

# Linked Open Data Infrastructure for Digital Humanities (LODI4DH)

**HELDIG Digital Humanities Summit 2020:  
Results of Today – Visions for Tomorrow, 11.12.2020**

*Jouni Tuominen*

*HELDIG – Helsinki Centre for Digital Humanities, University of Helsinki*

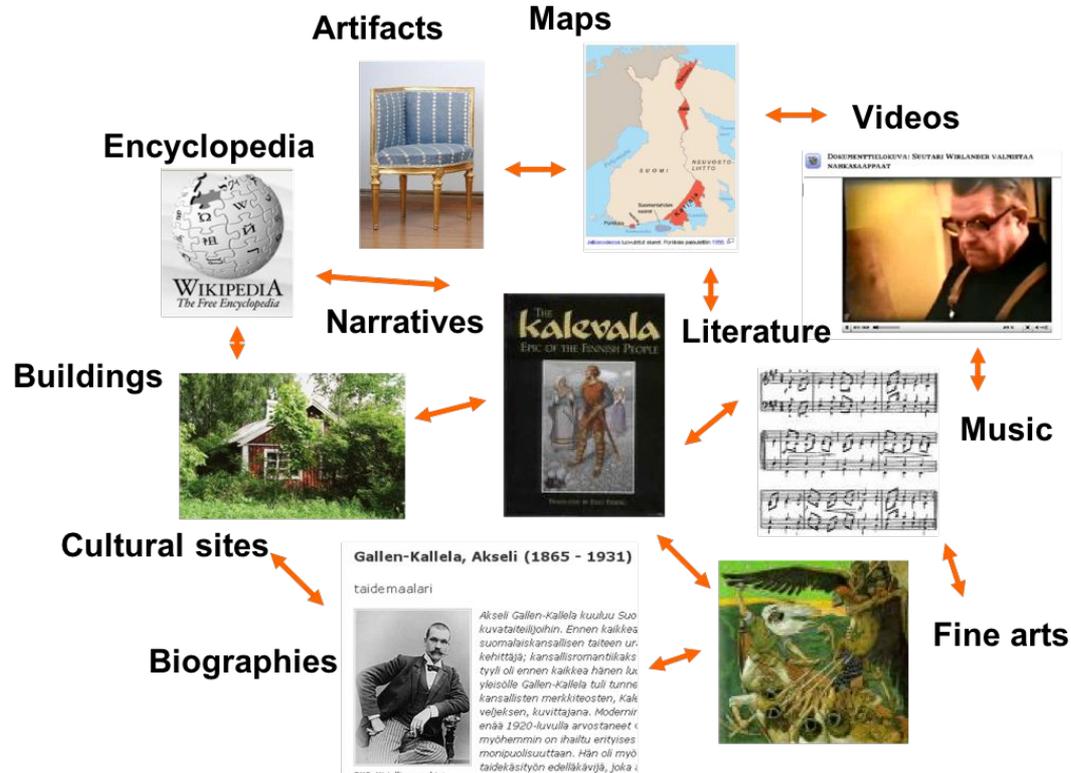
*Semantic Computing Research Group (SeCo), Aalto University*

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

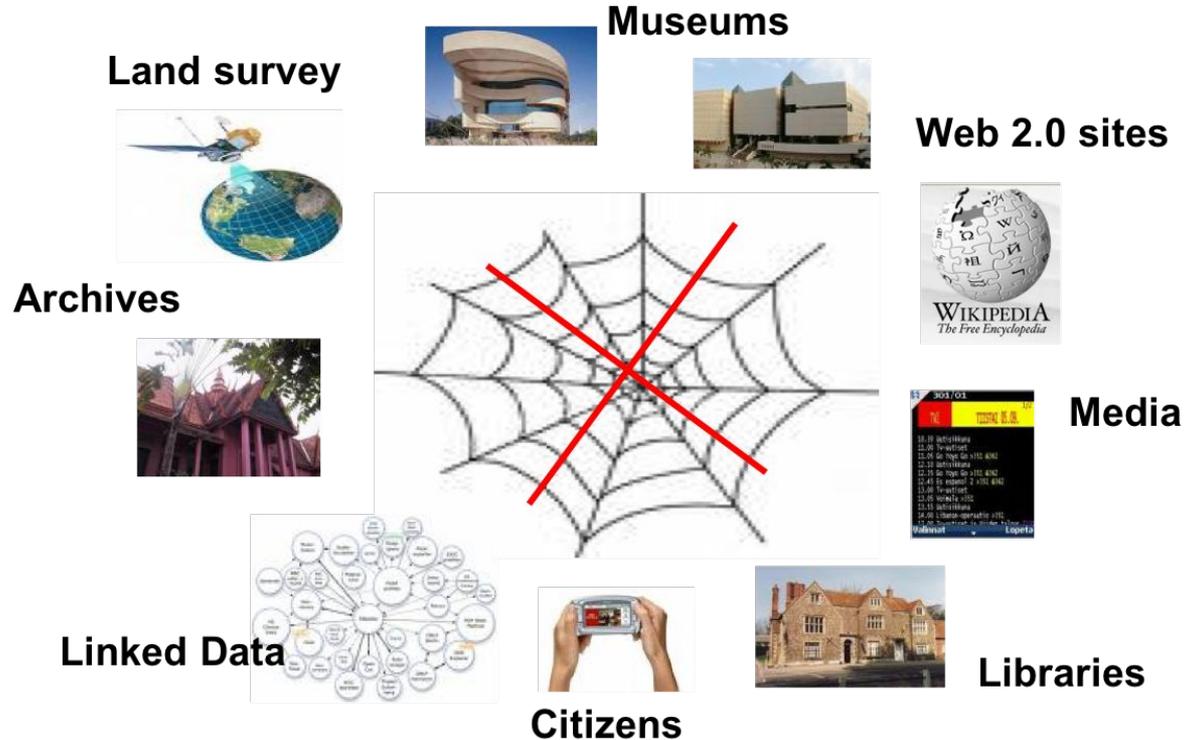
# Outline

- **Setting context: Digital Humanities (DH) & Linked Data**
- **Vision and work in progress**  
**for Linked Data Infrastructure for Digital Humanities**

# Cultural heritage content complexity – heterogeneous and interlinked



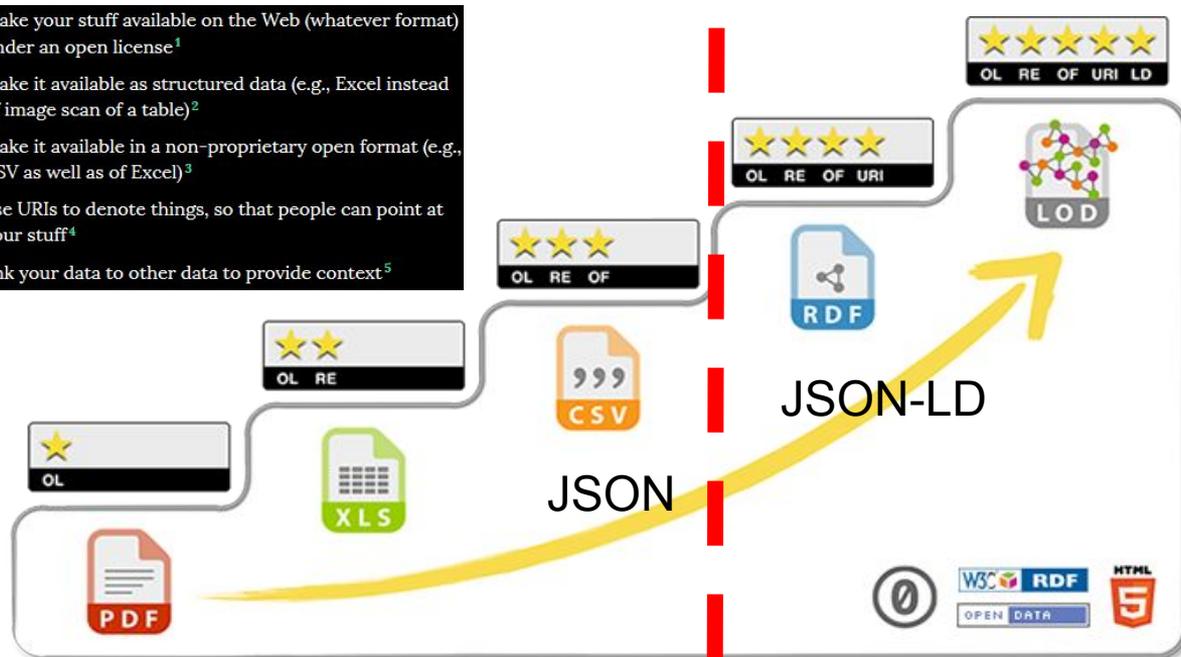
# Cultural heritage content production system – different independent organizations



# How to publish Linked Data?

## 5-star Linked Data model

- ★ make your stuff available on the Web (whatever format) under an open license<sup>1</sup>
- ★★ make it available as structured data (e.g., Excel instead of image scan of a table)<sup>2</sup>
- ★★★ make it available in a non-proprietary open format (e.g., CSV as well as of Excel)<sup>3</sup>
- ★★★★ use URIs to denote things, so that people can point at your stuff<sup>4</sup>
- ★★★★★ link your data to other data to provide context<sup>5</sup>



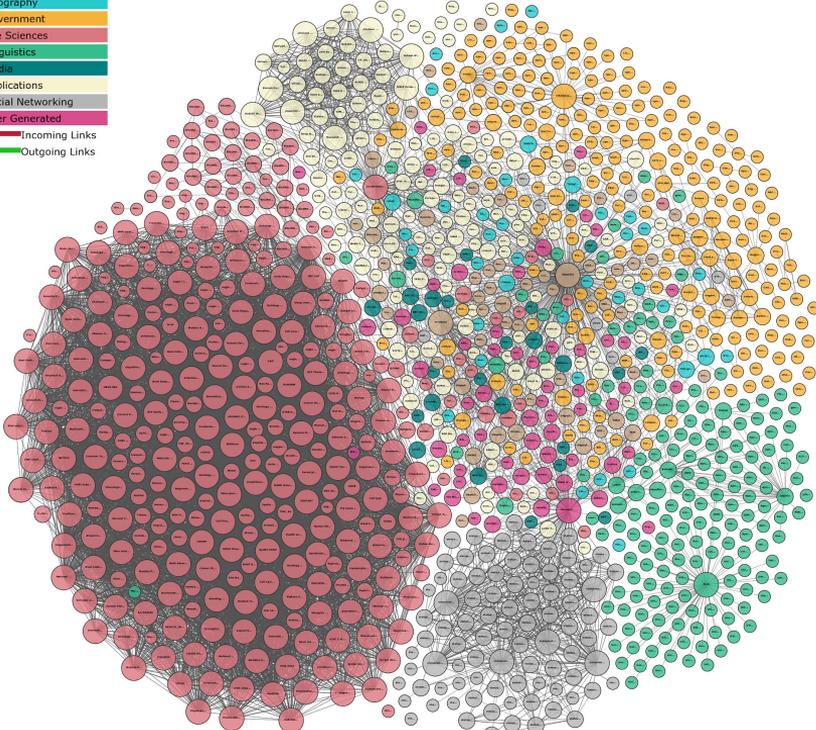
“Traditional” Open Data

Linked Open Data

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



- Legend
- Cross Domain
  - Geography
  - Government
  - Life Sciences
  - Linguistics
  - Media
  - Publications
  - Social Networking
  - User Generated
  - Incoming Links
  - Outgoing Links



Linking Open Data cloud diagram 2017, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cyganiak. <http://lod-cloud.net/>

Jery 1 x    Query 1 x    +

http://ldf.fi/warsa/sparql

```
# Events, photographs and articles that are situated in Vyborg
SELECT DISTINCT ?type ?resource ?prefLabel
WHERE {
  {
    # Events
    BIND (<http://ldf.fi/warsa/places/municipalities/m_place_614> as ?vyborg)
    ?resource crm:P7_took_place_at ?vyborg ;
      a ?typeURI ;
      crm:P4_has_time-span ?timespan .
    FILTER(?typeURI != wet:Photography)
    ?typeURI rdfs:subClassOf* crm:E5_Event .
  }
  UNION
  {

```

Table    Response    Pivot Table    Google Chart    Geo      

wing 1 to 50 of 11,147 entries (in 3.195 seconds)

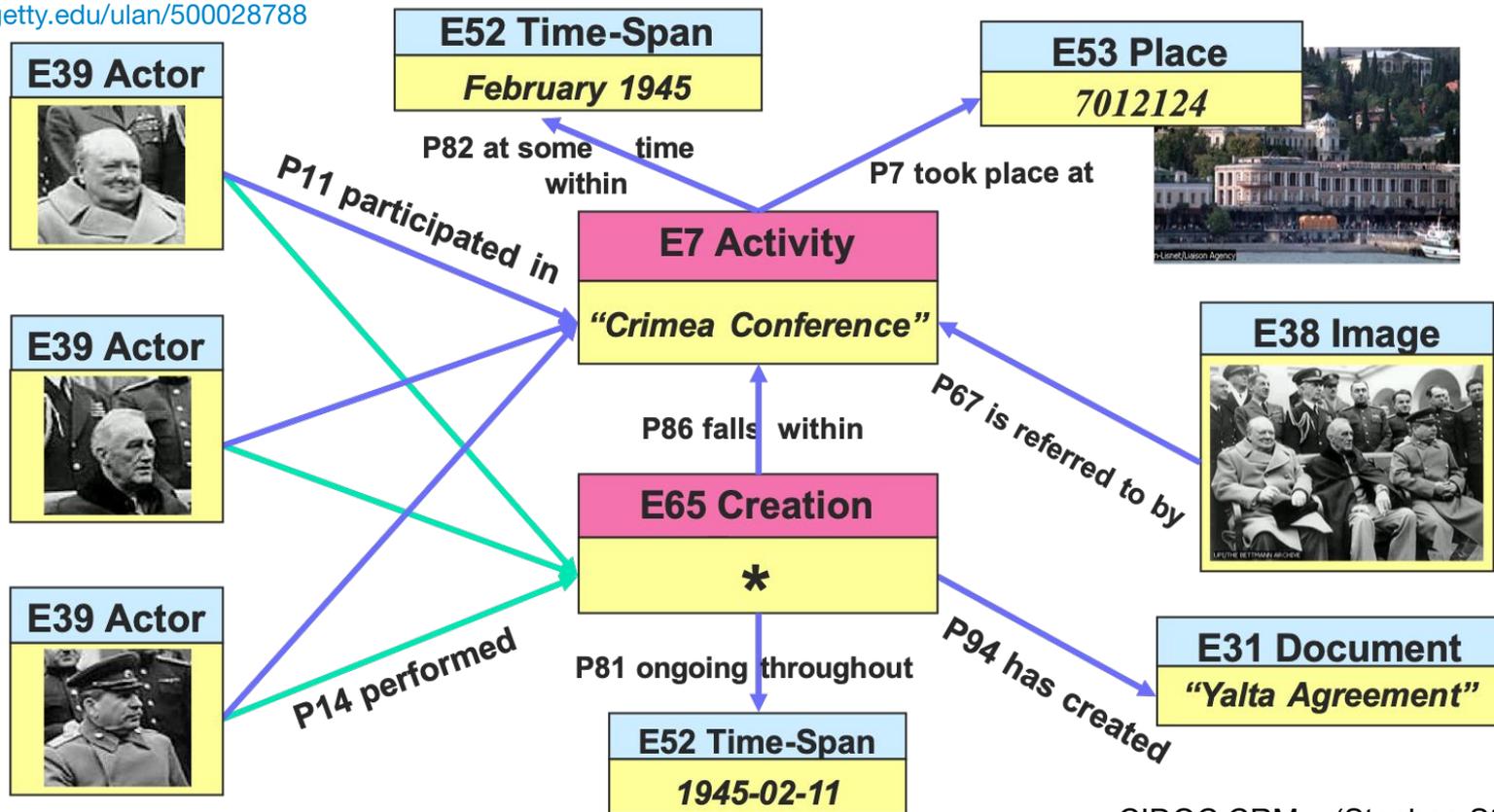
type	resource	prefLabel
"Battle"@en	<a href="http://ldf.fi/warsa/events/event_1078">http://ldf.fi/warsa/events/event_1078</a>	"Defensive battles of Viipuri"@en
"Battle"@en	<a href="http://ldf.fi/warsa/events/event_1080">http://ldf.fi/warsa/events/event_1080</a>	"Viipuri"@en
"Battle"@en	<a href="http://ldf.fi/warsa/events/event_1092">http://ldf.fi/warsa/events/event_1092</a>	"Battle of Vyborg, Pappilanniemi and Kärenmäki sub-blocks"@en
"Battle"@en	<a href="http://ldf.fi/warsa/events/event_1093">http://ldf.fi/warsa/events/event_1093</a>	"Battle of Vyborg, Pappilanniemi and Huhtiala sub-blocks"@en

# Linked Data for DH: what is needed?

- **Metadata models and ontologies for harmonizing data**
- **Tools for data creation and processing**
  - Ontology services
  - Linguistic (and other) processing tools
  - Linked Data publication services
- **Analysis and visualization tools**

# Metadata models

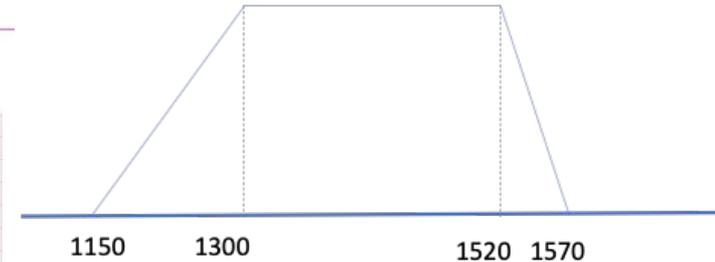
<http://vocab.getty.edu/ulan/500028788>



# Domain ontologies

Times and periods

Middle Age (Keskiaika) in Finland



## Actors

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



- Pjotr Tšaikovski (fi)
- Петр Ильич Чайковский (ru)
- Pyotr Ilyich Tchaikovsky (en)
- Pjotr Tšajkovskij (sv)
- Pjotr Tšajkovskij (no)
- Pjotr Iljitsch Tschaiowski (de)
- Pjotr Ilitch Tchaikowski (fr)
- Piotr Ilich Chaikowski (es)
- Pètr Il'ic Cajkovskij (it)
- Pjotr Iljitsj Tšajkovski (nl)
- Pjotr Ilitch Tchaikowski (pt)
- Pjotr Czajkowski (pl)
- Pjotr Iljici Csaikovski (ro)
- Pjotr Iljics Csaikowiczj (hu)

## Events

SAHA3 | historia - search   
<http://agricola.utu.fi/rd/phe863>

historiallinen tapahtuma: YYA-sopimus raukesi

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="#">Rurubulis, 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 53/1992</a>
tapahtuma-aika	1992-01-20
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>
liitty	<a href="#">YYA-sopimus solmittiin Neuvostoliiton kanssa</a> , <a href="#">YYA-sopimusta jatkettiin kahdeksanneksi vuodelle</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>

## Keyword concepts



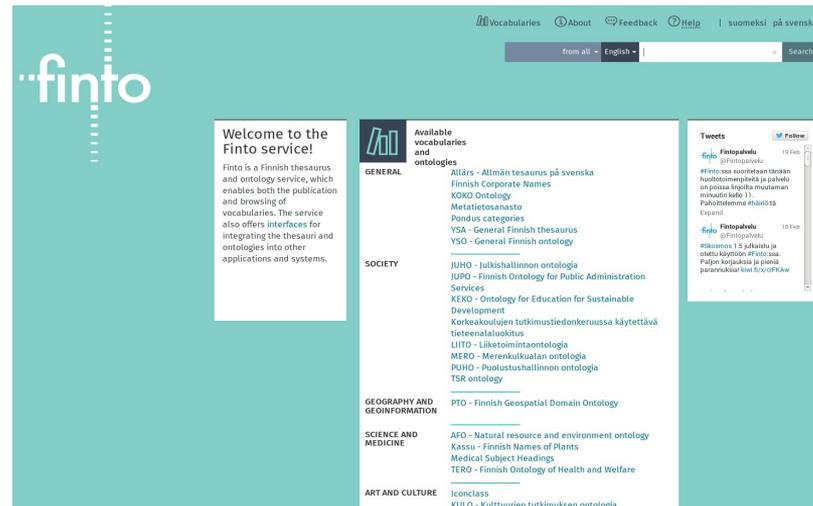
## Places

# Ontology services



## Available vocabularies and ontologies

- General**
  - General Finnish thesaurus
  - General Finnish upper ontology
  - KOKO Ontology
  - Allmän tesaurus på svenska - Allärs
  - AFO Ontology
  - Ontology of Applied Arts - TAO
  - Ontology for Museum Domain
  - Library of Congress Subject Headings - LCSH
- Health**
  - Finnish Ontology of Health and Welfare
- Culture**
  - 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**
  - Merenkulkualan ontologia - MERO
- Science**
  - Kassu - Kasvien suomenkieliset nimet



**Finto / 2019:**

**280 000 users (returning visits)**

**1,9 million page loads**

**32 million requests on programmatic application interface (API)**

# Linguistic processing tools



We provide the following JSON-returning Web Services:

- [Language Recognition](#)
- [Lemmatization](#)
- [Morphological Analysis](#)
- [Inflected Form Generation](#)
- [Hyphenation](#)

The tools backing these services are mostly not originally our own, but we've wrapped. For specifics, see the details of each service. For general questions about this service,

SeCo Text Annotation Service is a set of different tools that can be used for Finnish text and language analysis. The tools exchange text to JSON that contains results of each tool.

## Select a tool of choice

### FINER

Named entity recognition based on the FIN-CLARIN's FINER tool.

### FinBERT

TurkuNLP's named entity recognition application based on Google's BERT.

### APPI - Automatic annotation tool

Named entity linking based on voting scheme.

### Finnish dependency parser

Linguistic analysis based on the Turku NLP Finnish-dep-parser

### Date and registry number identifiers

Regular expression based application for identifying registry numbers, dates, and more.

### Person name finder service

Application that can pick up person names from text and offers links and contextual information (gender, lifespan) about each name.

### Gender Identification Service

Application that can determine person's gender based on a name.

# Linked Data publication services



## 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 number 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 do this by publishing our work on the [Linked Data](#) paradigm and stack of standards, which combines an expressive, semantic 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 data, consumers need more support in getting to know and access them, as well as a better grasp of their 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* 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

## WarSampo

### Linked Data Finland



This dataset includes harmonized data of different kinds concerning the Second World War in Finland, separated in different graphs representing events, actors, places, photographs, and other aspects and documentation of the war. To test and demonstrate its usefulness, this data service is in use in the semantic portal [WarSampo](#) explained in more detail in this [project page](#).

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](#)

#### License

[CC BY 4.0](#)



See possible graph-specific licenses below.

#### Detailed Dataset Contents

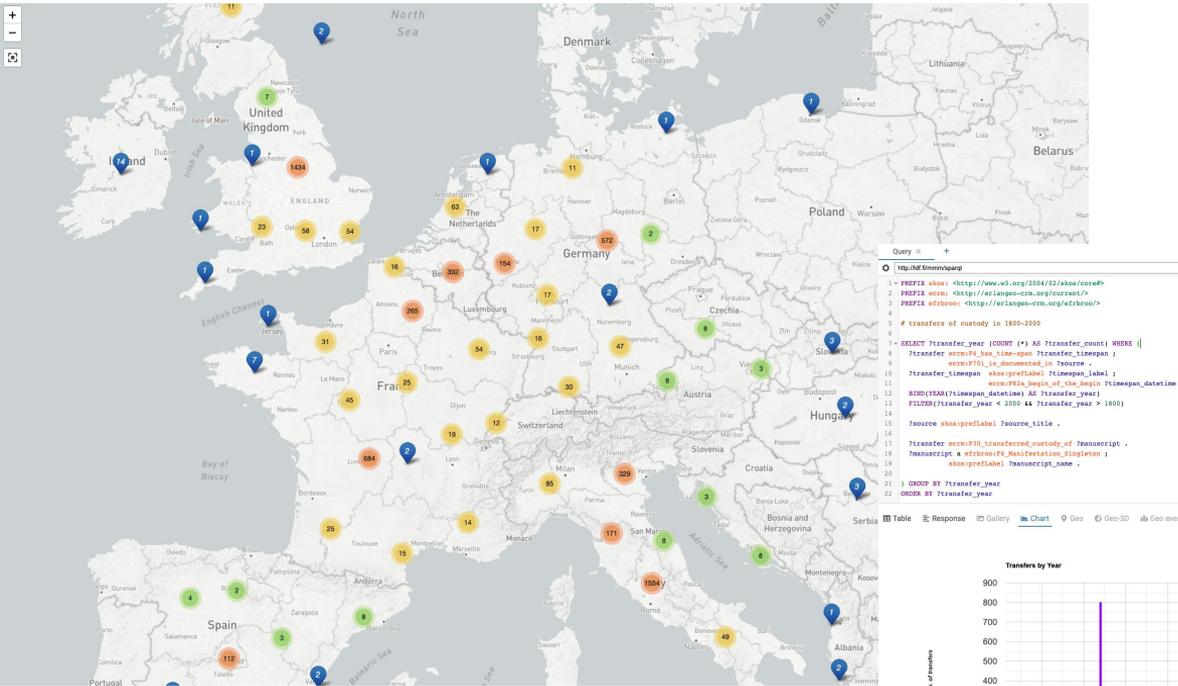
**Wartime Events (URI: <http://ldf.fi/warsa/events>)**

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

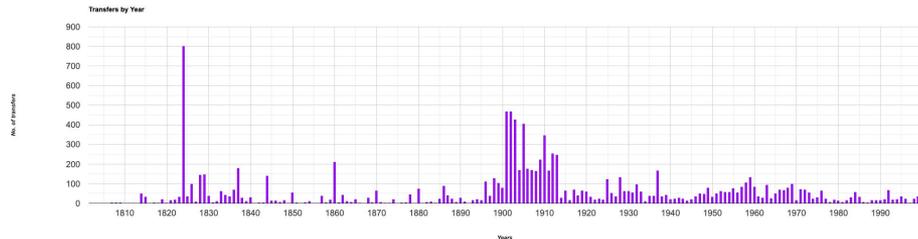
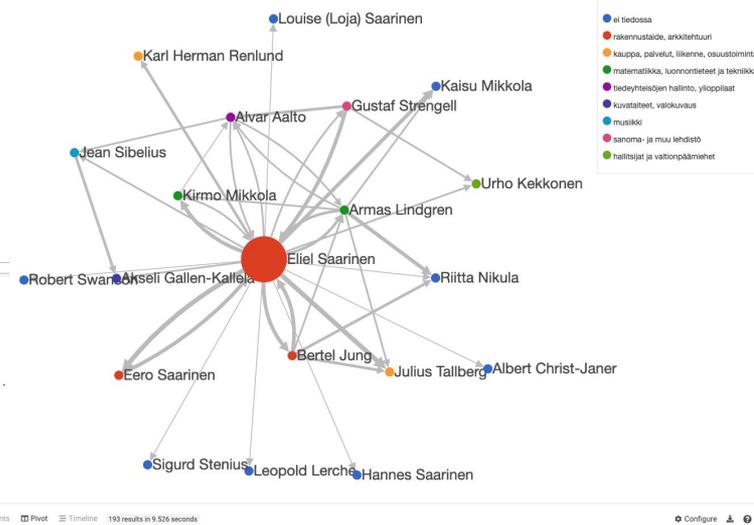


Information on the events of the Winter War and Continuation War were derived from the books Talvisodan pikkujättiläinen, Jatkosodan pikkujättiläinen, and the organization records of the Finnish Defence Force.

# Analysis and visualization tools



## Eiel Saarinen



# Thank you!

Linked Open Data Infrastructure for Digital Humanities  
(LODI4DH): <https://seco.cs.aalto.fi/projects/lodi4dh/>