Linked Open Data Infrastructure for Digital Humanities in Finland

Aalto Data Day, 25.5.2018
Espoo, Finland

Eero Hyvönen
Prof., Director
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

Aalto University

- **Background:** History
- **Vision:** Linked Open Data Infrastructure for Digital Humanities in Finland (LODI4DH)
- **Challenges:** Data Complexity & Production
- **Solution:** Components of LODI4DH

UNIVERSITY OF HELSINKI

SeCo semantic computing

Aalto University

Background: History 2001-

UNIVERSITY OF HELSINKI

SeCo
semantic computing

Semantic Web Activity at W3C Started 2001

Aalto University

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 »

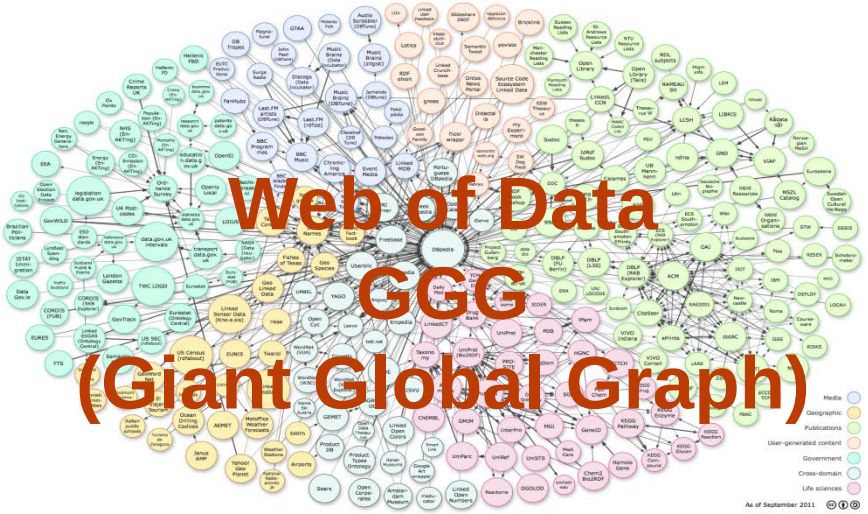
March 12, 2009

UNIVERSITY OF HELSINKI

SeCo semantic computing

Aalto University

Web of Pages
WWW

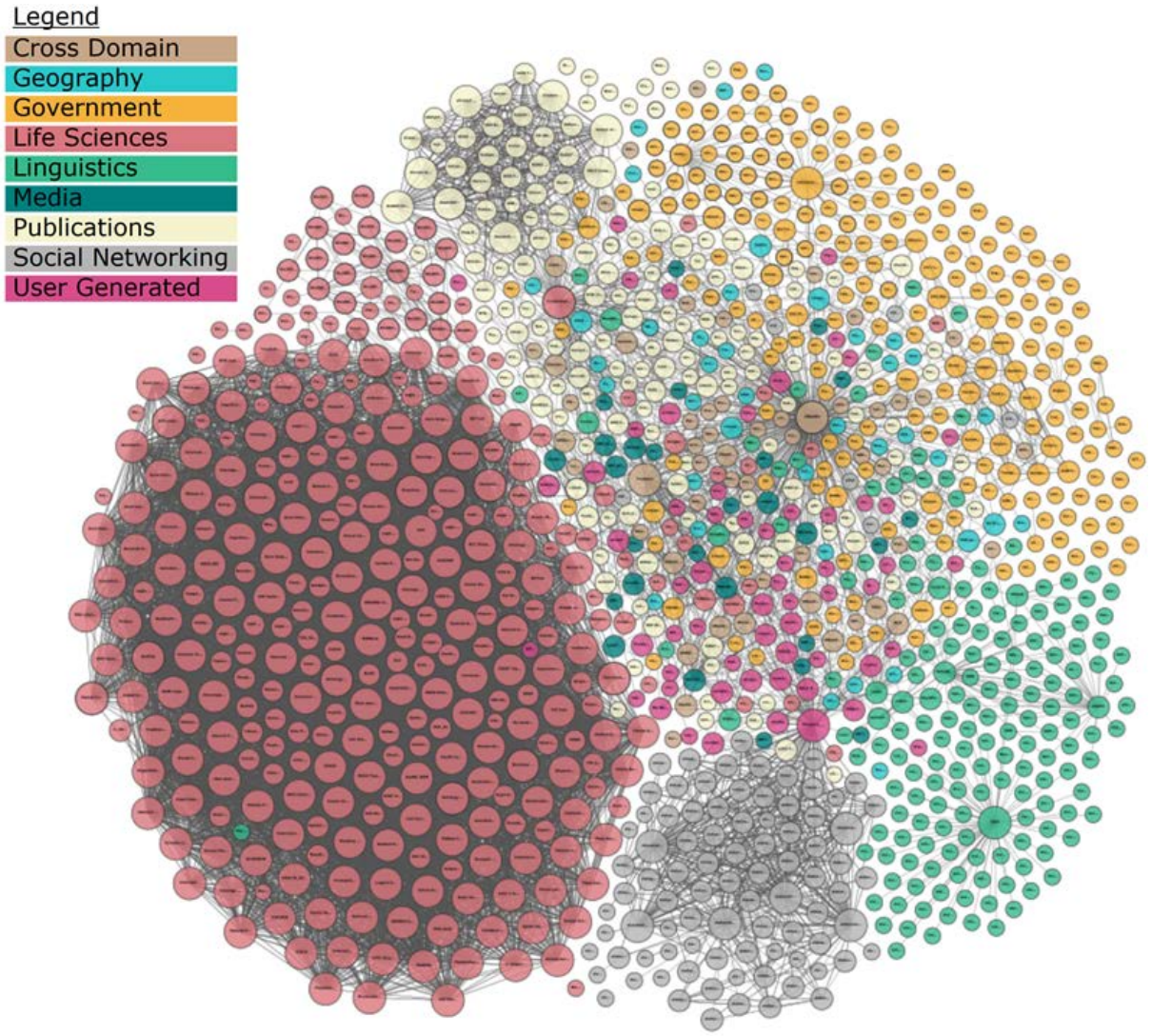


Web of Data
GGG
(Giant Global Graph)

UNIVERSITY OF HELSINKI

SeCo semantic computing

Linked Open Data Cloud 2018



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

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

UNIVERSITY OF HELSINKI

SeCo semantic computing

(Hyvönen, ed., 2002)

History behind this talk

- Semantic Portals for Cultural Heritage
  - 2004 MuseumFinland – Finnish Museums on the Semantic Web
    - » http://www.museosuomi.fi
  - 2008 CultureSampo – Finnish Culture on the Semantic Web 2.0
    - » http://www.kulttuurisampo.fi
  - 2011 BookSampo – Fiction Literature on the Semantic Web
    - » http://www.kirjasampo.fi
  - 2012 TravelSampo -- Mobile Contextualized Services of Cultural Tourism
  - 2015 WarSampo – Finnish World War II on the Semantic Web
    - » http://www.sotasampo.fi
  - 2018 Semantic National Biography
    - » To be published in September 2018

Aalto University

- Ontology and Data Services
  - 2009 National Ontology Library Service ONKI
    - » http://onki.fi
  - 2014 ONKI.fi -> Finto.fi of the National Library
    - » http://finto.fi
  - 2014 Linked Data Finland Data Service & Tools
    - » http://ldf.fi
  - 2016 Finnish Ontology Service for Historical Places and Maps
    - » http://hipla.fi
- Publications available at:
  - http://www.seco.tkk.fi/publications/

UNIVERSITY OF HELSINKI

SeCo
semantic computing

Joint Work During 2002-2018

Thanks

Aalto University

- Over 30 researchers at SeCo including Eetu Mäkelä, Tomi Kauppinen, Jouni Tuominen, Kim Viljanen, Tuukka Ruotsalo, Osma Suominen, Matias Frosterus, Suvi Kettula, Kaisa Hypen, Erkki Heino, Petri Leskinen, Minna Tamper, Esko Ikkala, Mikko Koho, …
- Over 50 organizations involved

UNIVERSITY OF HELSINKI

SeCo
semantic computing

2016-

Aalto University

Developing the Digital World Together

HELDIG

Helsinki Centre for Digital Humanities

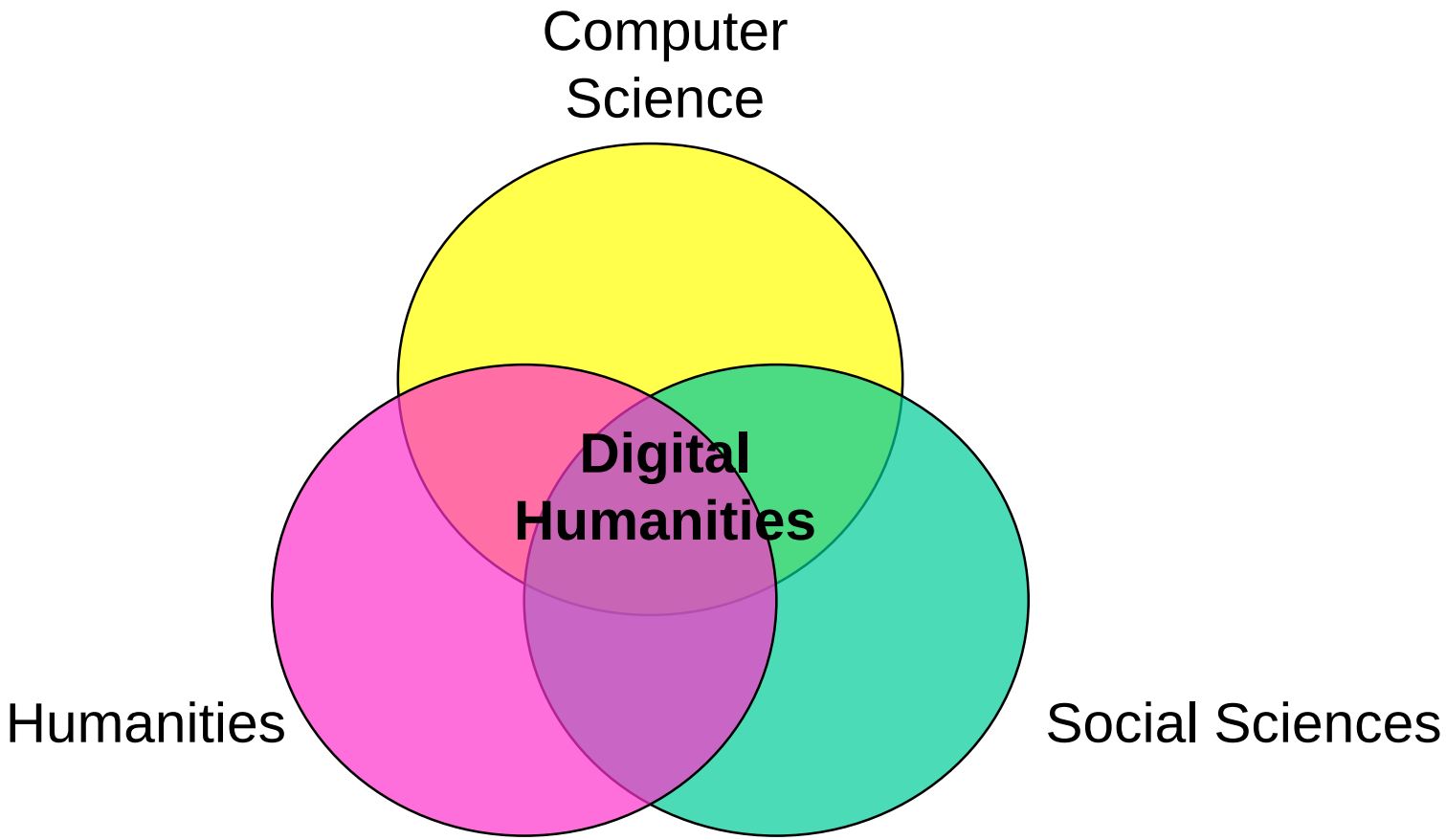
http://heldig.fi

UNIVERSITY OF HELSINKI

SeCo semantic computing

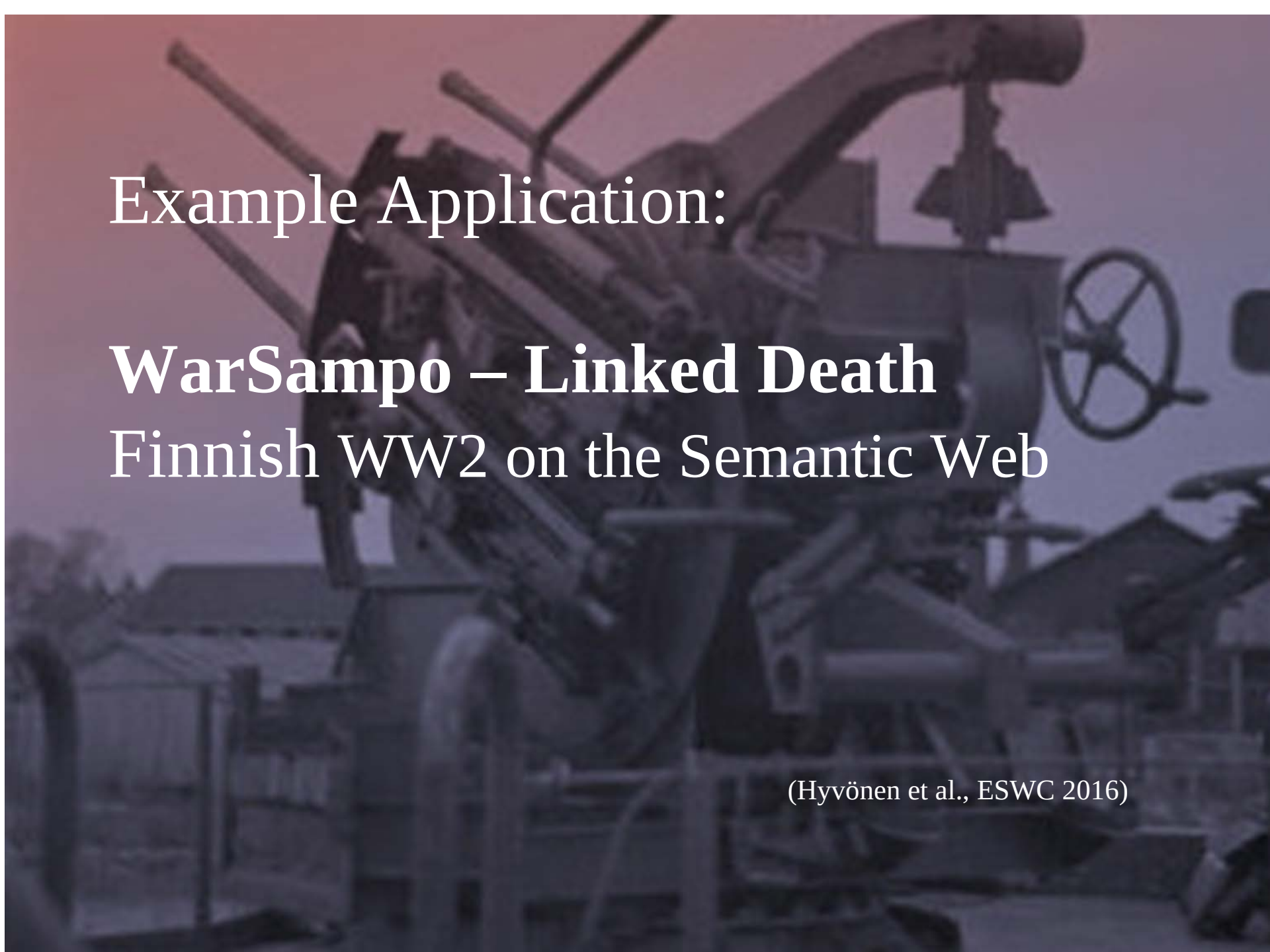
What Is Digital Humanities?

Aalto University



UNIVERSITY OF HELSINKI

SeCo semantic computing



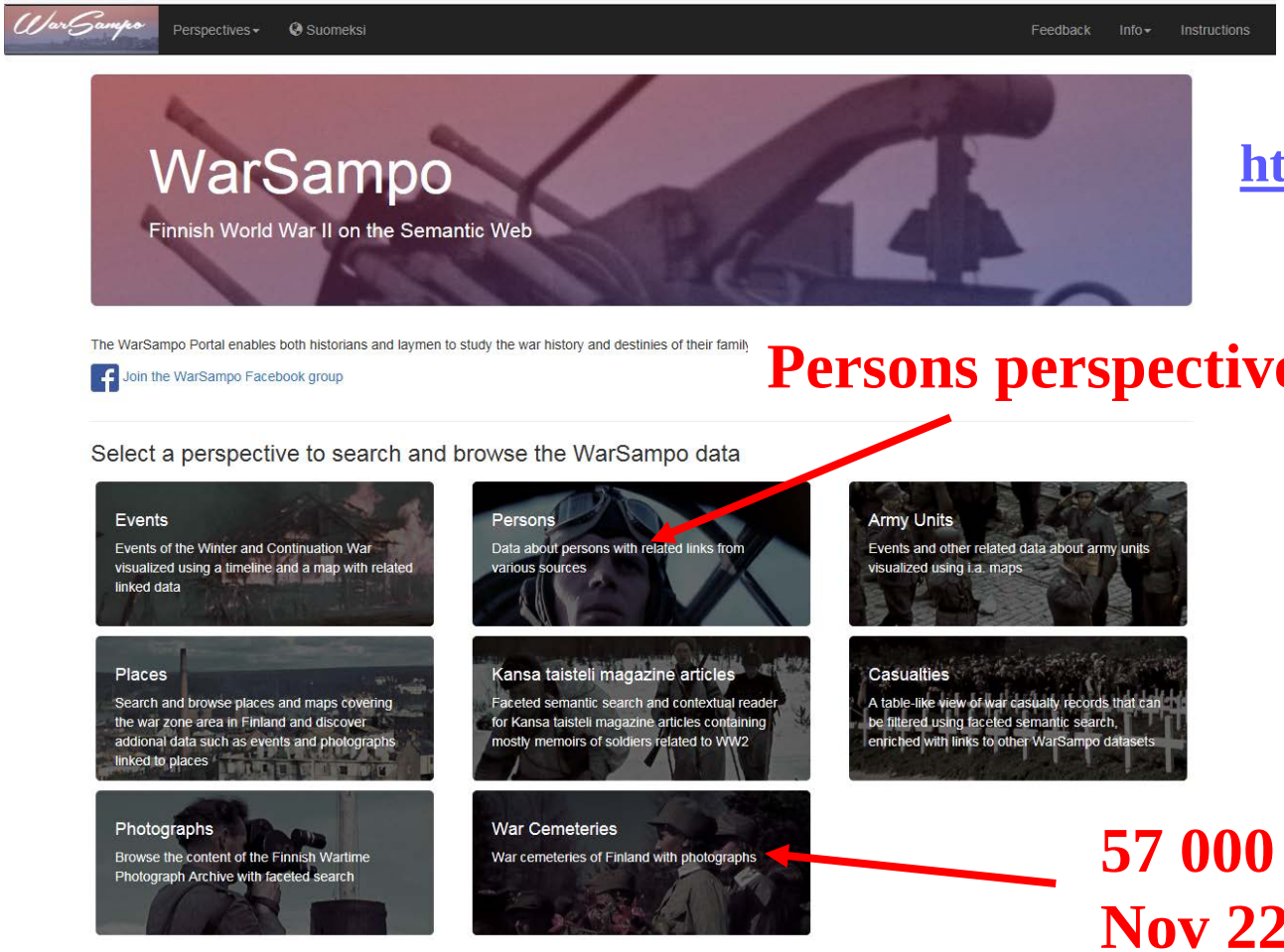
Example Application:

WarSampo – Linked Death
Finnish WW2 on the Semantic Web

(Hyvönen et al., ESWC 2016)

Semantic Portal: 8 Perspectives to WW2

Aalto University



http://sotasampo.fi

Persons perspective

57 000 users in a week
Nov 22, 2017

UNIVERSITY OF HELSINKI

SeCo semantic computing

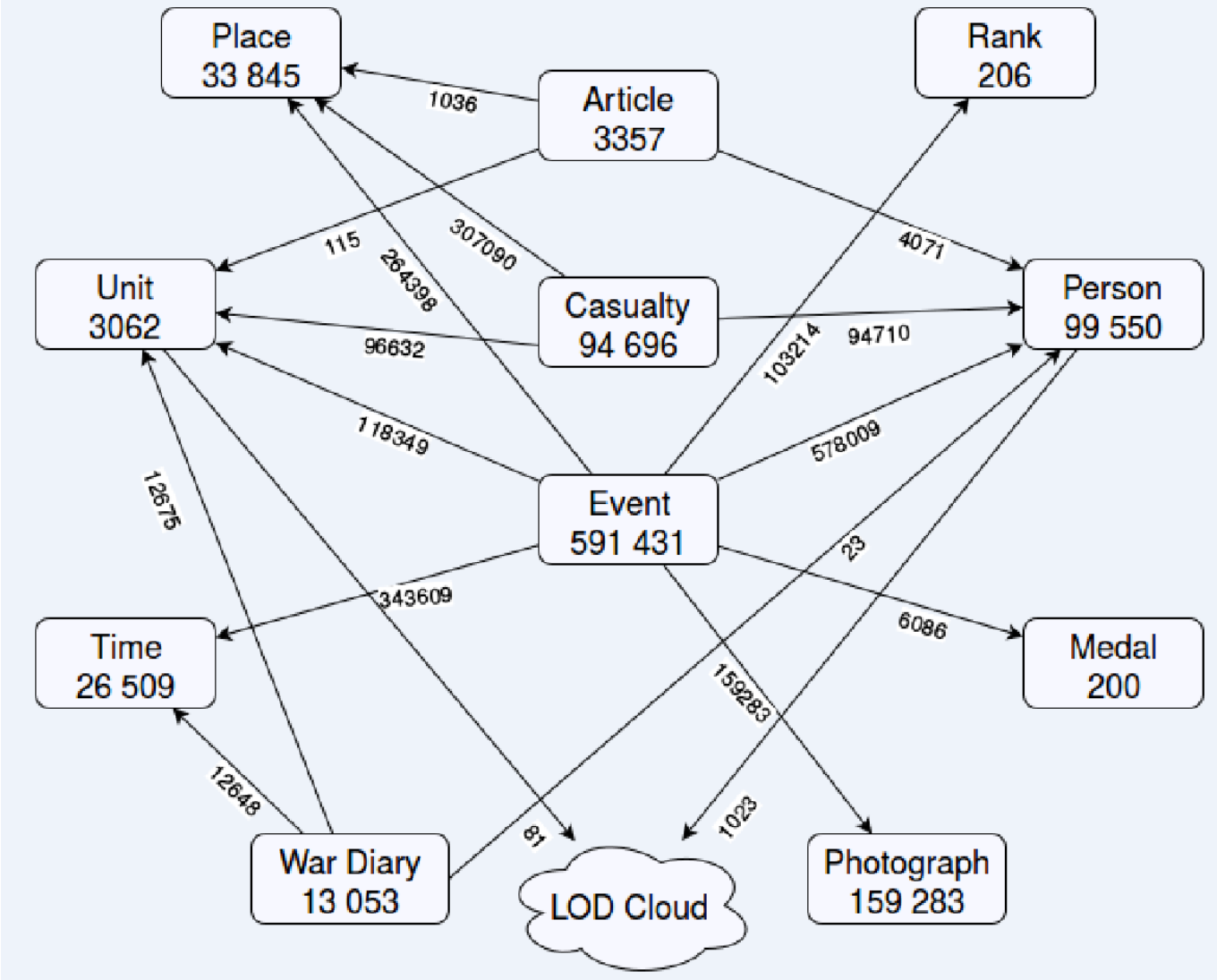
Aalto University

https://vimeo.com/212249404

UNIVERSITY OF HELSINKI

SeCo
semantic computing

WarSampo Linked Open Data Cloud



12M triples knowledge graph

Linked Data Infrastructure for Digital Humanities

Aalto University

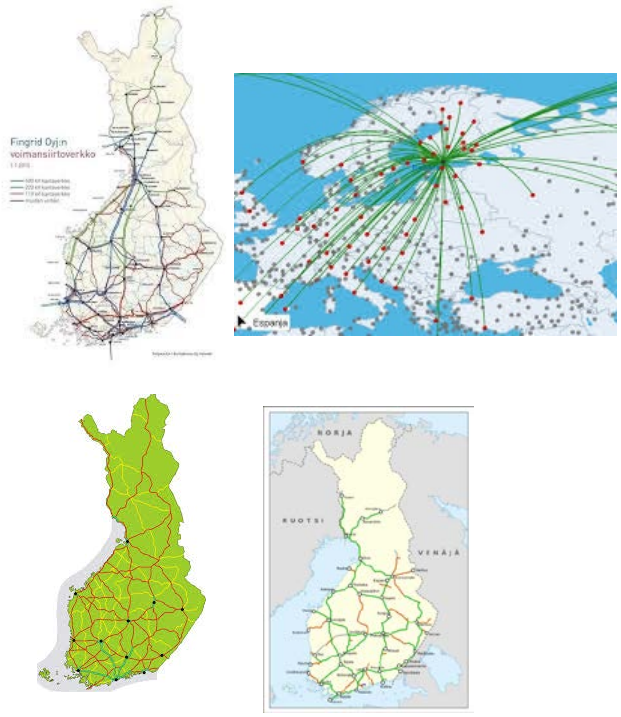
UNIVERSITY OF HELSINKI

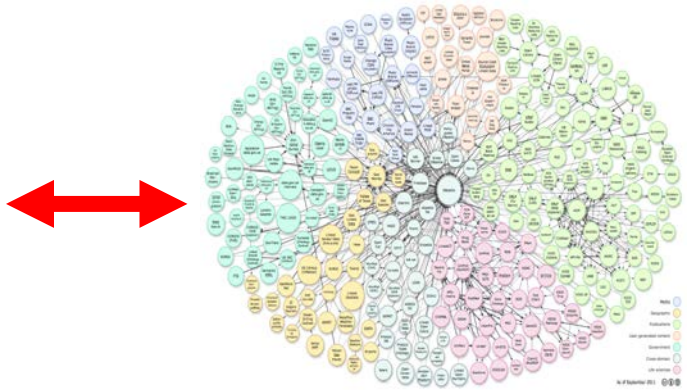
SeCo semantic computing

LODI4DH

Traditional Infras:
(rail)roads, electricity, …

Semantic Content Infra:
Ontologies, metadata, data





DARIAH-EU Digital Research Infrastructure for the Arts and Humanities

Aalto University

Challenges:
Content Complexity & Production

UNIVERSITY OF HELSINKI

SeCo
semantic computing

Problem 1: Cultural Content Compexity - Heterogenous and Interlinked

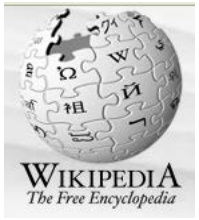
Aalto University

Artifacts

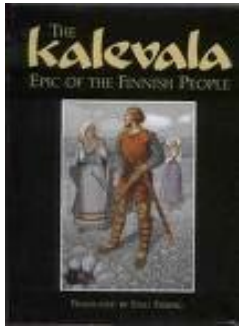
Maps

Videos

Encyclopedia



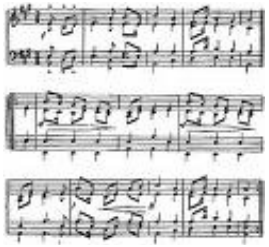
Narratives



Literature

Buildings

Music



Cultural sites

Biographies

Gallen-Kallela, Akseli (1865 - 1931)

taidemaalari

Akseli Gallen-Kallela kuuluu Suo
kuvataiteilijoihin. Ennen kaikkea
suomalaiskansallisen taiteen ur
kehittäjä; kansallisromantiikaks
tyyli oli ennen kaikkea hänen lu
yleisölle Gallen-Kallela tuli tunne
kansallisten merkkiteosten, Kal
veljeksen, kuvittajana. Modernir
enää 1920-luvulla arvostaneet
myöhemmin on ihailtu erityises
monipuolisuuttaan. Hän oli myö
taidekäsityön edelläkävijä, joka

Fine arts

UNIVERSITY OF HELSINKI

SeCo semantic computing

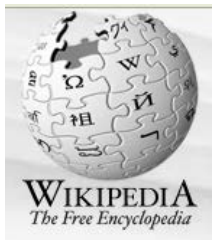
Problem 2: Cultural Content Production System - Distributed and Independent

Aalto University

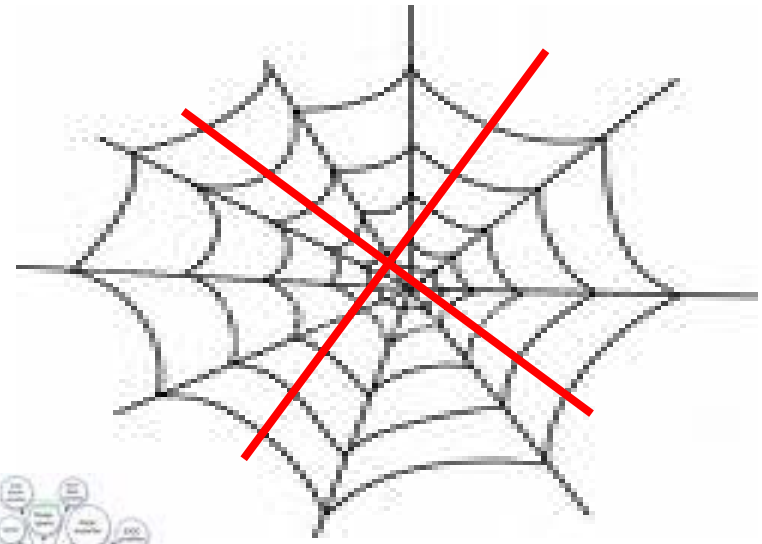
Museums

Land survey

Web 2.0 sites



Archieves



Media

Linked Data

Libraries

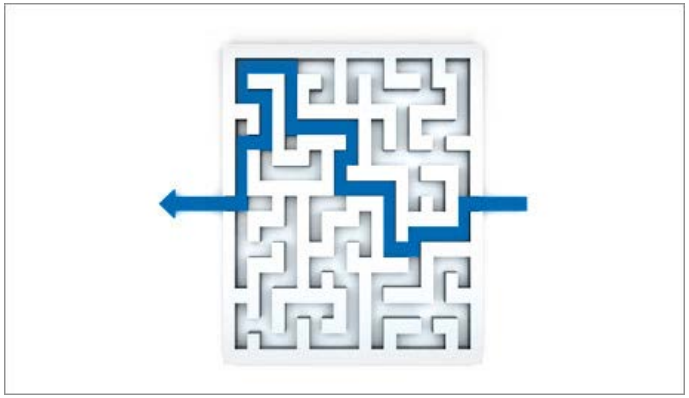
Citizens

UNIVERSITY OF HELSINKI

SeCo semantic computing

Aalto University

Solution:
Sampo Model



UNIVERSITY OF HELSINKI

SeCo
semantic computing

”Sampo” Model for Semantic Interoperabiity

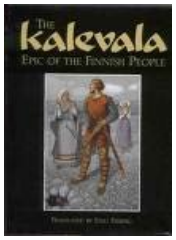
Land survey

Museums

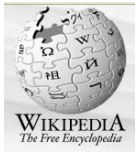
Content Providers

Semantic Metadata

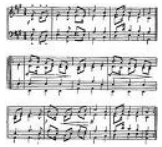
Web 2.0 sites



Linked Data Infrastructure



Archieves



Media

Linked Data

Citizens

Libraries

Aalto University

How Does This Work in Practise?

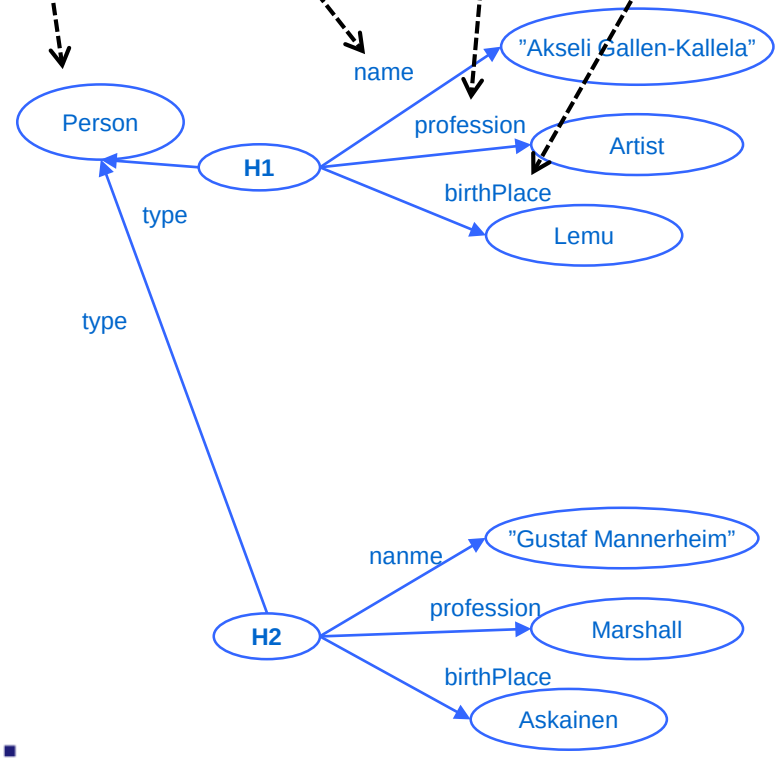
UNIVERSITY OF HELSINKI

SeCo
semantic computing

Biographical Registries Collect Data about Persons

Person Name Profession Birth Place

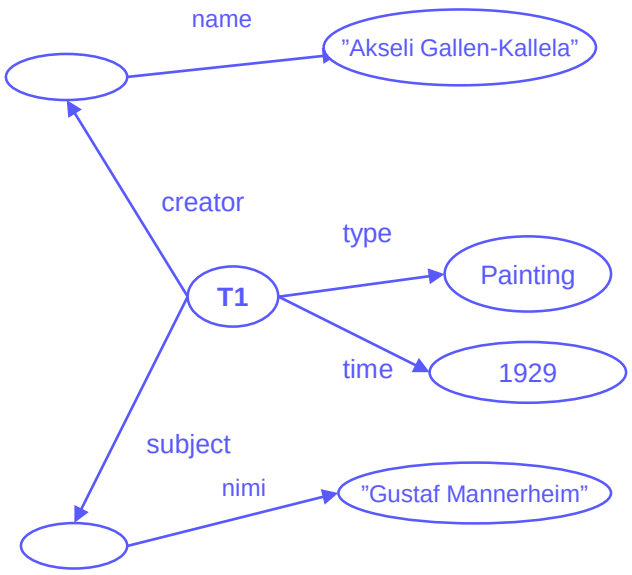
henkilö	nimi	ammatti	syntymapaikka	...
H1	Akseli Gallen-Kallela	taiteilija	Lemu	
H2	Gustaf Mannerheim	marsalkka	Askainen	
	..			



Biography Center

Art Museum Catalogs Paintings

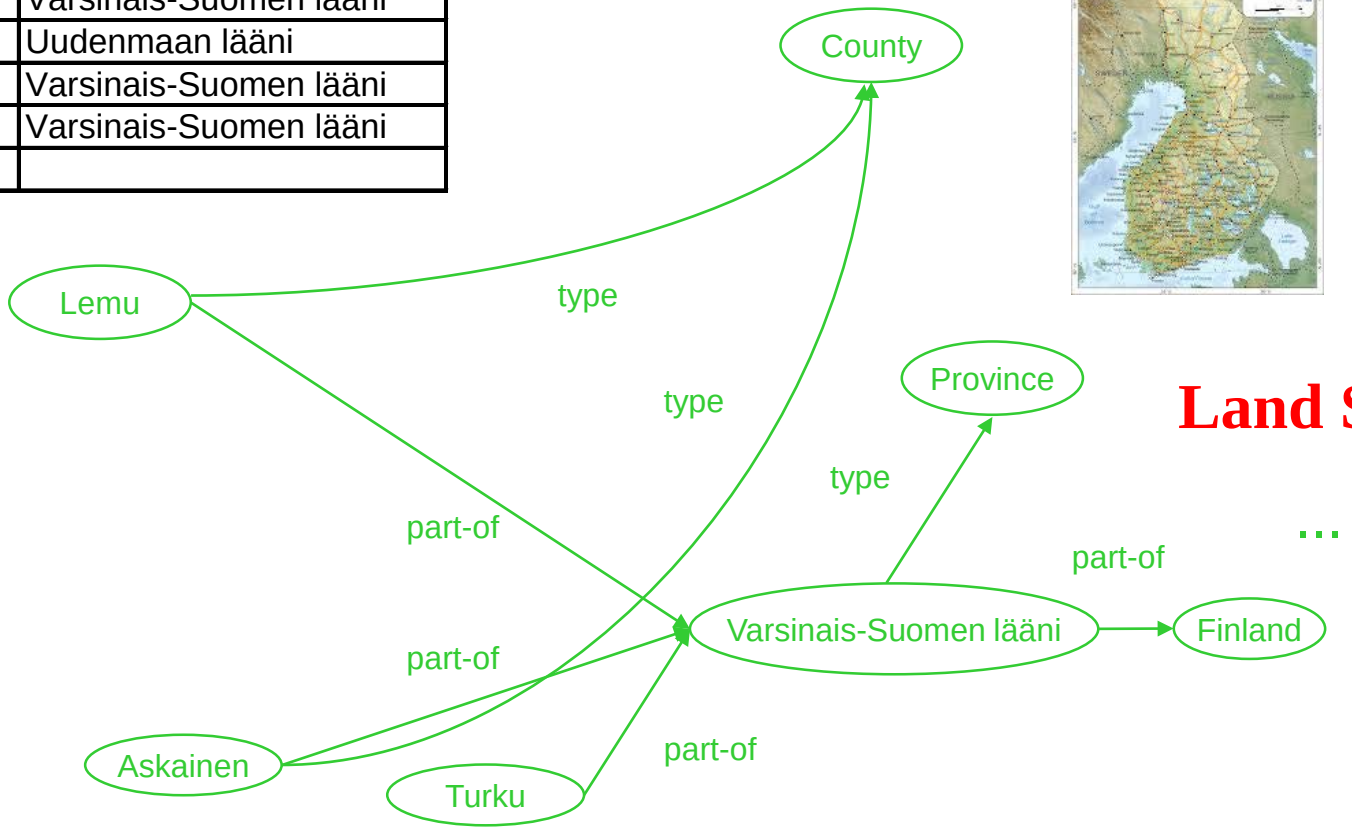
teos	nimi	tekijä	aika	aihe	...
T1	Mannerheimin muotokuva	Akseli Gallen-Kallela	1929	Gustaf Mannerheim	
T2	Aino-triptyykki	Akseli Gallen-Kallela	1891	Aino, Kalevala	
...					



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
...	



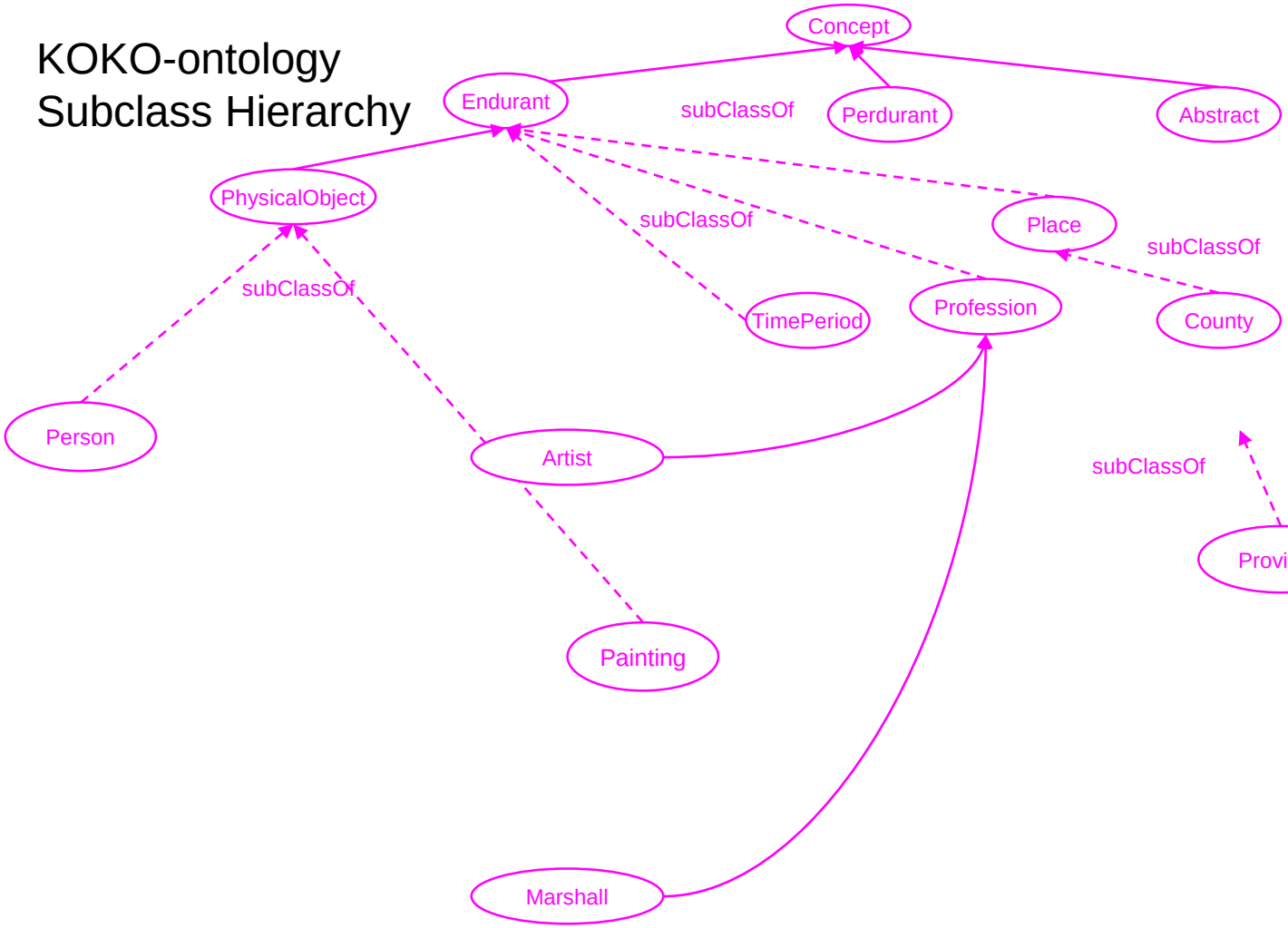
Land Survey

...

Ontologies are Developed by Semantic Web Researchers

Aalto University

KOKO-ontology
Subclass Hierarchy

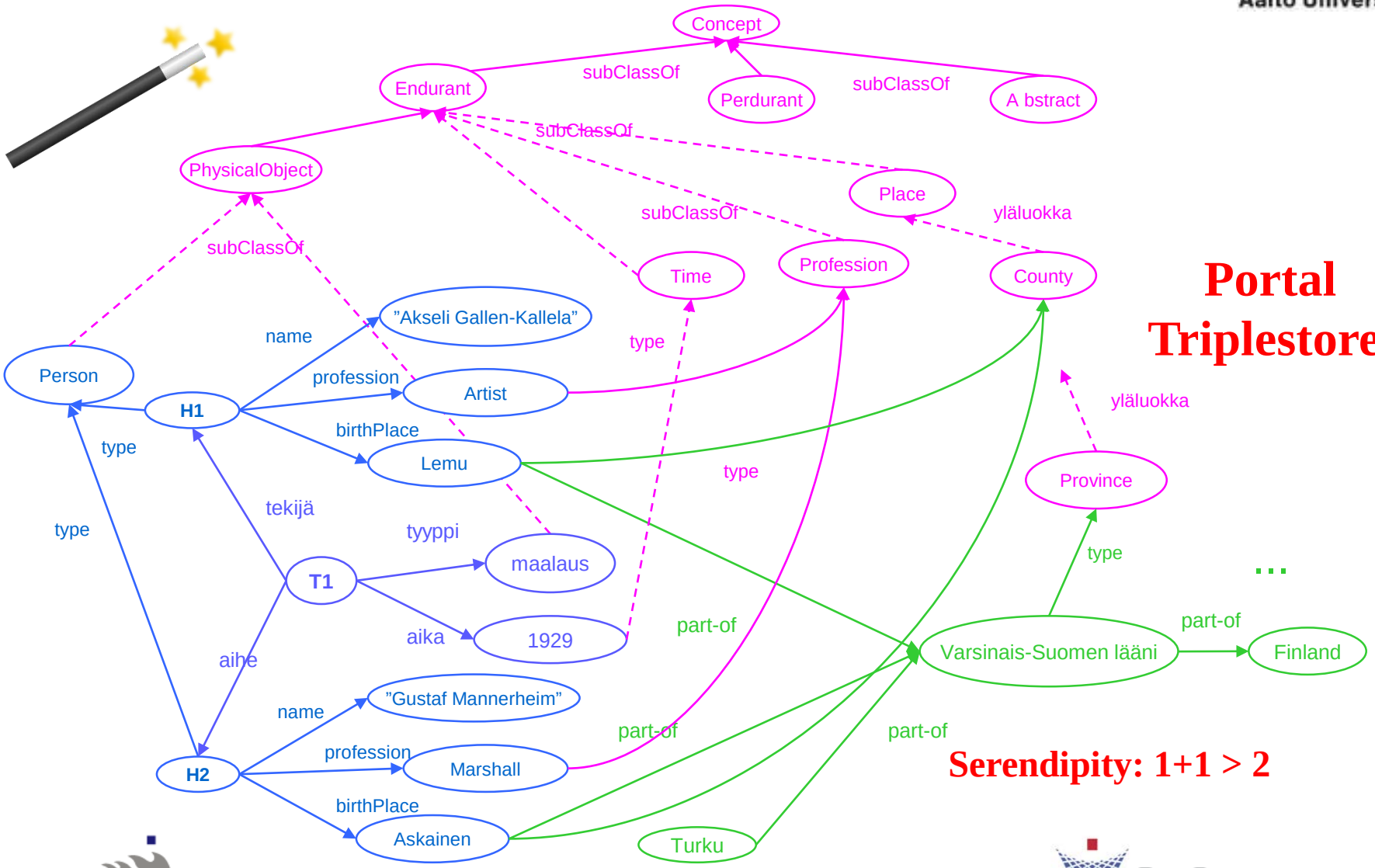


FinnONTO

UNIVERSITY OF HELSINKI

SeCo
semantic computing

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?

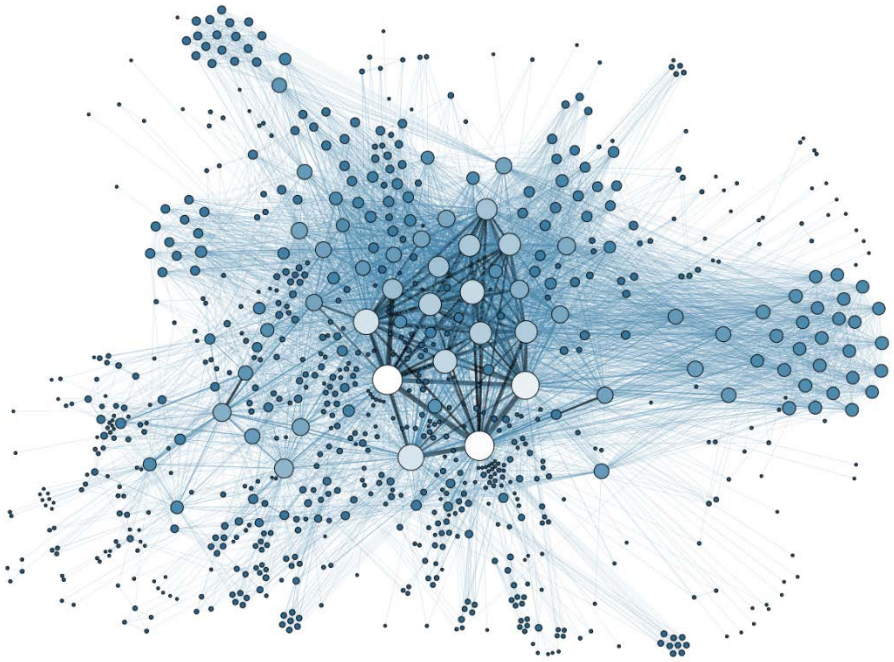
SHARED INFRA NEEDED!

UNIVERSITY OF HELSINKI

SeCo semantic computing

Aalto University

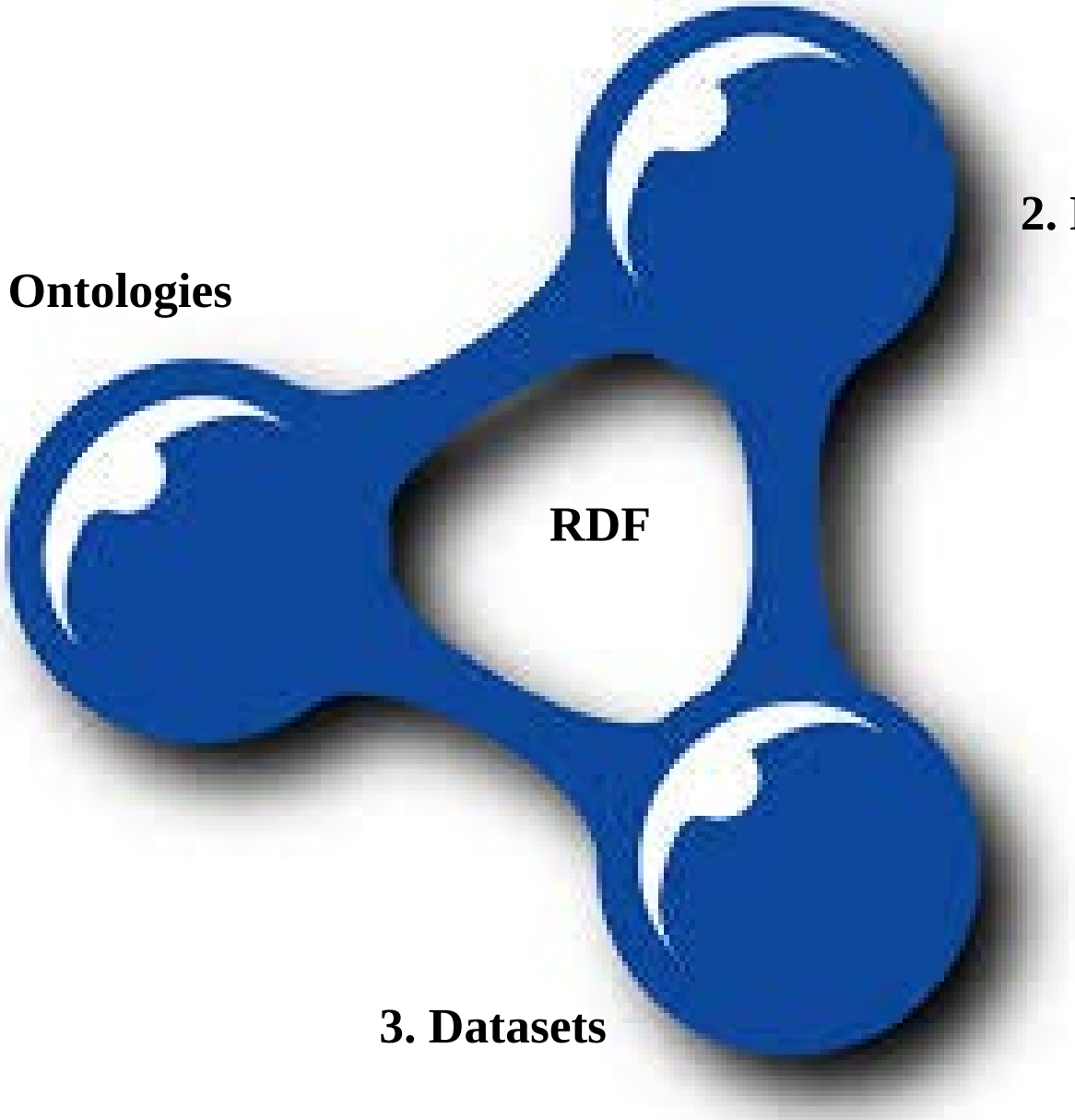
Major Components Needed



UNIVERSITY OF HELSINKI

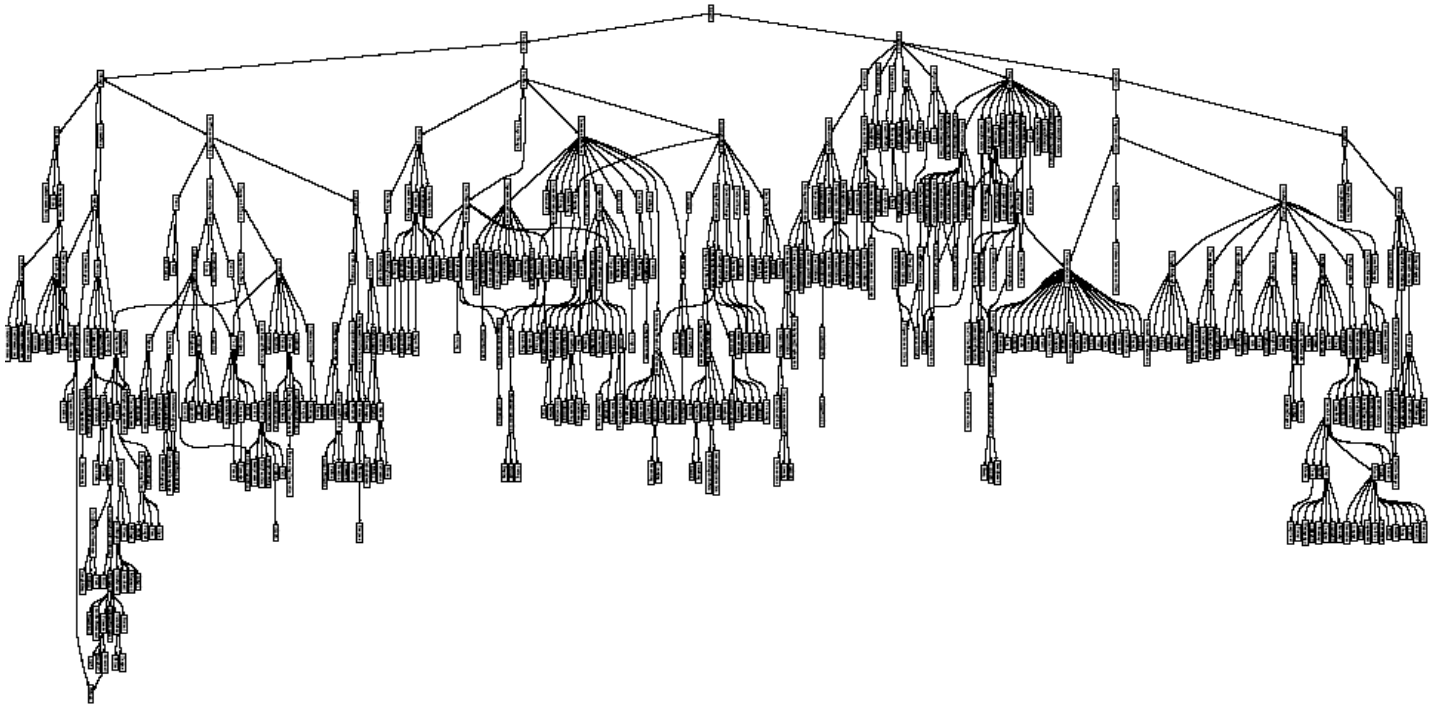
SeCo
semantic computing

Key Components of a Semantic Web Intrastructure



Aalto University

Component 1: Domain Ontologies



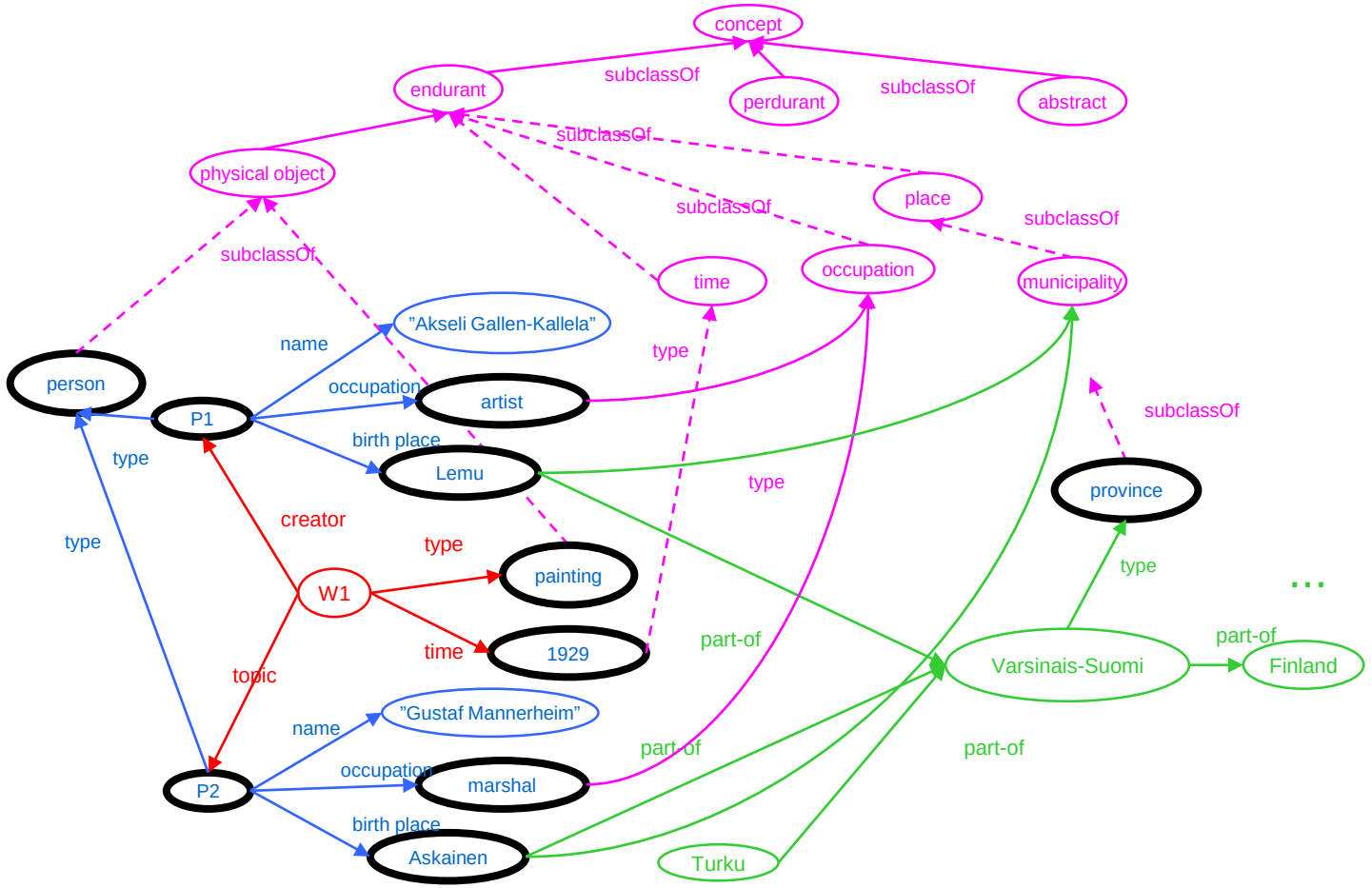
UNIVERSITY OF HELSINKI

SeCo
semantic computing

Concepts (IDs) and Data Model Glue Data Together

Solution: Shared Ontology Infrastructure

Aalto University



UNIVERSITY OF HELSINKI

SeCo semantic computing

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

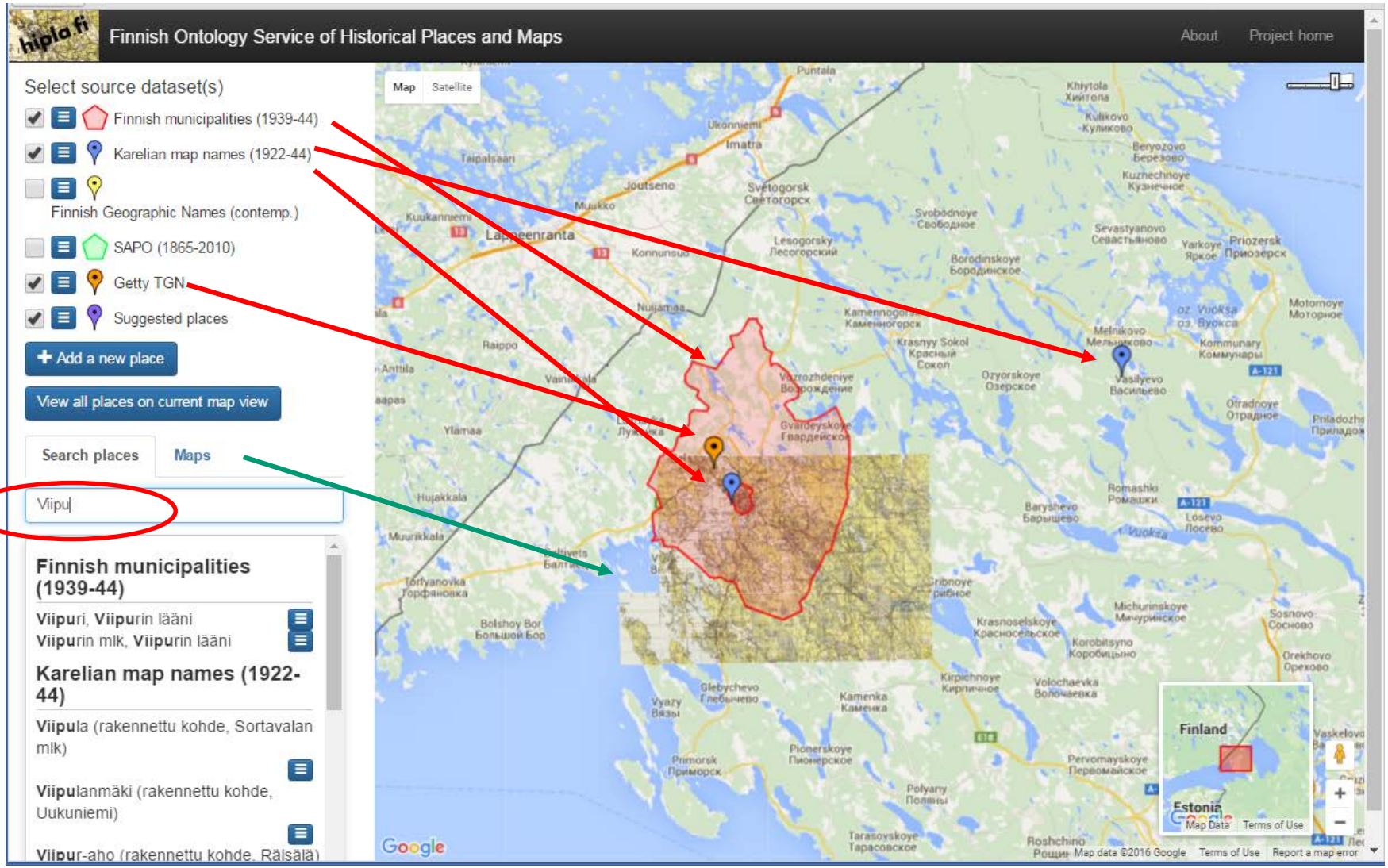
Aalto University

Ontology of Historical Places and Maps

UNIVERSITY OF HELSINKI

SeCo
semantic computing

Hipla.fi: Searching for Place ”Viipu…”



Aalto University

Ontology of Historical Events

”Semantic glue” of everything

UNIVERSITY OF HELSINKI

SeCo
semantic computing

Some Works of SeCo Group

Aalto University

- WW1 history ontology
- Finnish WW2 history in WarSampo
- Finnish History Ontology HISTO
  - Based on the Agricola Network timeline
  - Work continues …

UNIVERSITY OF HELSINKI

SeCo
semantic computing

Example of an Event on HISTO Ontology

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

Ontology of Historical Keywords

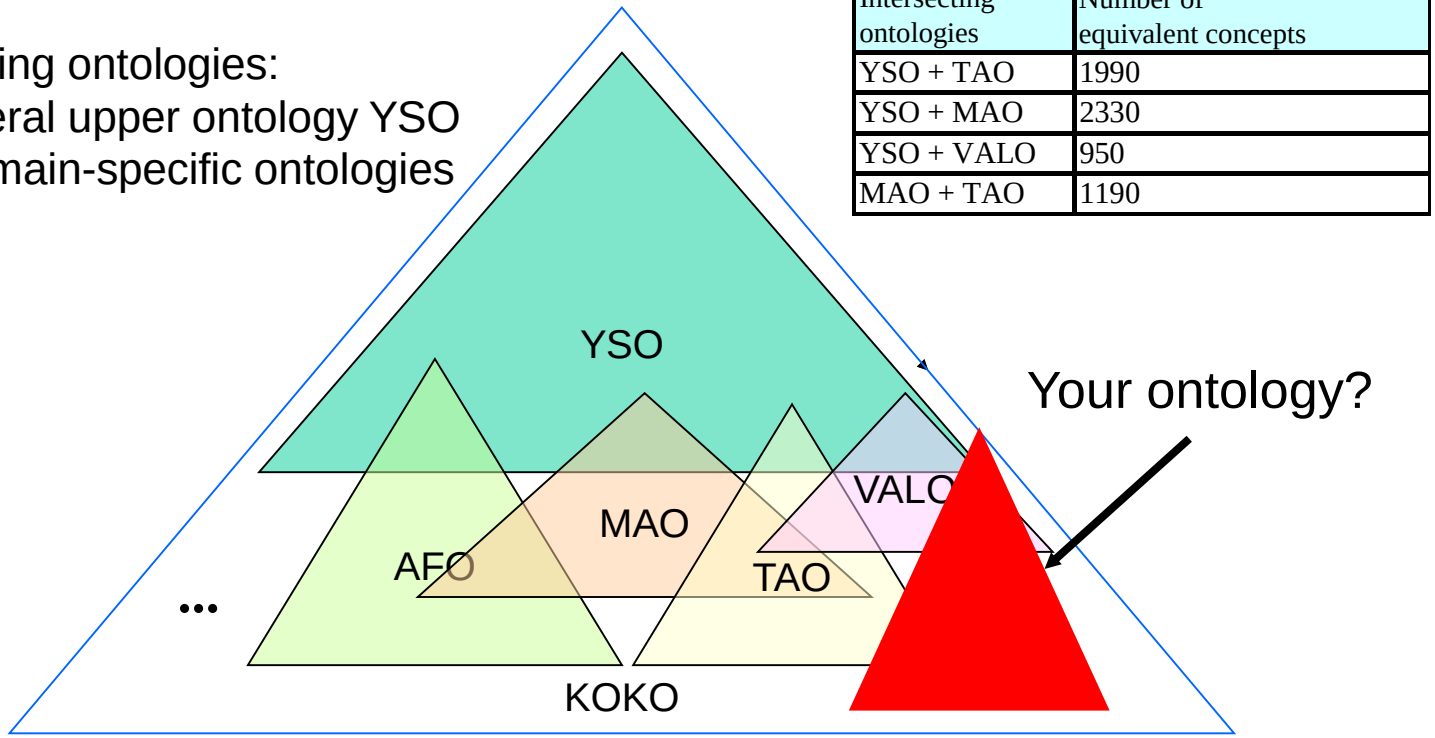
UNIVERSITY OF HELSINKI

SeCo
semantic computing

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]

http://finto.fi

Examples

- Professions
- Archeological concepts
- Object types, materials etc.
- Mappings with international thesauri

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 (= biography)
  - New standards for data harmonization and enriching
    - CIDOC CRM (museums) + IFLA LRM (libraries)
- Process-centric?
  - Next level after events

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

About…

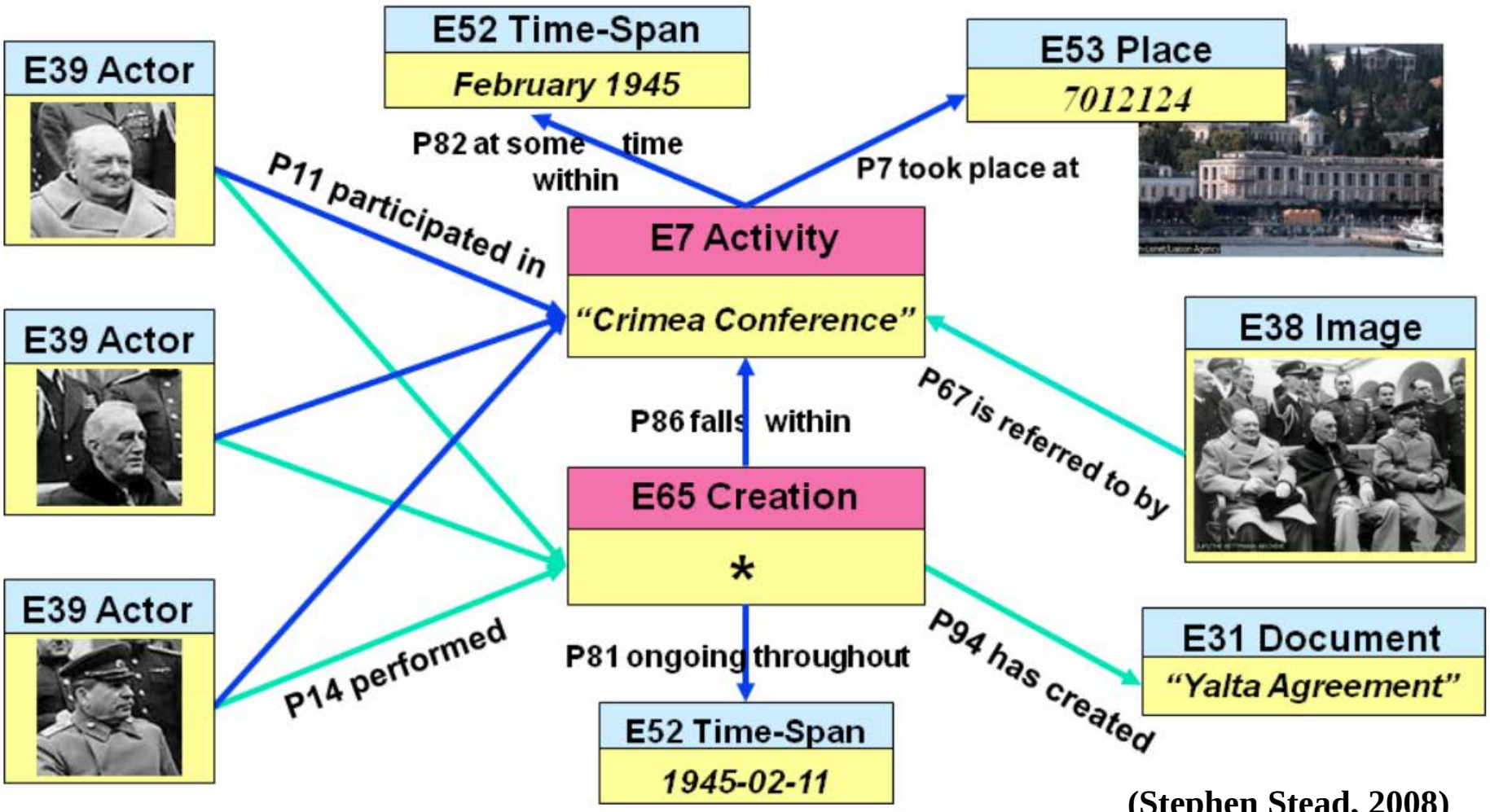
Photos, Persons

[Slide by: Stephen Stead]

Event-centric view: CIDOC CRM

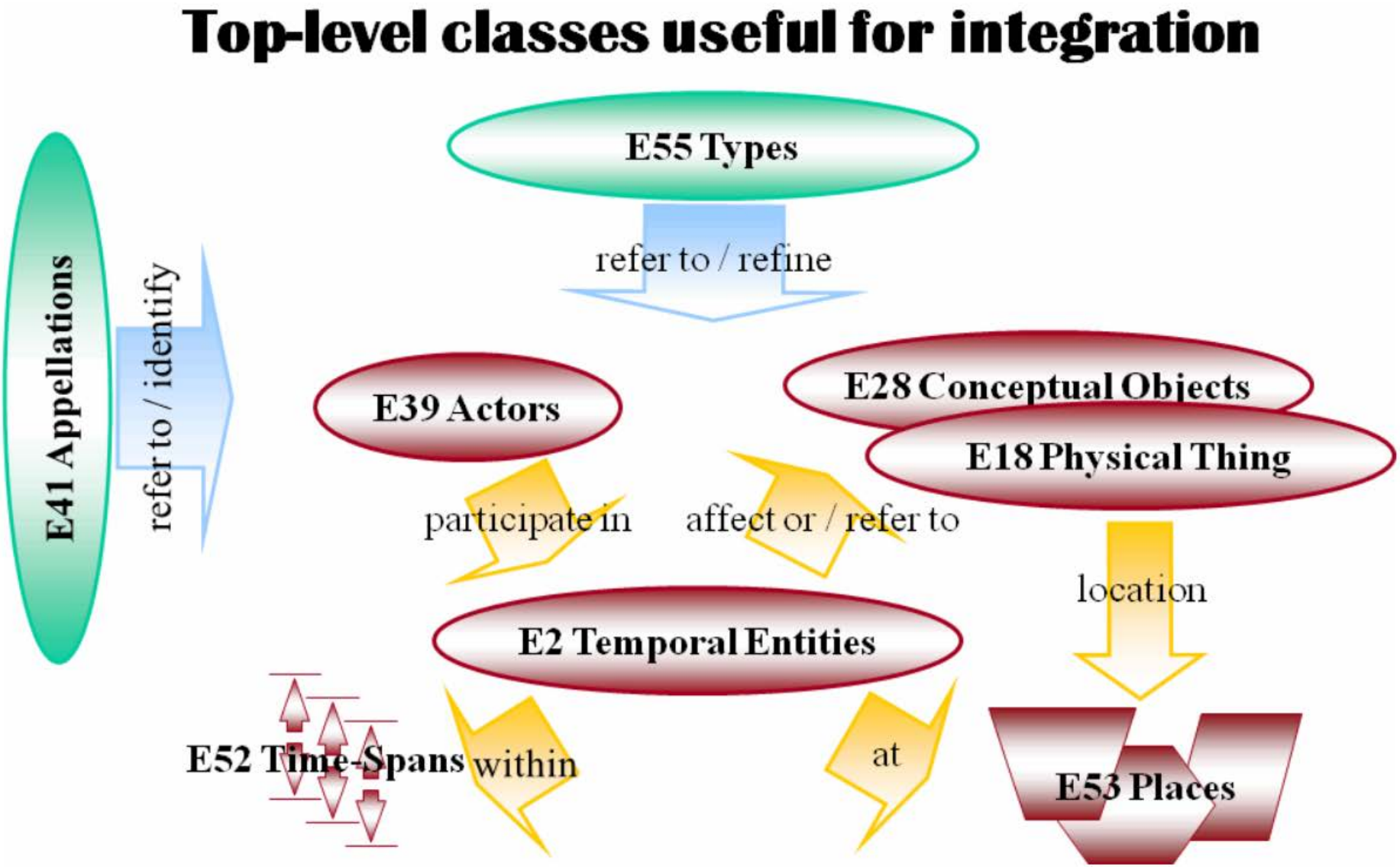
Aalto University

Explicit Events, Object Identity, Symmetry



(Stephen Stead, 2008)

CIDOC CRM: Event Based



UNIVERSITY OF HELSINKI

SeCo semantic computing

(Stephen Stead, 2008)

Aalto University

Component 3: Datasets for the Finnish DH Cloud

UNIVERSITY OF HELSINKI

SeCo
semantic computing

Examples

- Persons
  - Semantic National Biography
- Places
  - Nimiarkisto (Kotus)
  - Map collections (National Archives, Land Survey,…)
- Events
  - Acricola and other timelines of Finnish history
- Museums and Libraries
  - Finnish National Gallery
  - Finna, Fennica
  - BookSampo, CultureSampo
- Media
  - YLE and KAVA achives
- Legislation
  - Semantic Finlex

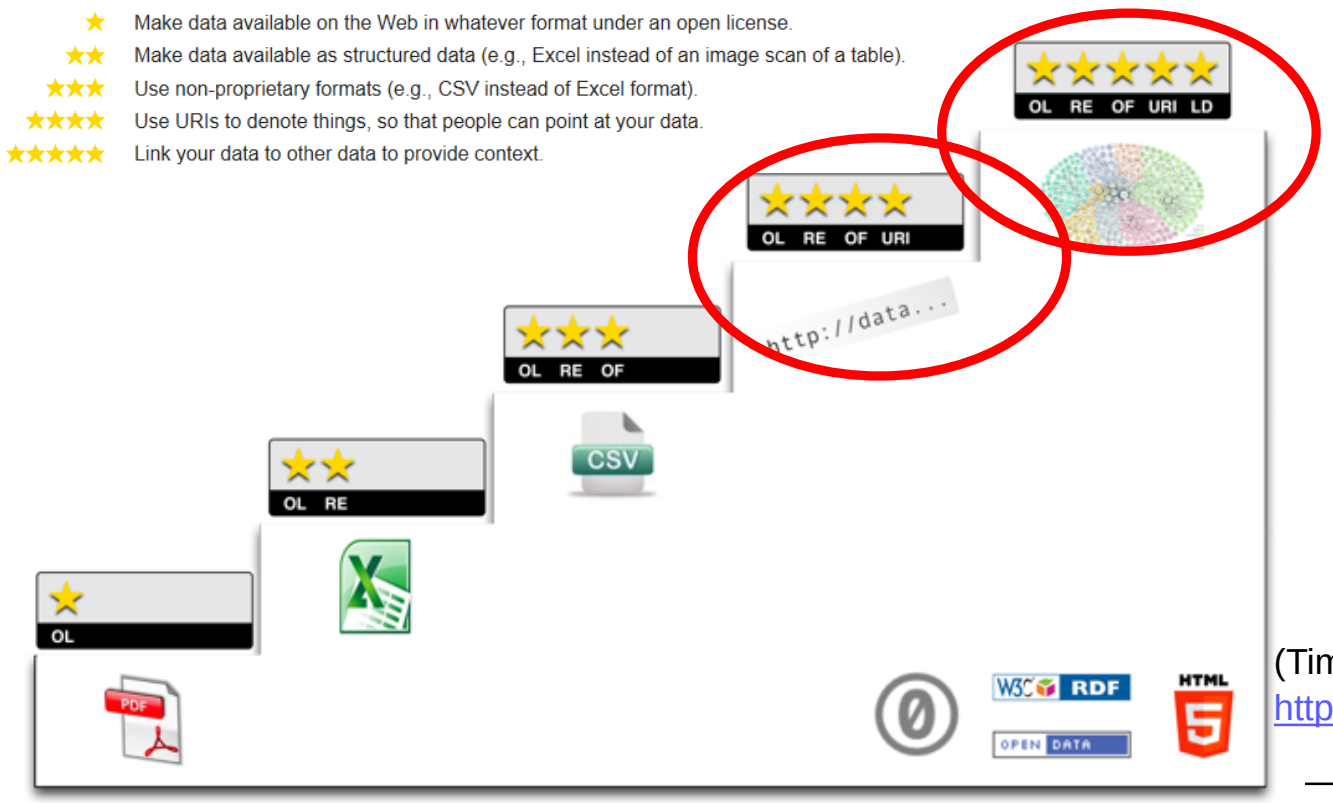
Aalto University

How to Deploy the Infrastructure for the Clients to Use?

UNIVERSITY OF HELSINKI

SeCo
semantic computing

How to publish Linked Data?
5-star Linked Data model



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

Our ”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** (e.g., ONKI http://onki.fi)
- But for **data** and **schemas**

Data Services for

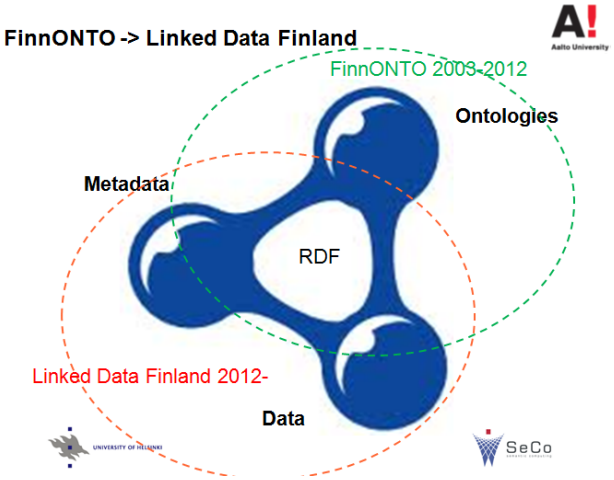
- Linked datasets
- Schemas

Links to

- Related services
- Related applications

Learning Center

- For publishing and using Linked Data



Linked Data Finland Living Lab

- 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.

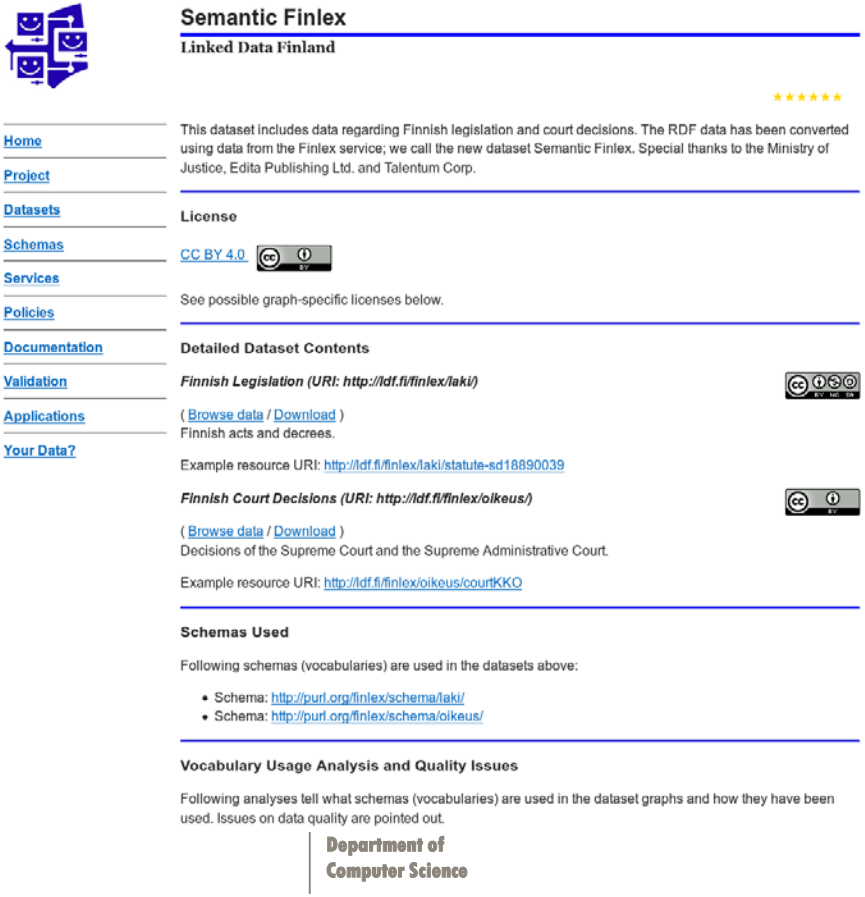
7-star Linked Data Service

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: Finnish Law as Linked Data



Semantic Finlex

Linked Data Finland

★★★★★★

Home
Project
Datasets
Schemas
Services
Policies
Documentation
Validation
Applications
Your Data?

This dataset includes data regarding Finnish legislation and court decisions. The RDF data has been converted using data from the Finlex service; we call the new dataset Semantic Finlex. Special thanks to the Ministry of Justice, Edita Publishing Ltd. and Talentum Corp.

License

CC BY 4.0

See possible graph-specific licenses below.

Detailed Dataset Contents

Finnish Legislation (URI: http://ldf.fi/finlex/laki/)

(Browse data / Download)
Finnish acts and decrees.

Example resource URI: http://ldf.fi/finlex/laki/statute-sd18890039

Finnish Court Decisions (URI: http://ldf.fi/finlex/oikeus/)

(Browse data / Download)
Decisions of the Supreme Court and the Supreme Administrative Court.

Example resource URI: http://ldf.fi/finlex/oikeus/courtKKO

Schemas Used

Following schemas (vocabularies) are used in the datasets above:

- Schema: http://purl.org/finlex/schema/laki/
- Schema: http://purl.org/finlex/schema/oikeus/

Vocabulary Usage Analysis and Quality Issues

Following analyses tell what schemas (vocabularies) are used in the dataset graphs and how they have been used. Issues on data quality are pointed out.

Department of Computer Science

Aalto University
School of Science
UNIVERSITY OF HELSINKI

SeCo
semantic computing

SeCo

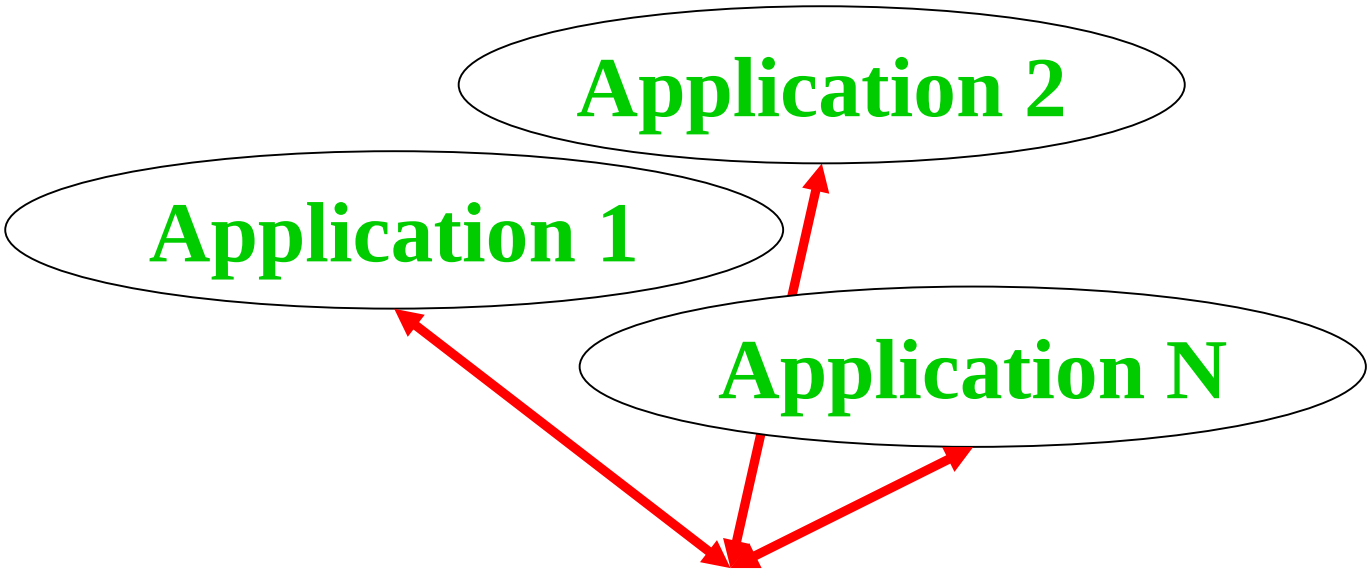
Additional Services

Aalto University

- 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*

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

Aalto University



Client Side (Browser)

SPARQL End Point

Server Side

Linked Data Service

WWW Standard Model

UNIVERSITY OF HELSINKI

SeCo semantic computing

WarSampo: two components

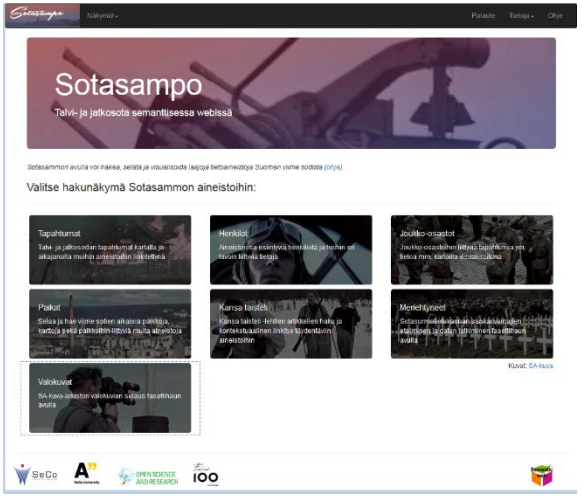
Aalto University

Linked Data Service

http://ldf.fi

Applications

http://sotasampo.fi



UNIVERSITY OF HELSINKI

SeCo semantic computing

Aalto University

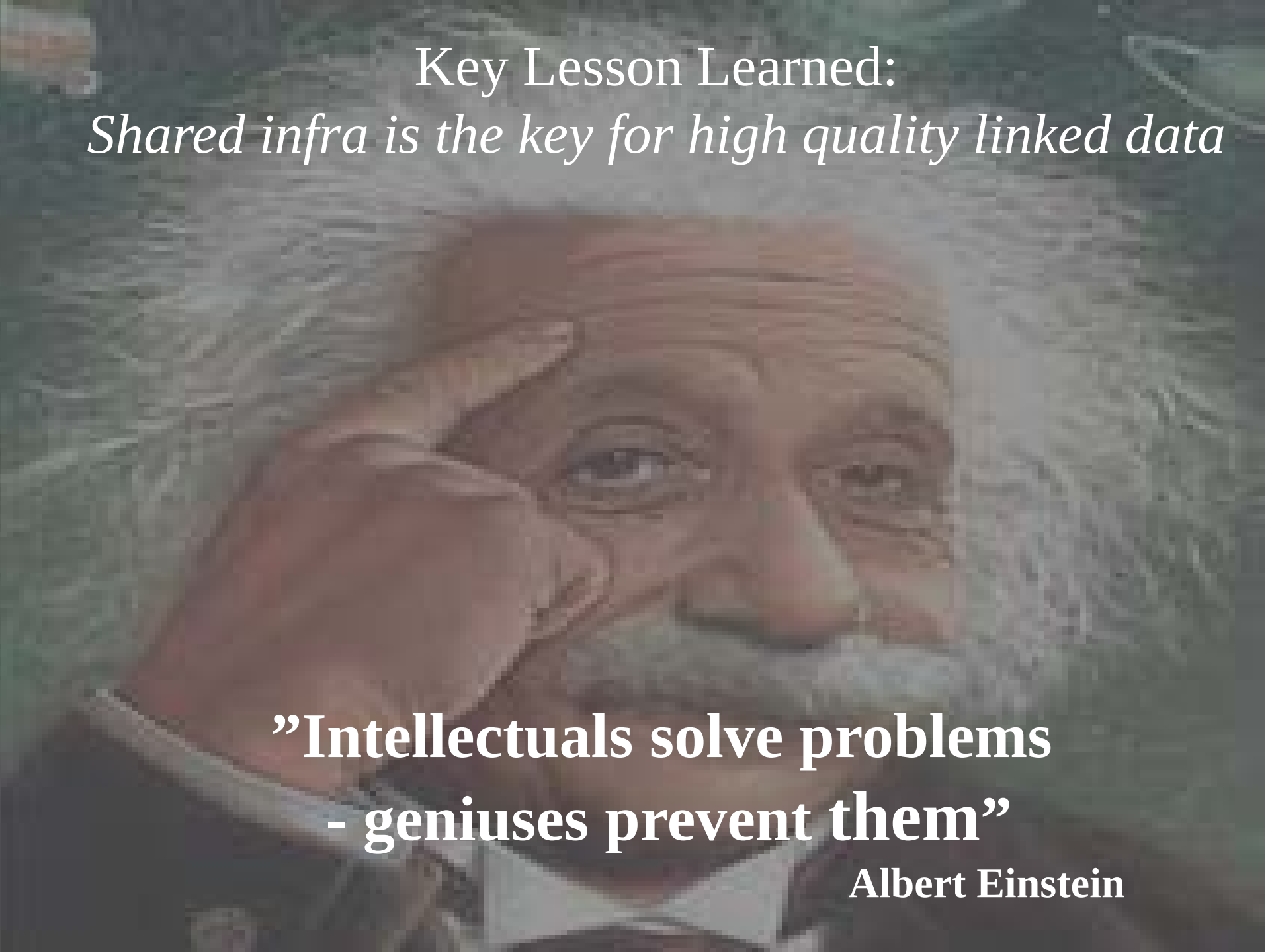
Conclusions

UNIVERSITY OF HELSINKI

SeCo
semantic computing

Key Lesson Learned:

Shared infra is the key for high quality linked data

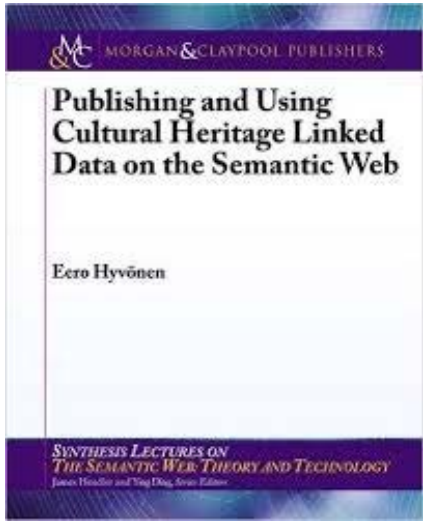


”Intellectuals solve problems - geniuses prevent them”

Albert Einstein

More Info – Questions?

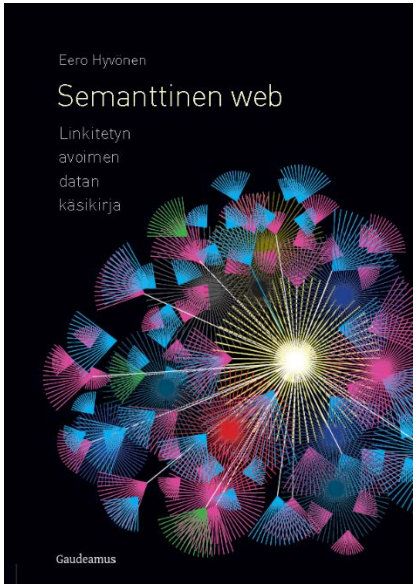
LODI4DH
https://seco.cs.aalto.fi/projects/lodi4dh/
Semantic Web & Linked Data
http://www.w3.org/standards/semanticweb/
Sampo Model & Applications
http://seco.cs.aalto.fi/publications



In English

2012

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



In Finnish

2018

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