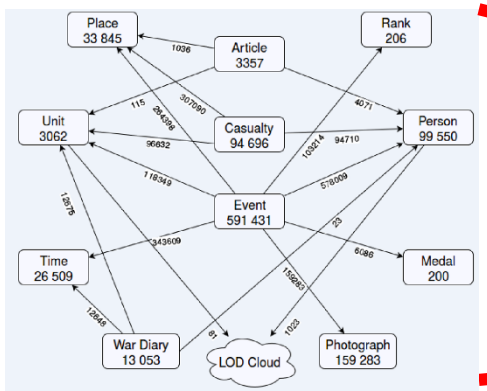
[Leskinen et al.]

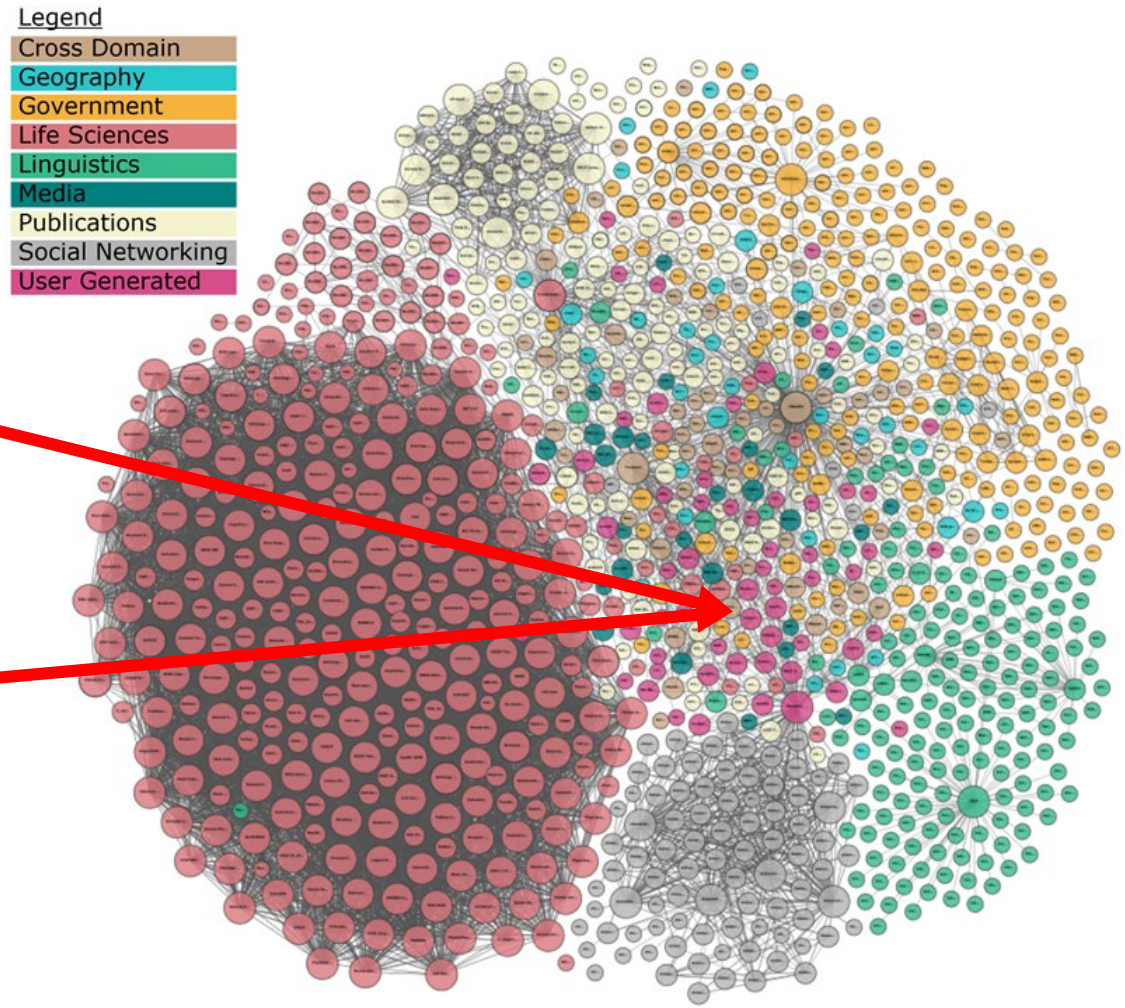
Semantic Portal http://sotasampo.fi/en
8 Perspectives to War History

https://vimeo.com/212249404

Linked Open Data Cloud 2018 http://linkeddata.org

WarSampo





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
- …

Aalto University

The Vision of LODI4DH

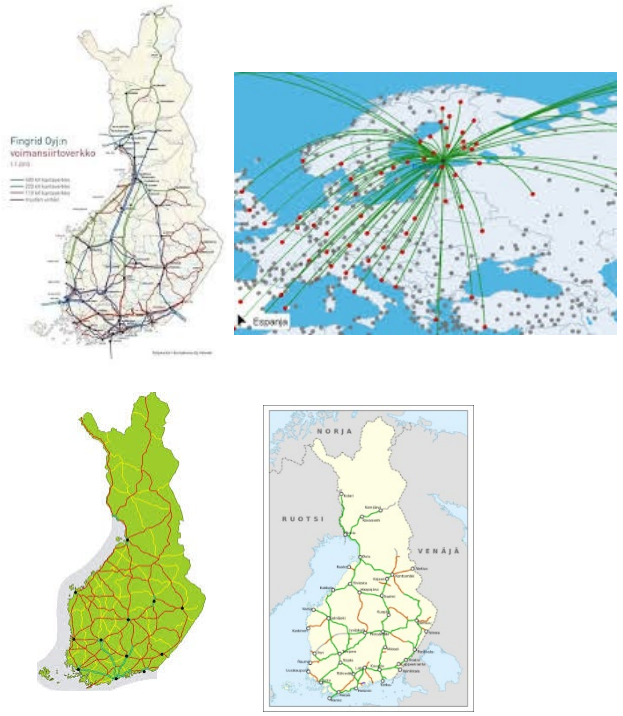
Linked Open Data Infrastructure for Digital Humanities in Finland

UNIVERSITY OF HELSINKI

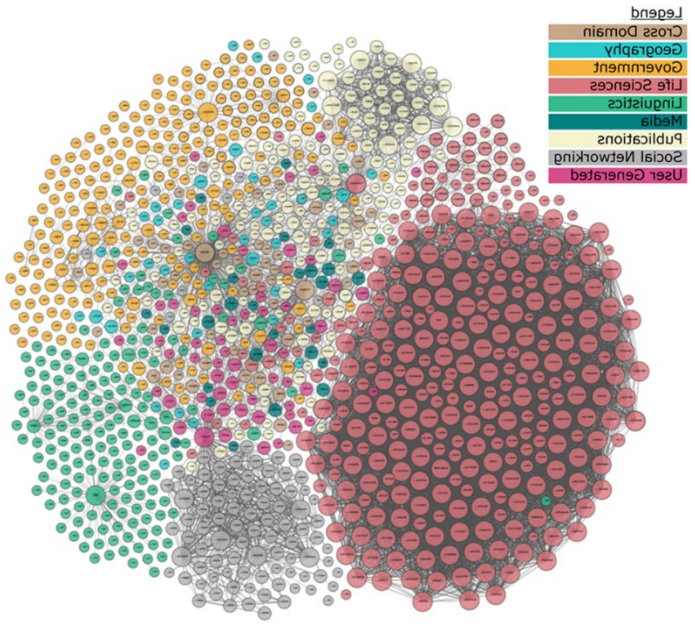
SeCo semantic computing

Linked Data Infrastructure

Traditional Infras



Linked Data Infra



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

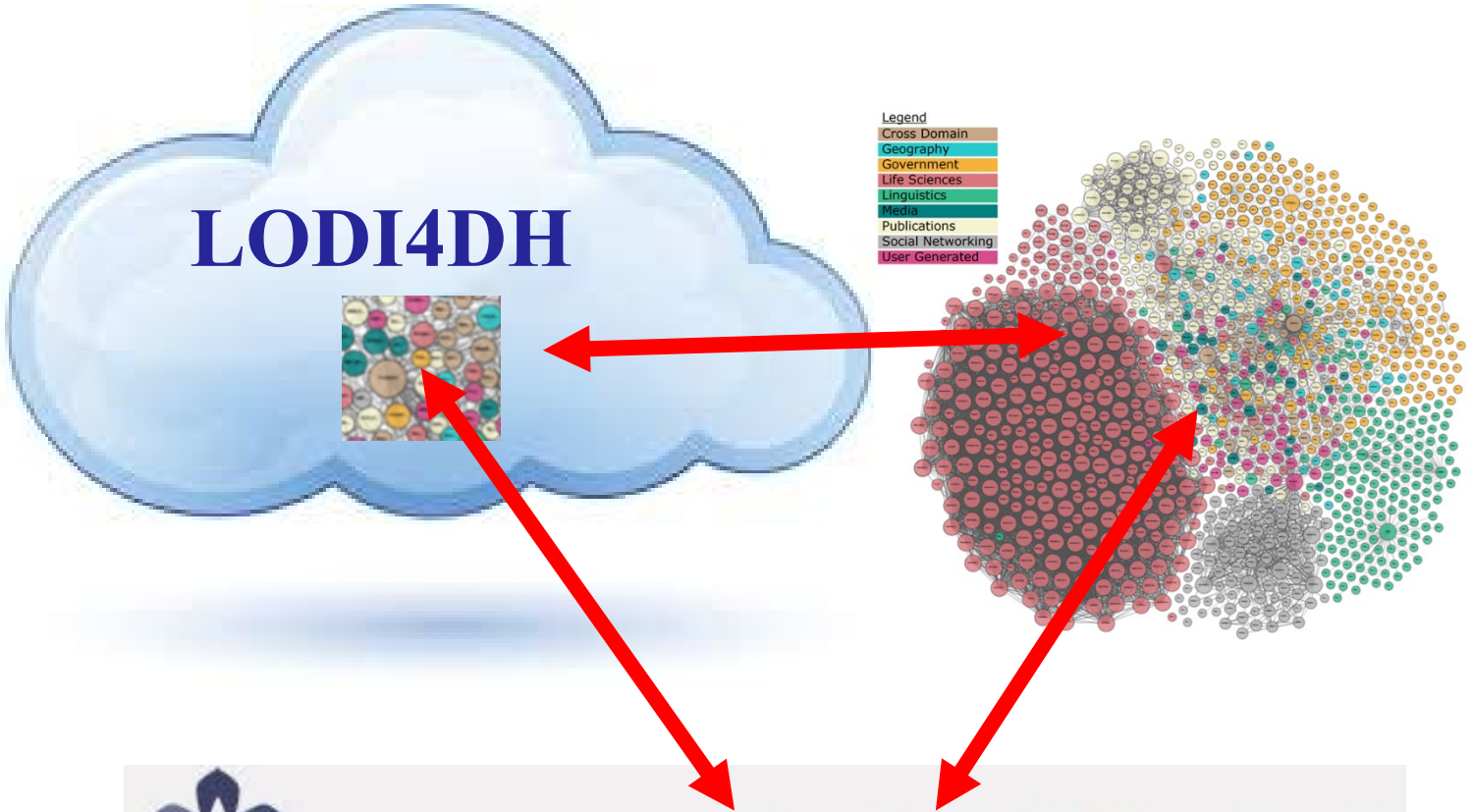
- National Finnish Ontology Project 2003-2012
- Linked Data Finland 2012-2014
- Linked Open Science 2014-2017
- LODI4DH 2018-

Linked Open Data

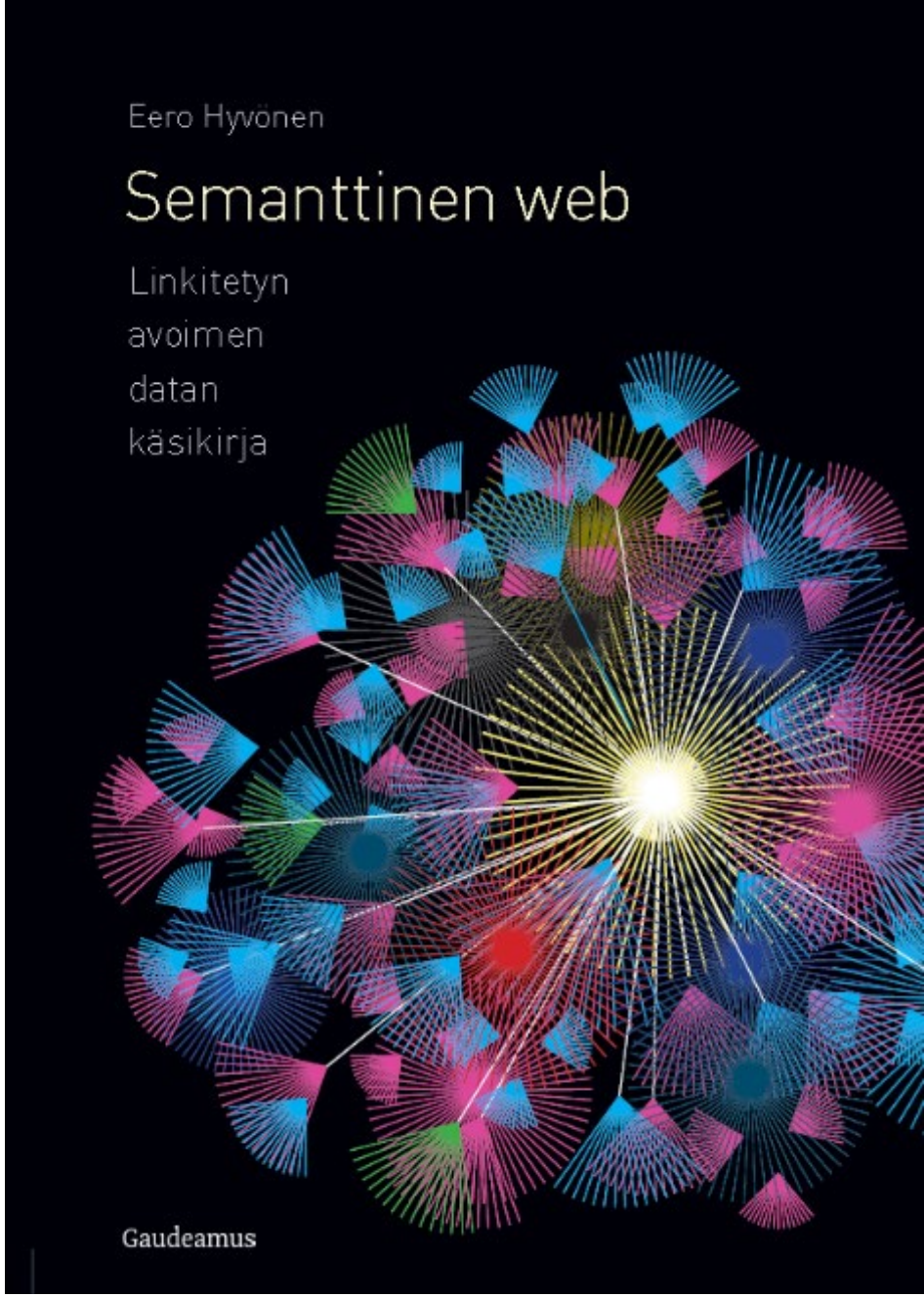
http://linkddata.org

Word Wide Web Consortium

LODI4DH



DARIAH-EU Digital Research Infrastructure for the Arts and Humanities



Aalto University

How Does Linked Data Work in Practise?

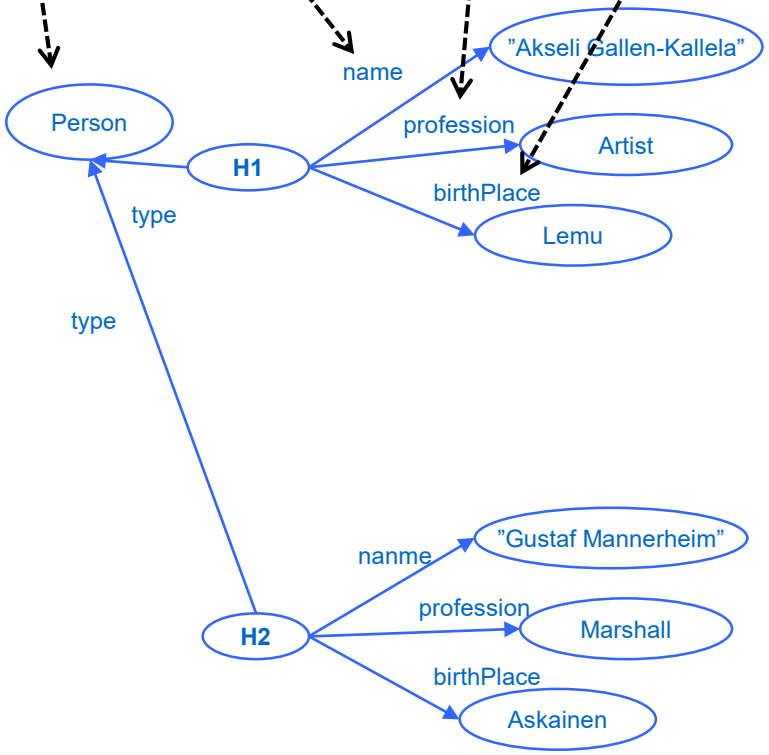
UNIVERSITY OF HELSINKI

SeCo
semantic computing

Biographical Registries Collect Data about Persons

Person Name Profession Birth Place

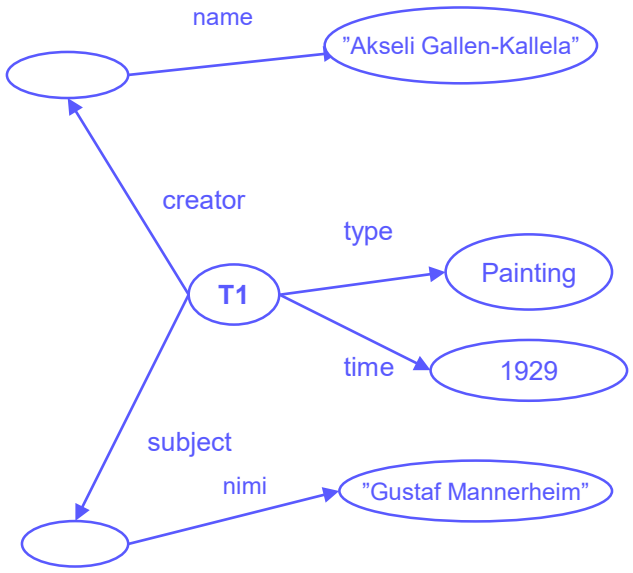
henkilö	nimi	ammatti	syntymapaikka	...
H1	Akseli Gallen-Kallela	taiteilija	Lemu	
H2	Gustaf Mannerheim	marsalkka	Askainen	
	..			



Biography Center

Art Museum Catalogs Paintings

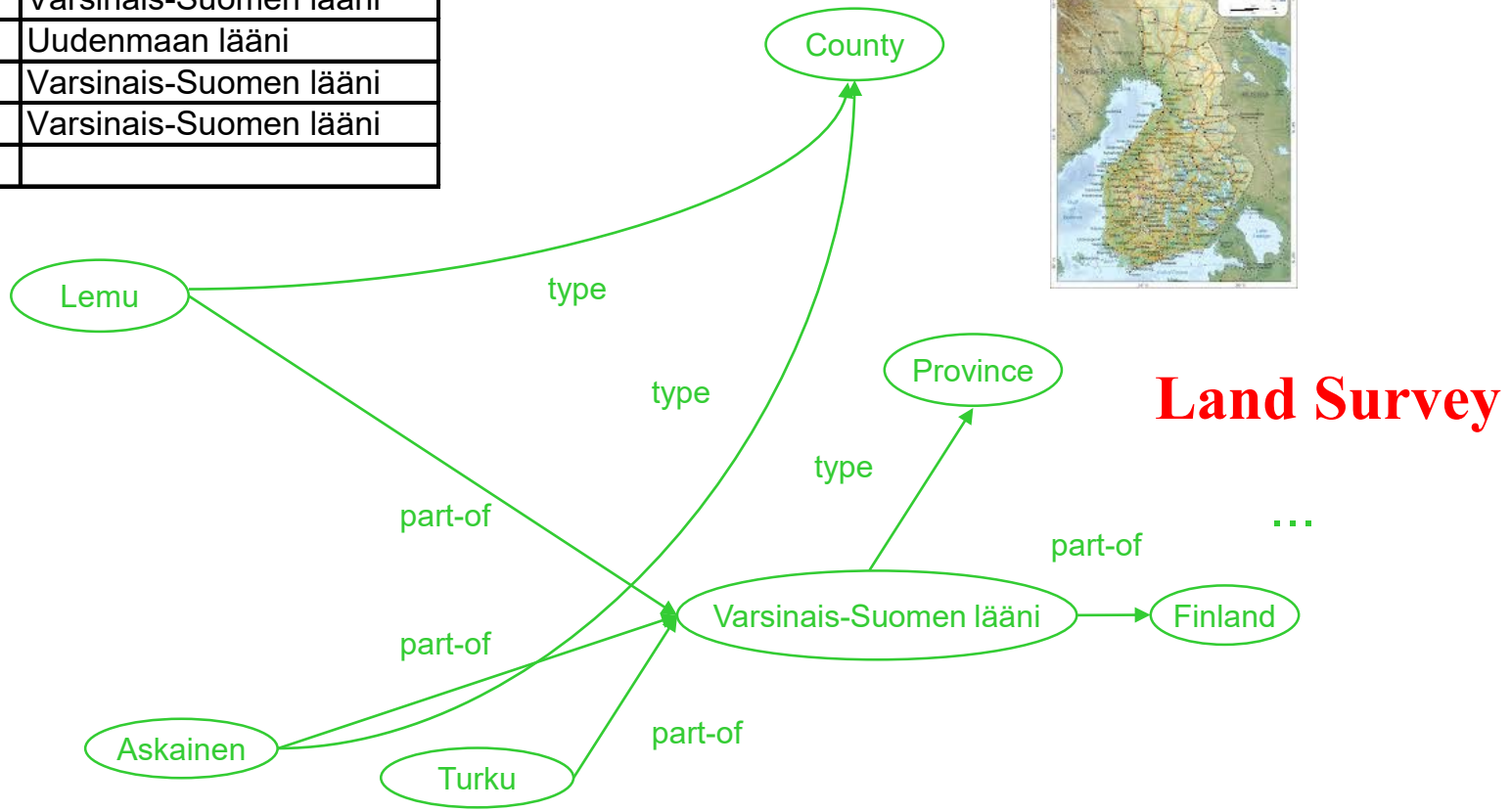
teos	nimi	tekijä	aika	aihe	...
T1	Mannerheimin muotokuva	Akseli Gallen-Kallela	1929	Gustaf Mannerheim	
T2	Aino-triptyykki	Akseli Gallen-Kallela	1891	Aino, Kalevala	
...					



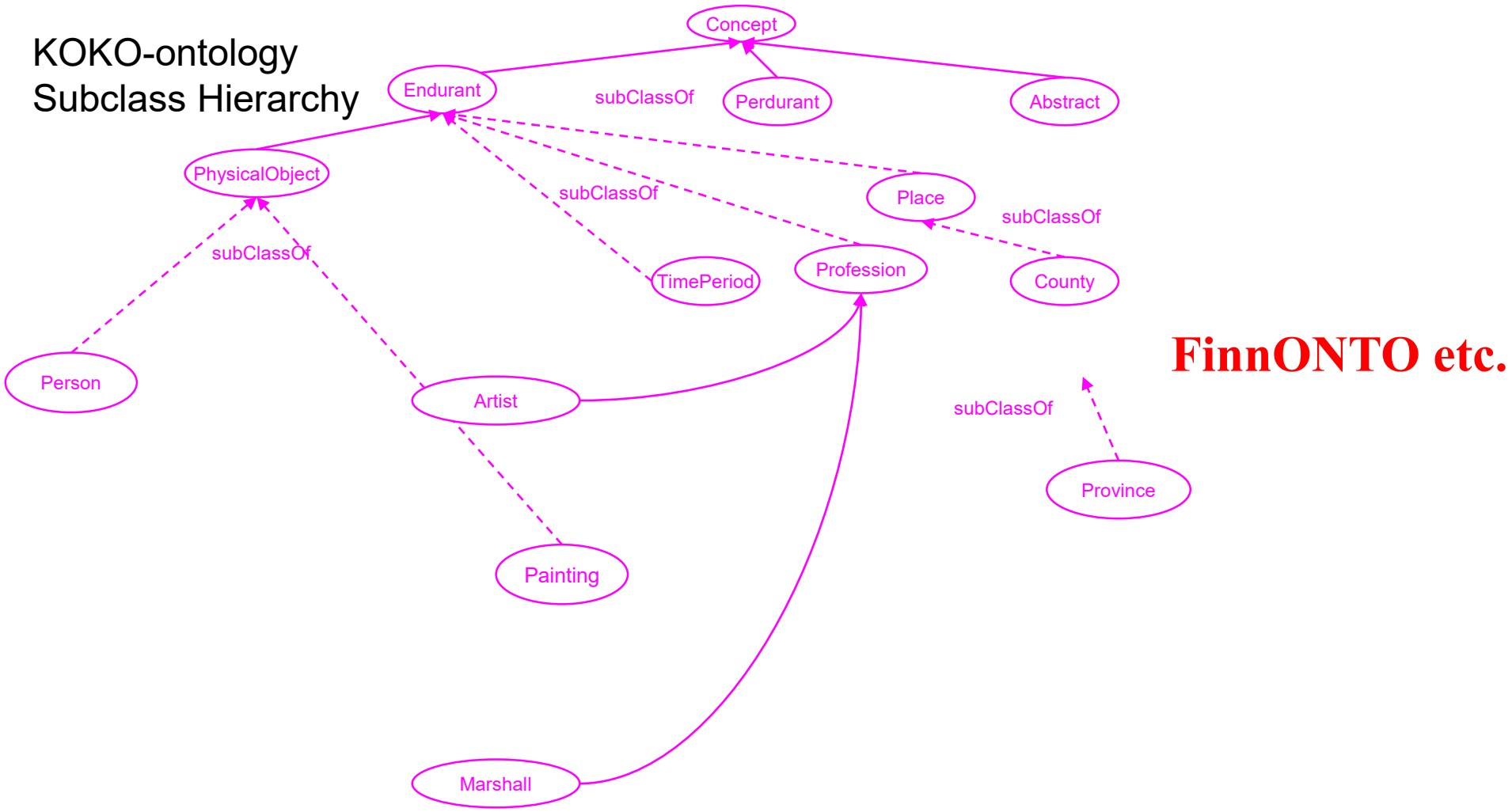
Art Museum Collection

Land Survey Organizations Know Places

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

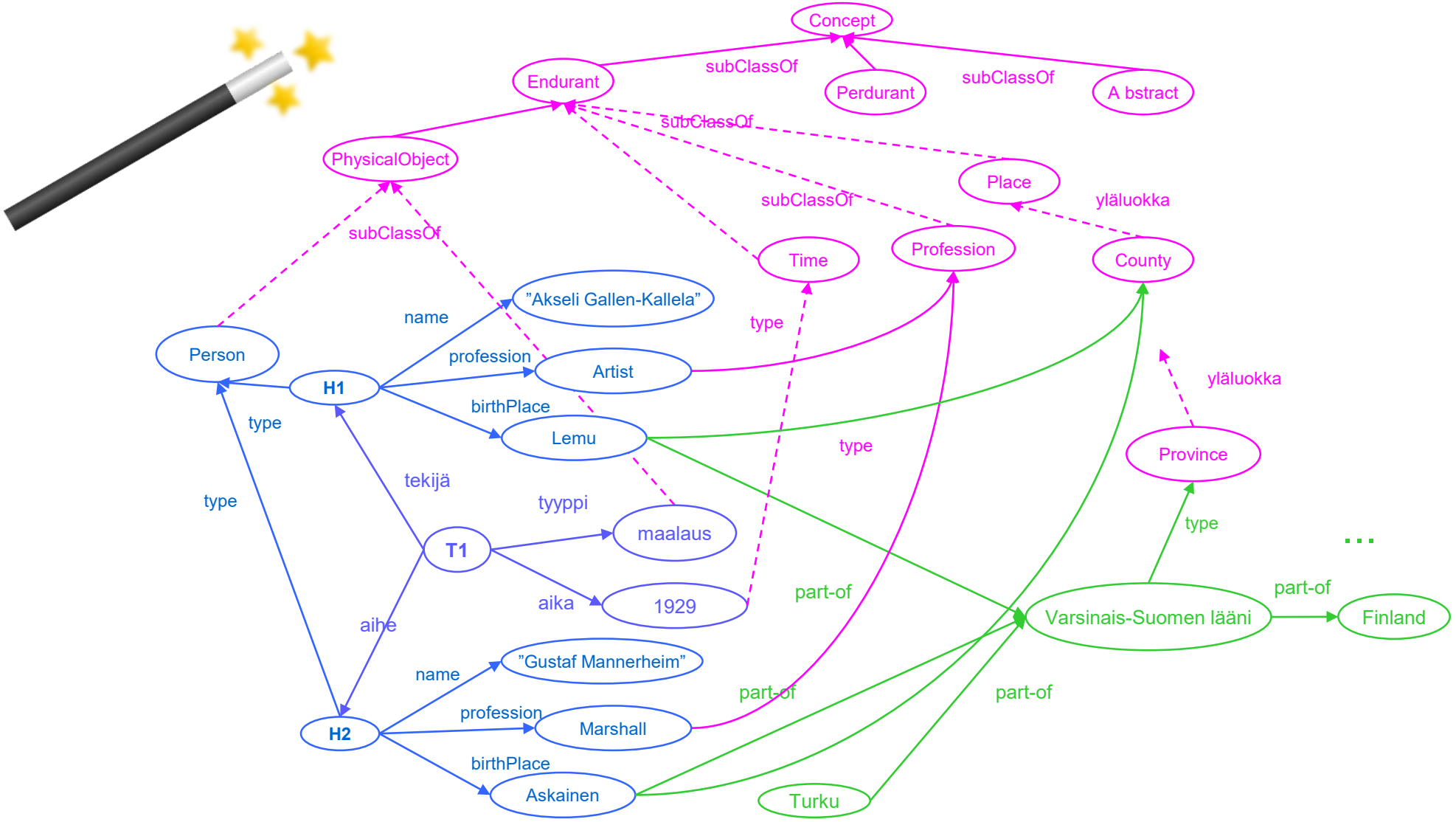


Ontologies Developed by Semantic Web Researchers



FinnONTO etc.

RDF Connects and Harmonizes Linked Data into a GGG



Aalto University

1+1>2

UNIVERSITY OF HELSINKI

SeCo
semantic computing

Aalto University

AI

UNIVERSITY OF HELSINKI

SeCo
semantic computing

Aalto University

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?

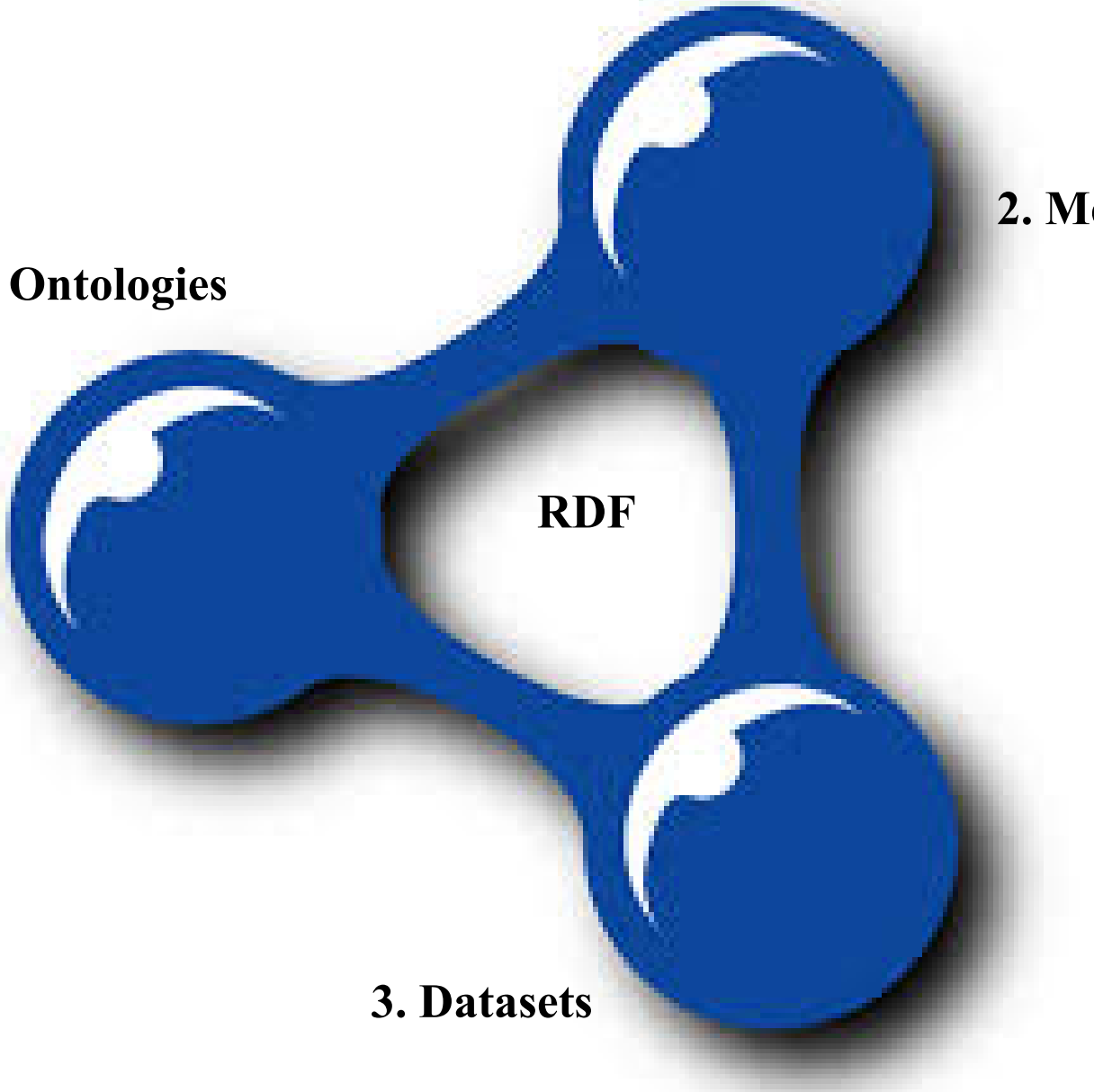
SHARED INFRA NEEDED!

UNIVERSITY OF HELSINKI

SeCo semantic computing

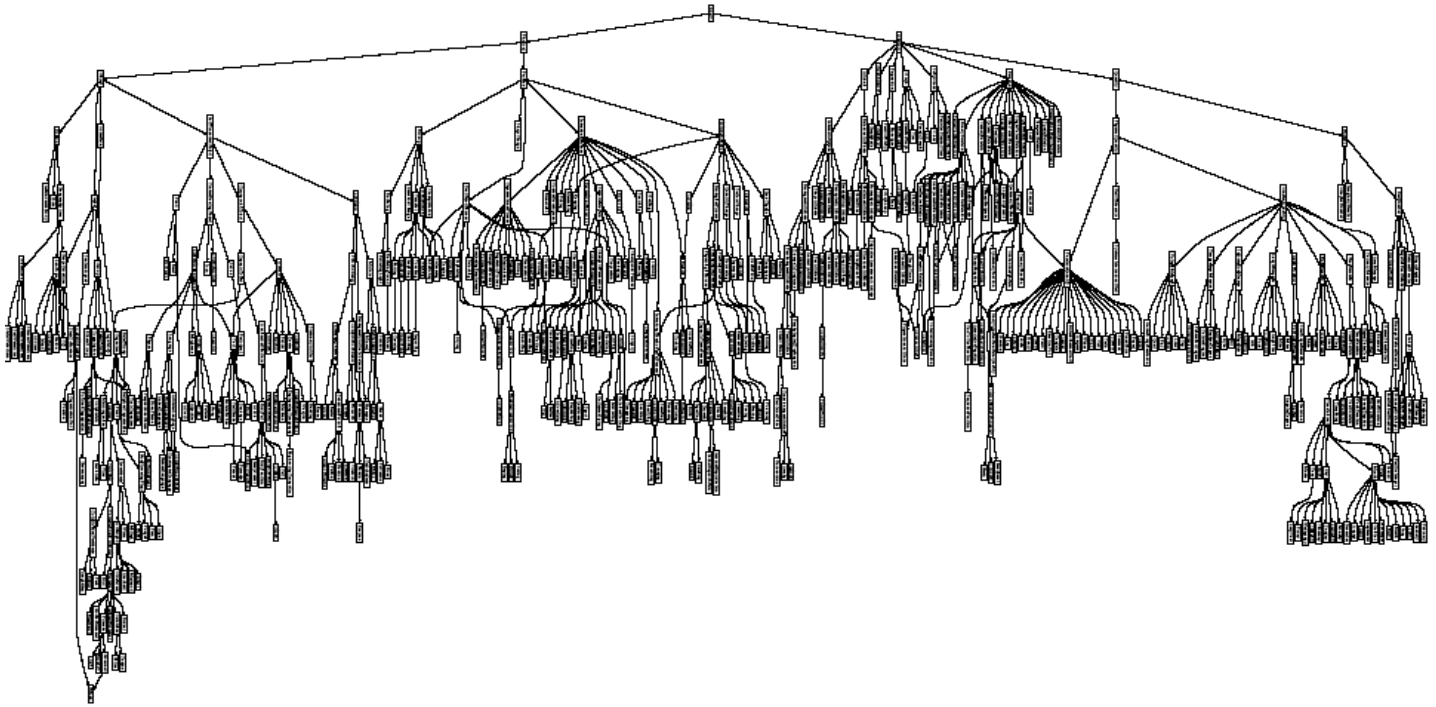
Major Components Needed

Key Components of a Linked Data Intrastructure



Aalto University

Component 1: Domain Ontologies

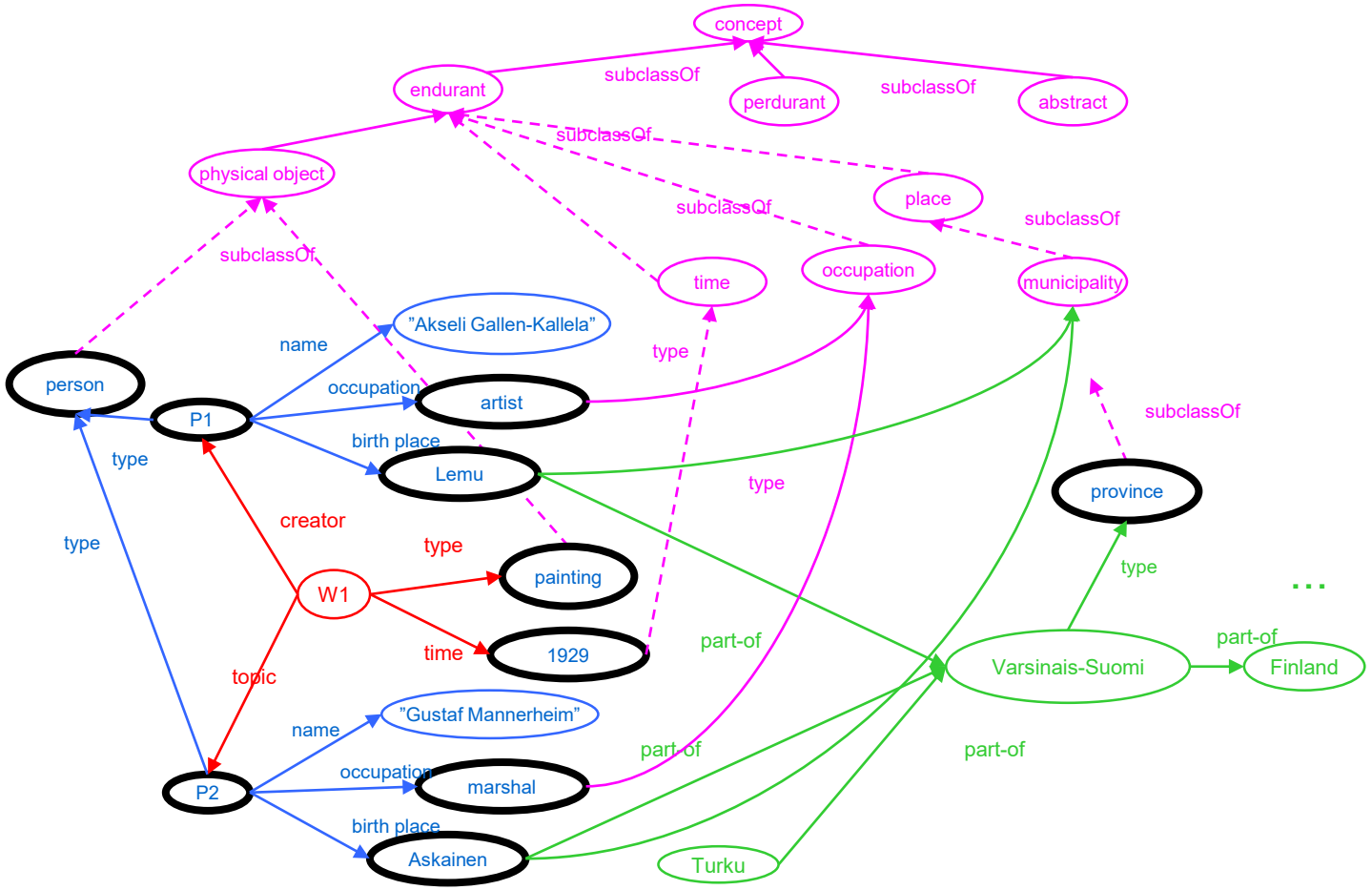


UNIVERSITY OF HELSINKI

SeCo
semantic computing

Concepts (IDs) and Data Model Glue Data Together

Solution: Shared Ontology Infrastructure



Aalto University

Ontology of Historical Persons

UNIVERSITY OF HELSINKI

SeCo
semantic computing

Resolving Identities

Aalto University

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

Pjotr Tšaikovski (fi)
Пётр Ильи́ч Чайко́вский (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)

UNIVERSITY OF HELSINKI

SeCo semantic computing

Aalto University

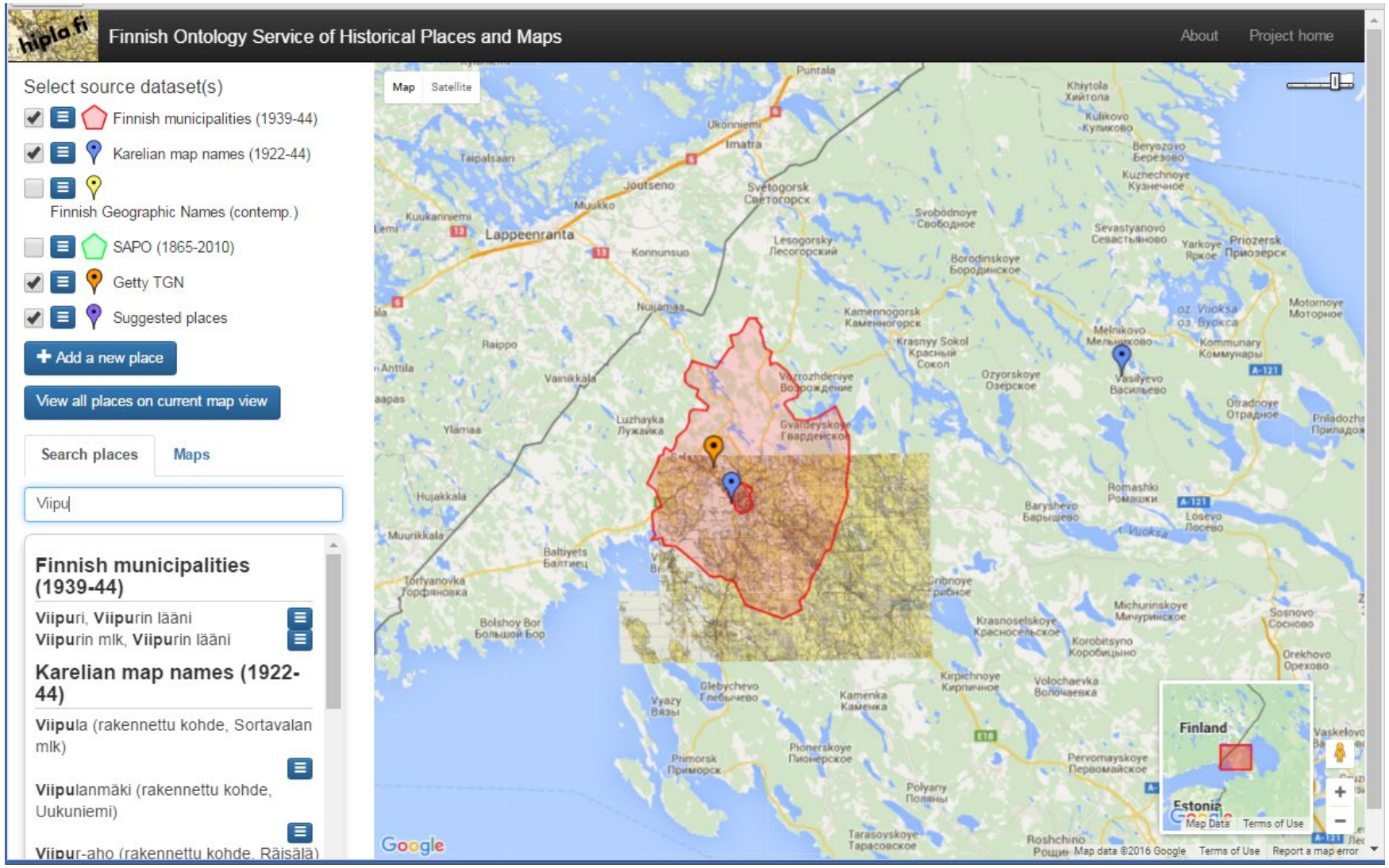
Ontologies of Historical Places and Maps

UNIVERSITY OF HELSINKI

SeCo
semantic computing

Finnish Ontology Service of Historical Places and Maps

http://hipla.fi



Aalto University

Ontology of Historical Events

”Semantic glue” of everything

UNIVERSITY OF HELSINKI

SeCo
semantic computing

Example of an Event on HISTO Ontology of Finnish History

Aalto University

SAHA3 | historia - search

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

historiallinen tapahtuma: YYA-sopimus raukesi

[edit]

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

UNIVERSITY OF HELSINKI

SeCo semantic computing

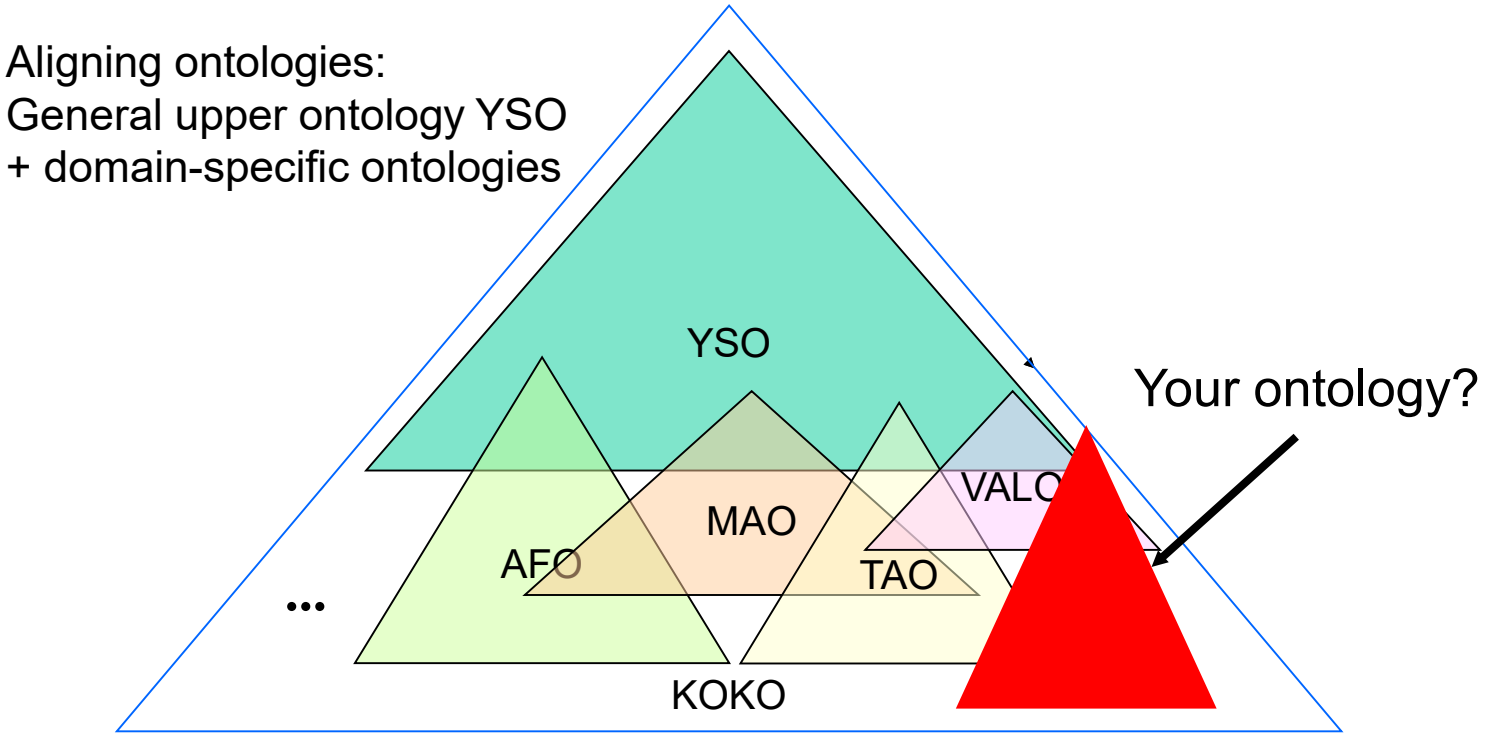
Aalto University

Ontologies of Historical Keyword Concepts

UNIVERSITY OF HELSINKI

SeCo
semantic computing

Holistic Collaborative Finnish Ontology KOKO



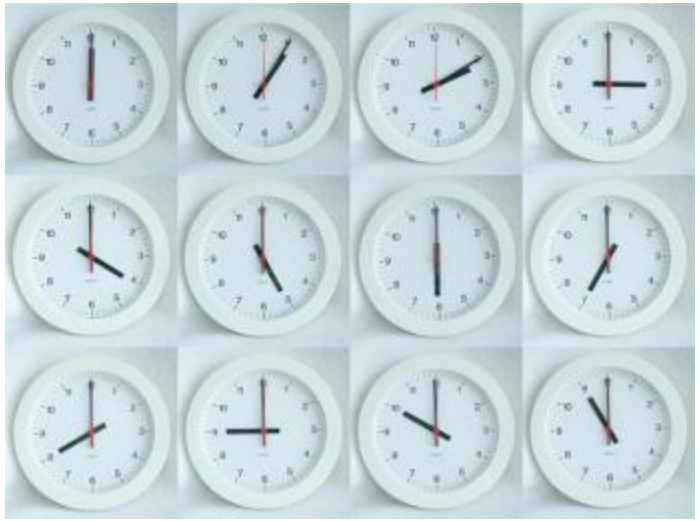
[Hyvönen et al., ESWC 2009]

http://finto.fi

FINNONTO

Aalto University

Ontology of Historical Times and Periods



UNIVERSITY OF HELSINKI

SeCo
semantic computing

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

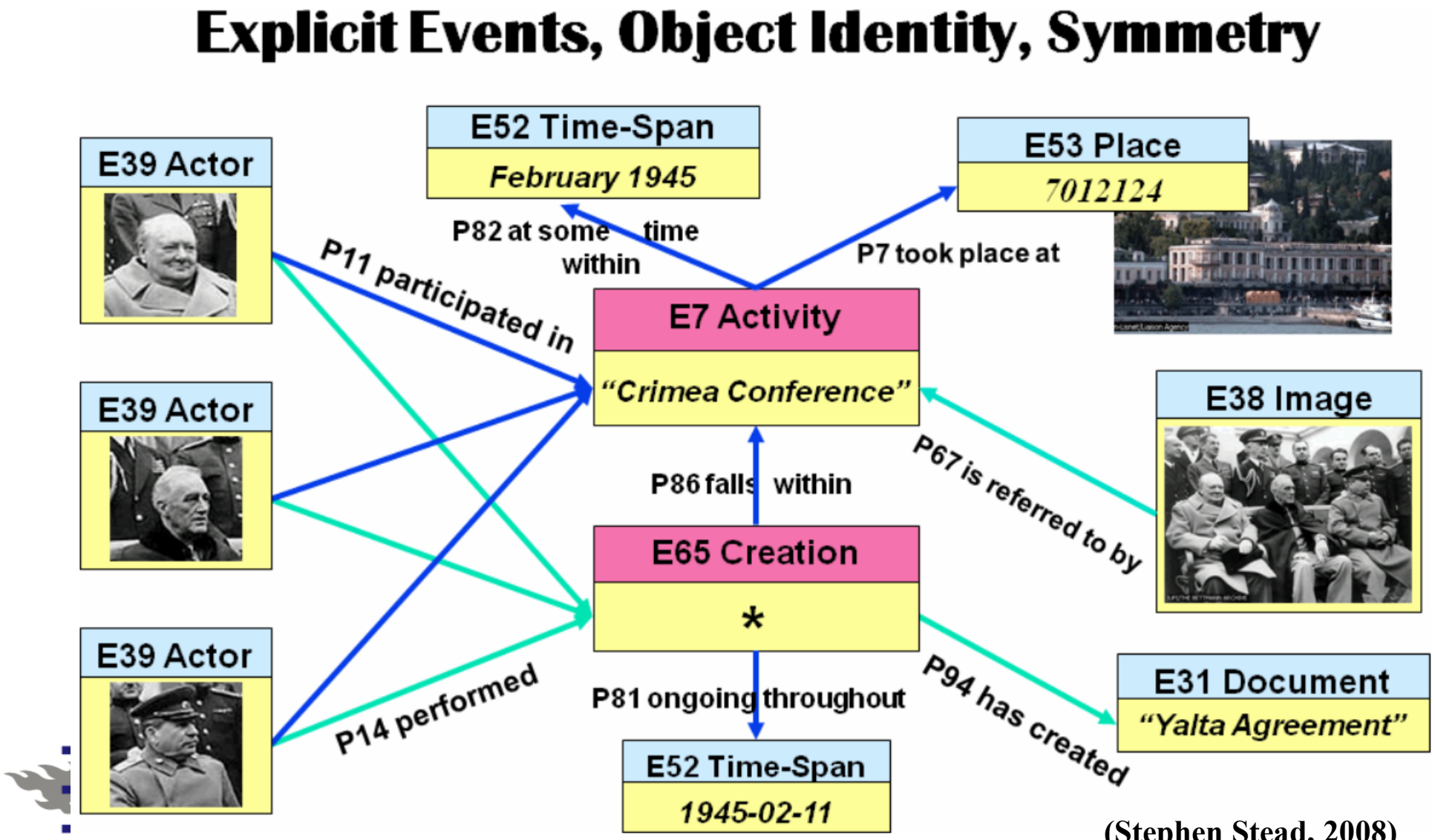
Metadata

Photos, Persons

About…

[Slide by: Stephen Stead]

Event-centric view: CIDOC CRM



(Stephen Stead, 2008)

Aalto University

Component 3: Datasets for the Finnish DH Cloud

UNIVERSITY OF HELSINKI

SeCo semantic computing

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

Aalto University

LODI4DH Initiative Outline

UNIVERSITY OF HELSINKI

SeCo
semantic computing

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)

Aalto University

How to Deploy the Infrastructure for the Clients to Use?

UNIVERSITY OF HELSINKI

SeCo
semantic computing

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.

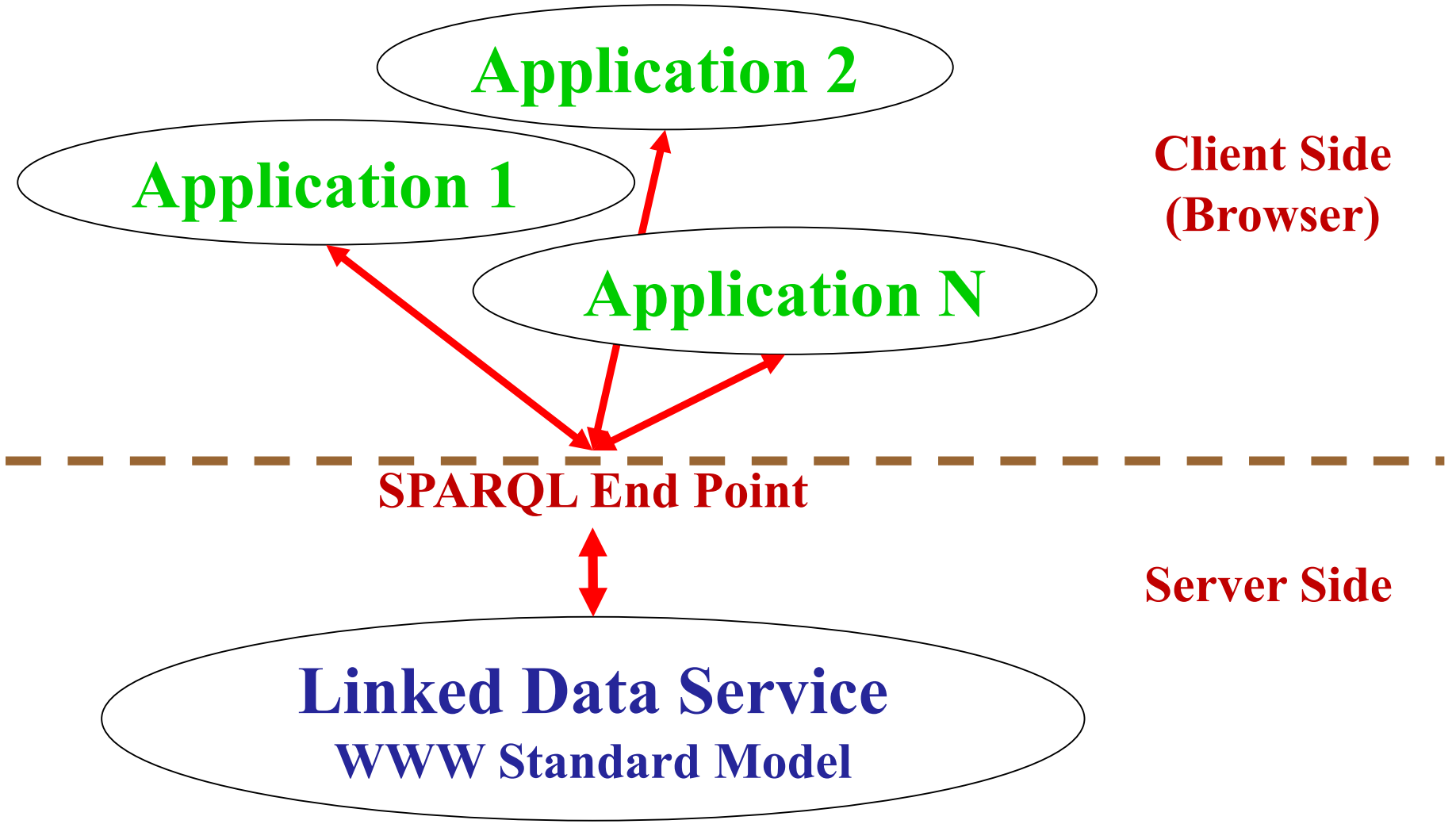
7-star Linked Data Service

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)



Aalto University

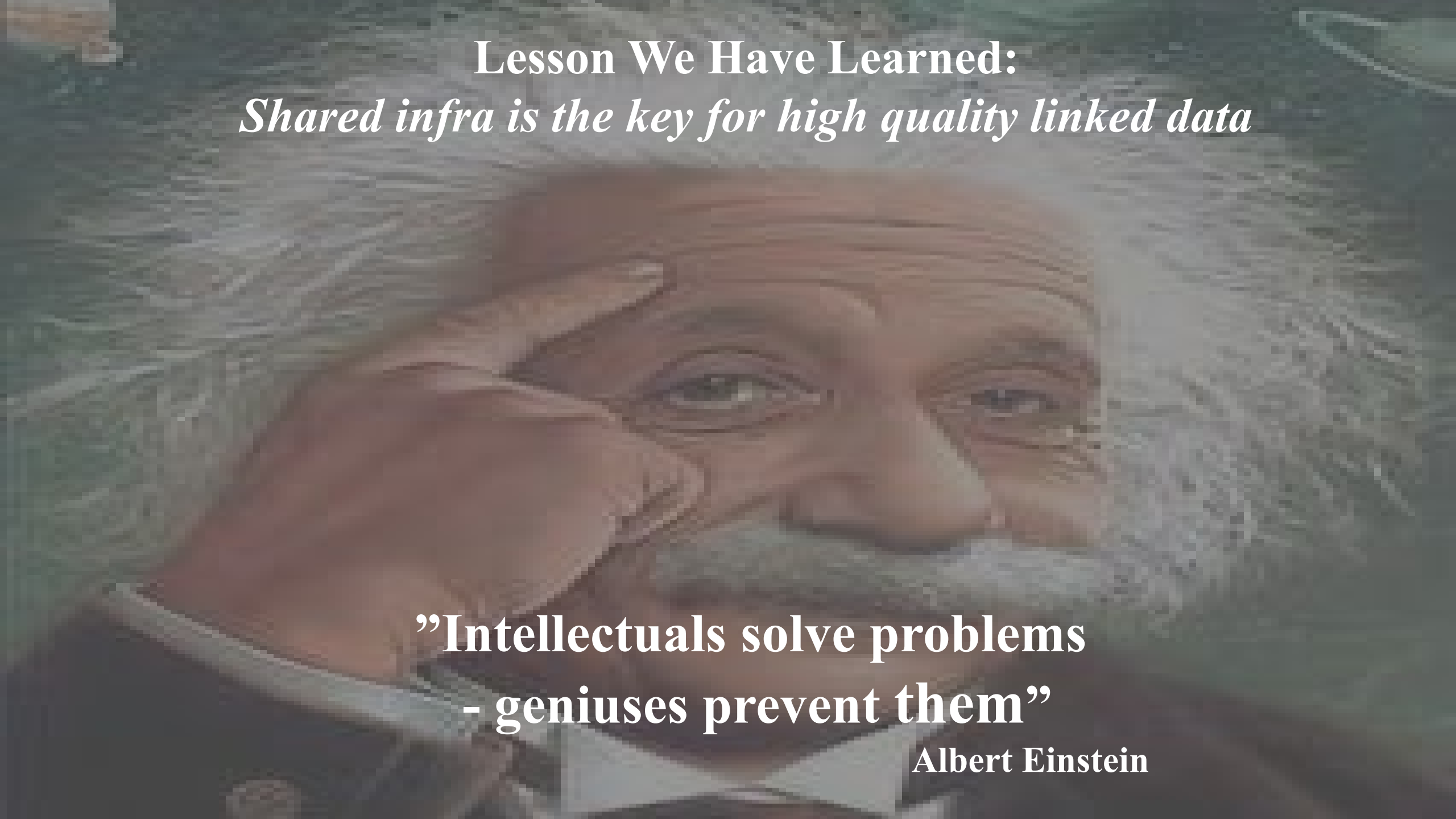
Conclusions

UNIVERSITY OF HELSINKI

SeCo
semantic computing

Lesson We Have Learned:

Shared infra is the key for high quality linked data

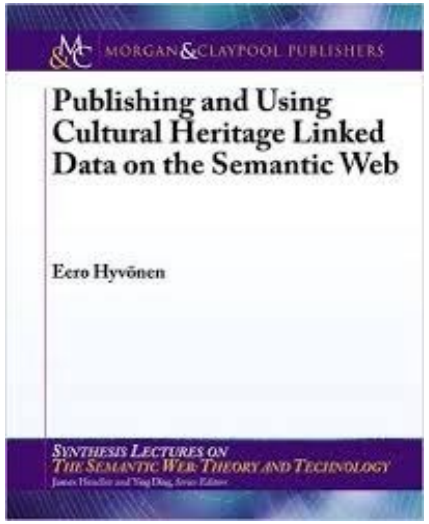


”Intellectuals solve problems
- geniuses prevent them”

Albert Einstein

More Info – Questions?

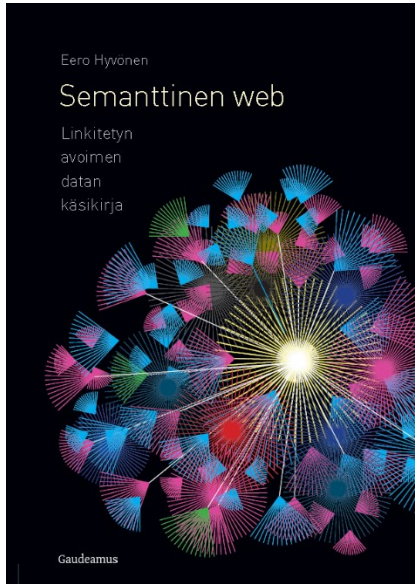
LODI4DH
https://seco.cs.aalto.fi/projects/lodi4dh/
Semantic Web & Linked Data
http://www.w3.org/standards/semanticweb/
Sampo Model & Applications
http://seco.cs.aalto.fi/publications



In English

2012

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



In Finnish

2018

https://www.gaudeamus.fi/semanttinen-web/