

Linked Open Data Infrastructure for Digital Humanities in Finland (LODI4DH)

**Digital Humanities in Action:
Sampo Model and Portals for Cultural Heritage, 29.10.2020**

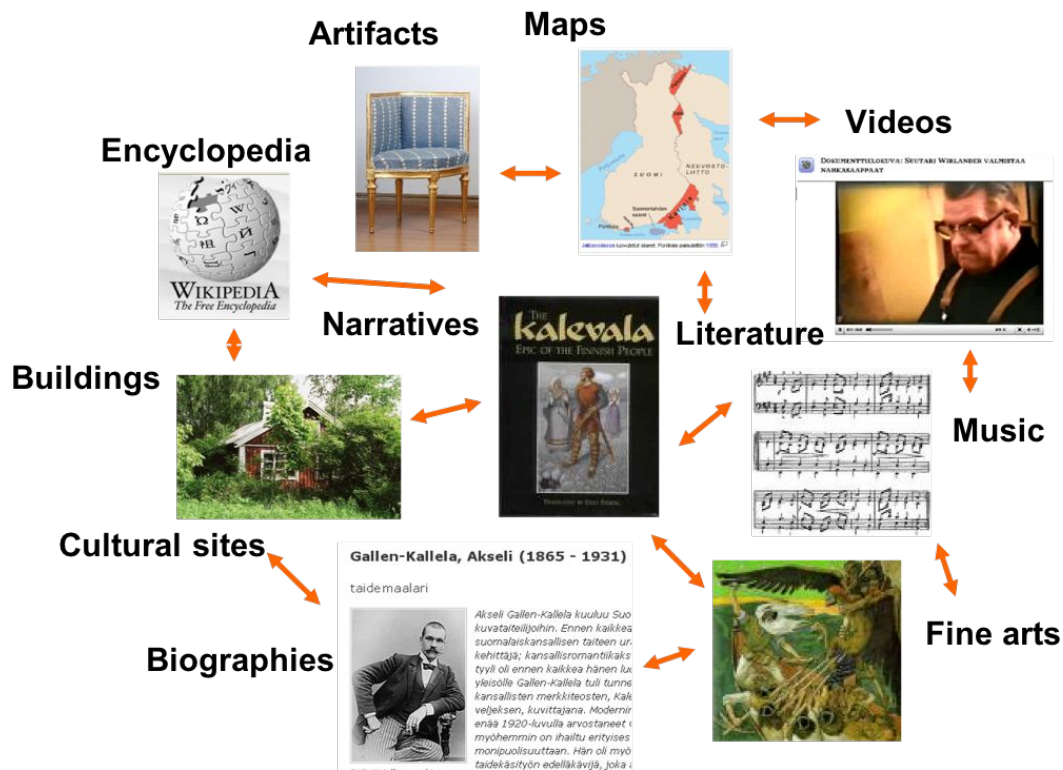
Jouni Tuominen

*HELDIG – Helsinki Centre for Digital Humanities, University of Helsinki, <http://heldig.fi>
Semantic Computing Research Group (SeCo), Aalto University, <http://seco.cs.aalto.fi>
jouni.tuominen@helsinki.fi*

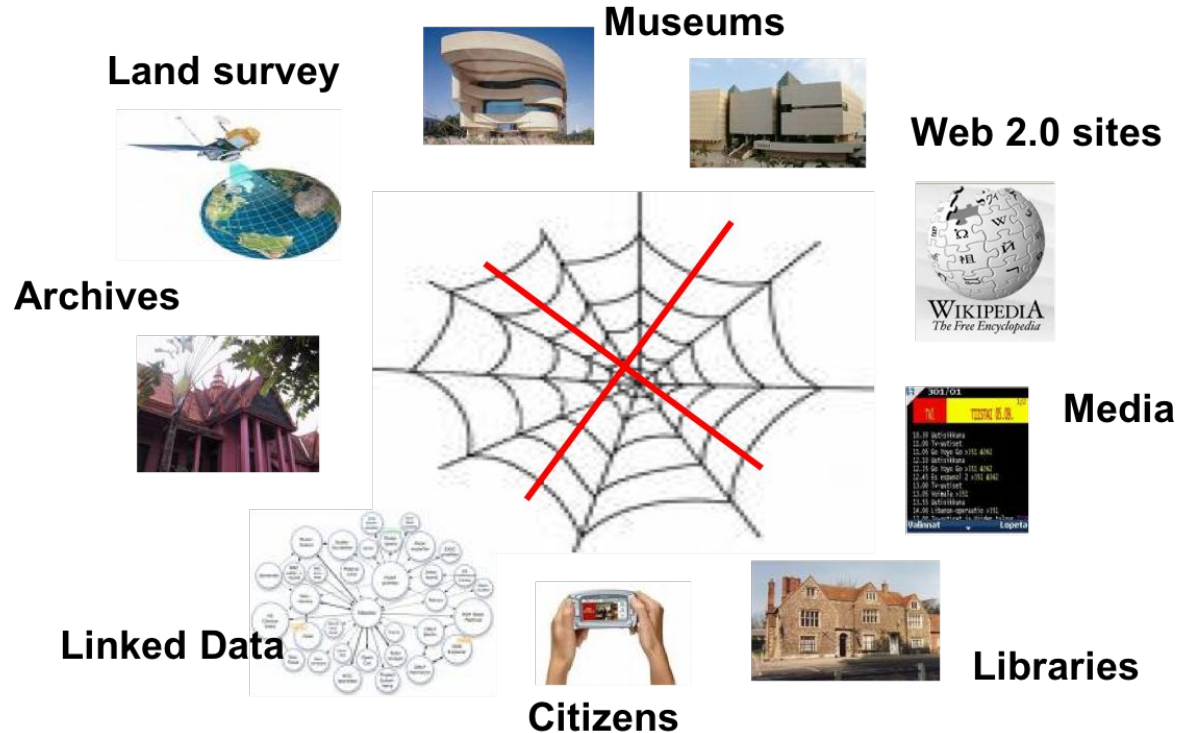
Outline

- **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



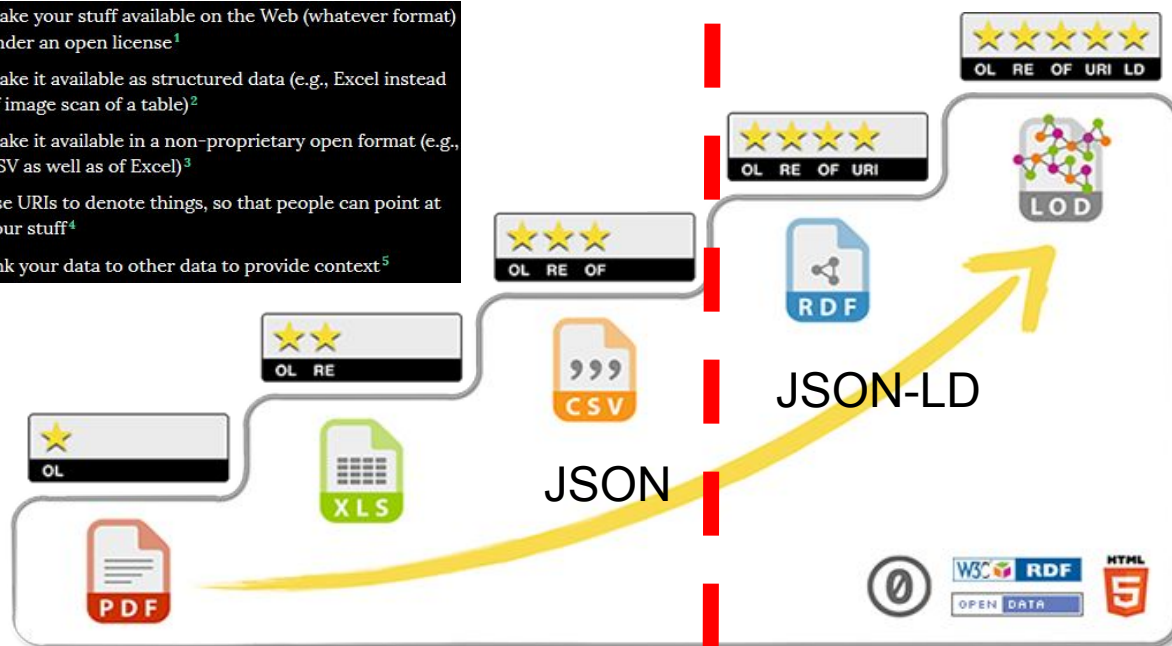
Cultural heritage content production system – different independent organizations



How to publish Linked Data?

5-star Linked Data model

- ★ make your stuff available on the Web (whatever format) under an open license¹
- ★★ make it available as structured data (e.g., Excel instead of image scan of a table)²
- ★★★ make it available in a non-proprietary open format (e.g., CSV as well as of Excel)³
- ★★★★ use URIs to denote things, so that people can point at your stuff⁴
- ★★★★★ link your data to other data to provide context⁵



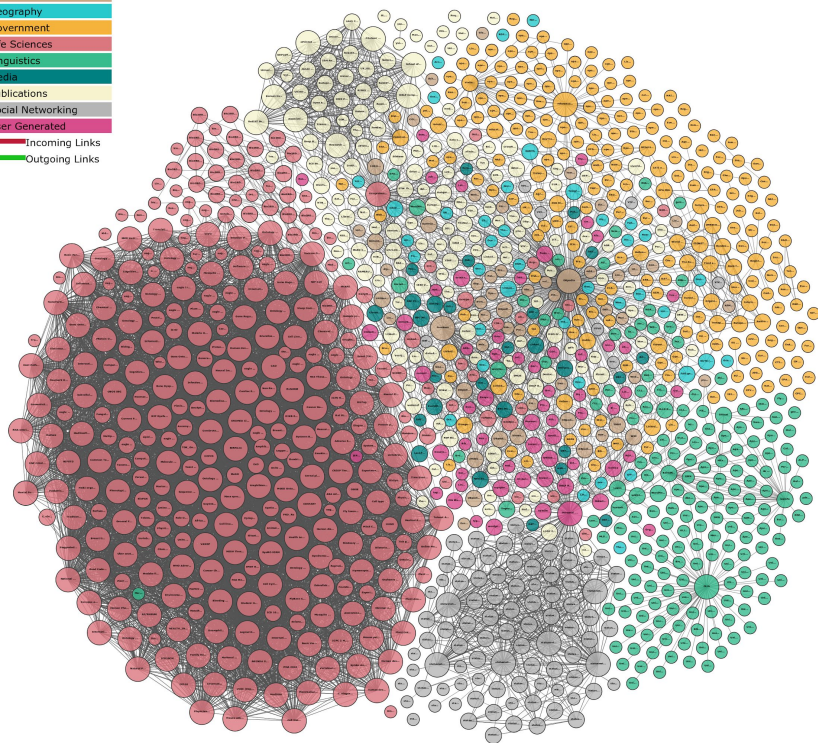
“Traditional” Open Data

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



Query 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)

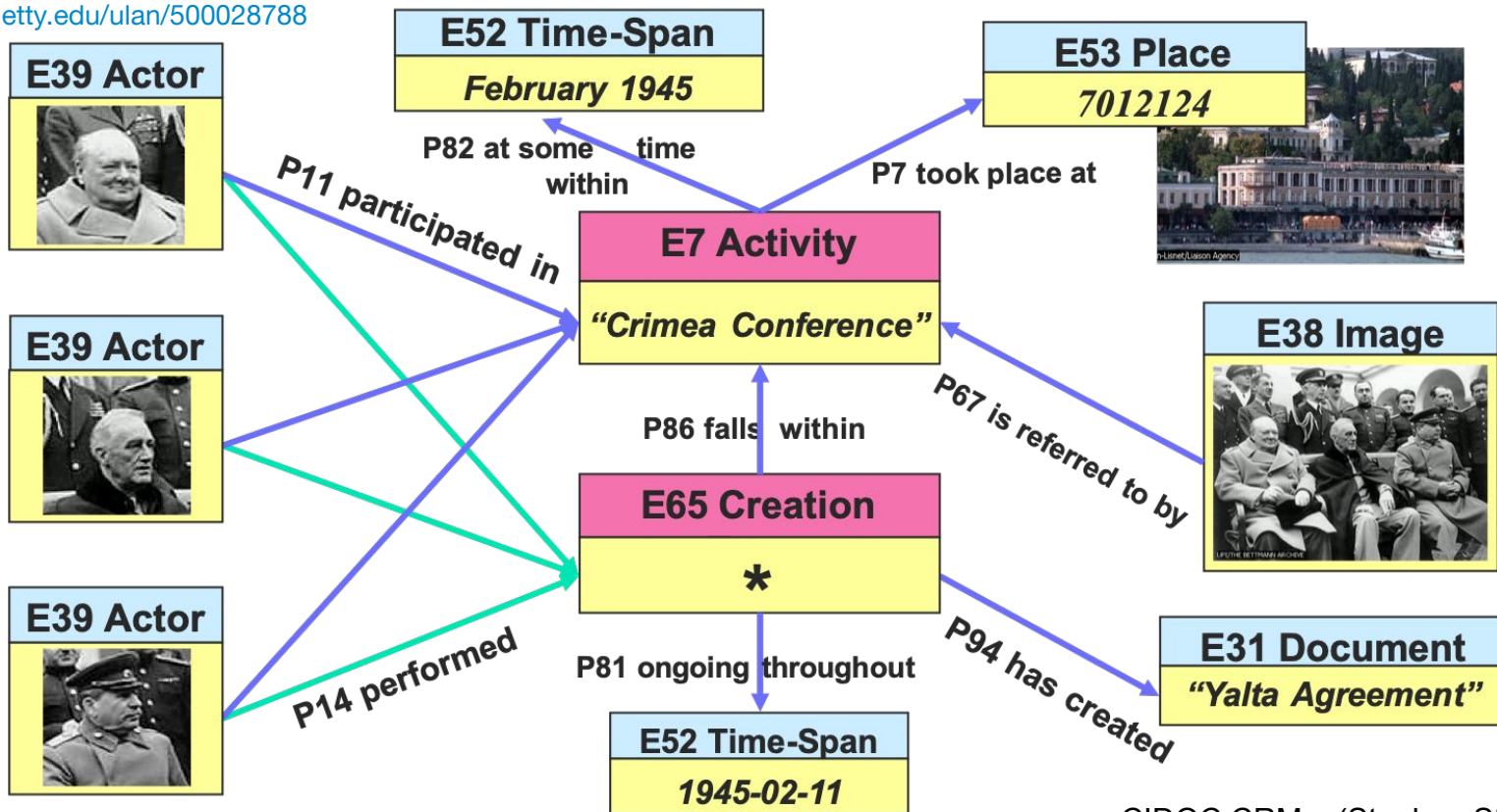
type	resource	prefLabel
"Battle"@en	http://ldf.fi/warsa/events/event_1078	"Defensive battles of Viipuri"@en
"Battle"@en	http://ldf.fi/warsa/events/event_1080	"Viipuri"@en
"Battle"@en	http://ldf.fi/warsa/events/event_1092	"Battle of Vyborg, Pappilanniemi and Kärenmäki sub-blocks"@en
"Battle"@en	http://ldf.fi/warsa/events/event_1093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

Actors

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



Pyotr Tšaikovski (fi)
Пётр Ильич Чайковский (ru)
Pyotr Ilyich Tchaikovsky (en)
Pyotr Tšajkovskij (sv)
Pyotr Tšajkovskij (no)
Pyotr Iljitsch Tschaiowski (de)
Pyotr Ilitch Tchaikovsky (fr)
Pyotr Ilitch Chaikovsky (es)
Pëtr Il'ič Čajkovskij (it)
Pyotr Iljitsj Tšajkovski (nl)
Pyotr Ilitch Tchaikovsky (pt)
Piotr Czajkowski (pl)
Piotr Iljics Csaikovski (ro)
Piotr Iljics Csaikovszkij (hu)

Events

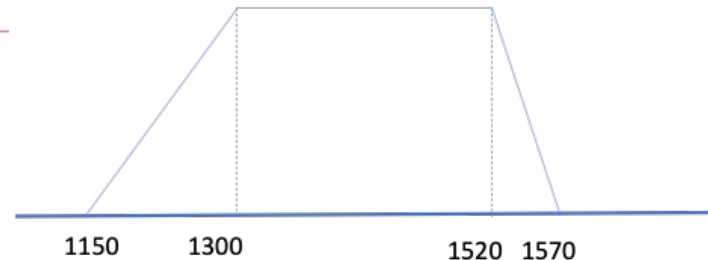
SAHA3 | historia - search
<http://agricola.utu.fi/rdf/phe863>

historiallinen tapahtuma: YYA-sopimus raukesi
[\[edit\]](#)

nimi	(fi) YYA-sopimus raukesi
tapahtumatyyppi	kahdenkeskiset kansainväliset sopimukset
kuuluu teemaan	poliittinen historia
kuka	Aho, Esko , Burbulis, Gennadi Eduardovič (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)
aiasana	diplomatia , kansainväliset sopimukset , kansainväliset suhteet , ulkopolitiikka , YYA-sopimus
liitty	YYA-sopimus solmittiin Neuvostoliiton kanssa , YYA-sopimusta jatkettiin kahdeksanneksi vuodeksi
liittyvä web-sivu	(fi) http://www.finlex.fi/fi/sopimukset/sopsteksti/1992/19920063
syy	Neuvostoliitto lakkautetaan
tyyppi	historiallinen tapahtuma

Times and periods

Middle Age (Keskiaika) in Finland

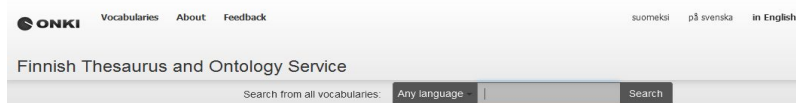


Keyword concepts



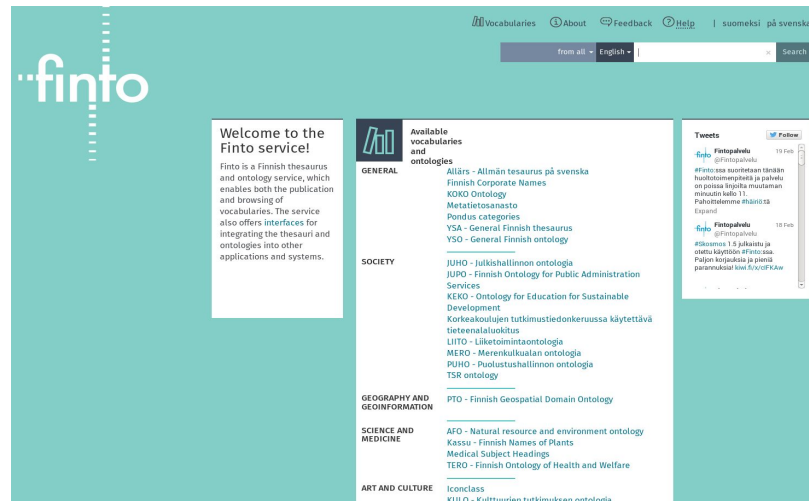
Places

Ontology services



Available vocabularies and ontologies

- General concepts**
 - General Finnish thesaurus
 - General Finnish upper ontology
 - KOKO Ontology
 - Allmän tesaurus på svenska - Allars
 - 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 - KITO
- 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



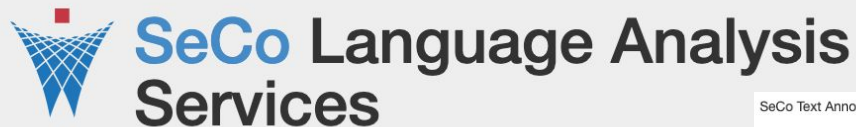
Finto / 2019:

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 [Ser 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. 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

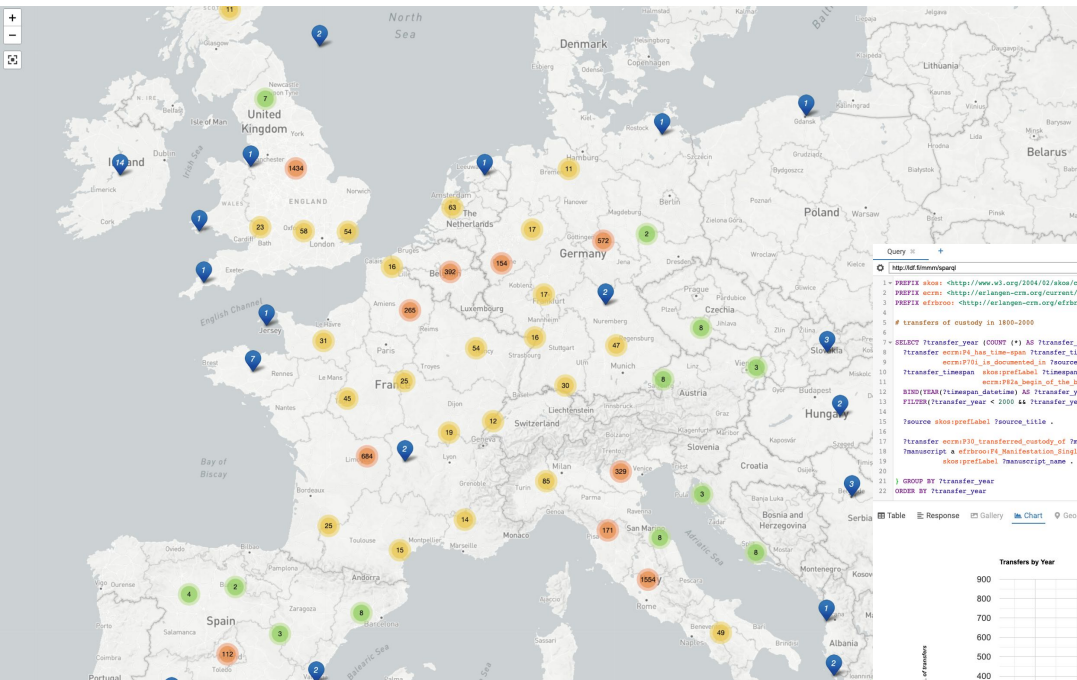
Wartime Events (URI: <http://ldf.fi/warsa/events>)

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

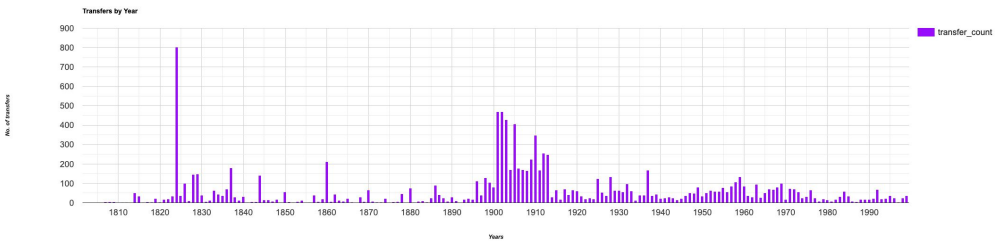
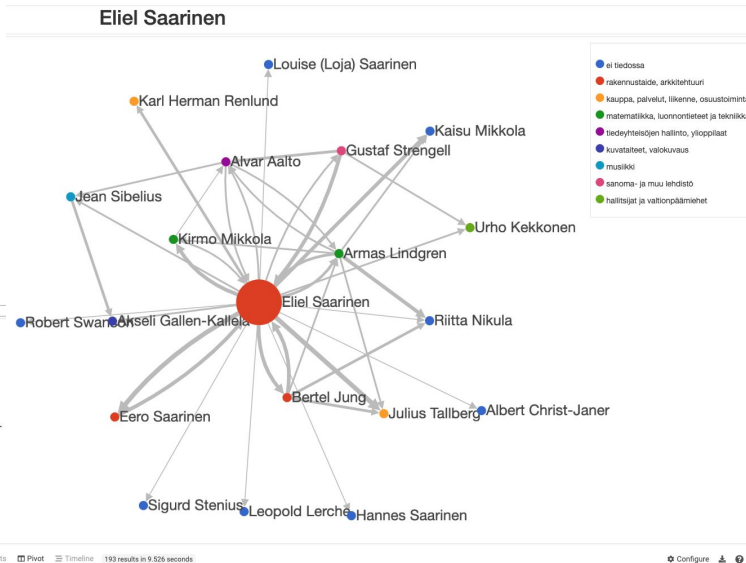


Information on the events of the Winter War and Continuation War were derived from the books Talvisodan pikkujähttiläinen, Jatkosodan pikkujähttiläinen, and the organization records of the Finnish Defence Force.

Analysis and visualization tools



```
Query in: +  
1 = PREFIX skos: <http://www.w3.org/2004/02/skos/core#>  
2 PREFIX ecrr: <http://ec.europa.eu/ecrr/ontology/>  
3 PREFIX ecrr: <http://ec.europa.eu/ecrr/ontology/>  
4 # transfers of custody in 1800-2000  
5  
6  
7 = SELECT ?transfer_year (COUNT(*) AS ?transfer_count) WHERE {  
8   ?transfer ecrr:P30 ?in_document ?in_document .  
9   ?transfer ecrr:P30 ?in_document ?in_document .  
10  ?transfer ecrr:P30 ?in_document ?in_document .  
11  ?transfer ecrr:P30 ?in_document ?in_document .  
12  ?transfer ecrr:P30 ?in_document ?in_document .  
13  ?transfer ecrr:P30 ?in_document ?in_document .  
14  ?transfer ecrr:P30 ?in_document ?in_document .  
15  ?transfer ecrr:P30 ?in_document ?in_document .  
16  ?transfer ecrr:P30 ?in_document ?in_document .  
17  ?transfer ecrr:P30 ?in_document ?in_document .  
18  ?transfer ecrr:P30 ?in_document ?in_document .  
19  ?transfer ecrr:P30 ?in_document ?in_document .  
20  ?transfer ecrr:P30 ?in_document ?in_document .  
21  ?transfer ecrr:P30 ?in_document ?in_document .  
22  ?transfer ecrr:P30 ?in_document ?in_document .  
23  ?transfer ecrr:P30 ?in_document ?in_document .  
24  ?transfer ecrr:P30 ?in_document ?in_document .  
25  ?transfer ecrr:P30 ?in_document ?in_document .  
26  ?transfer ecrr:P30 ?in_document ?in_document .  
27  ?transfer ecrr:P30 ?in_document ?in_document .  
28  ?transfer ecrr:P30 ?in_document ?in_document .  
29  ?transfer ecrr:P30 ?in_document ?in_document .  
30  ?transfer ecrr:P30 ?in_document ?in_document .  
31  ?transfer ecrr:P30 ?in_document ?in_document .  
32  ?transfer ecrr:P30 ?in_document ?in_document .  
33  ?transfer ecrr:P30 ?in_document ?in_document .  
34  ?transfer ecrr:P30 ?in_document ?in_document .  
35  ?transfer ecrr:P30 ?in_document ?in_document .  
36  ?transfer ecrr:P30 ?in_document ?in_document .  
37  ?transfer ecrr:P30 ?in_document ?in_document .  
38  ?transfer ecrr:P30 ?in_document ?in_document .  
39  ?transfer ecrr:P30 ?in_document ?in_document .  
40  ?transfer ecrr:P30 ?in_document ?in_document .  
41  ?transfer ecrr:P30 ?in_document ?in_document .  
42  ?transfer ecrr:P30 ?in_document ?in_document .  
43  ?transfer ecrr:P30 ?in_document ?in_document .  
44  ?transfer ecrr:P30 ?in_document ?in_document .  
45  ?transfer ecrr:P30 ?in_document ?in_document .  
46  ?transfer ecrr:P30 ?in_document ?in_document .  
47  ?transfer ecrr:P30 ?in_document ?in_document .  
48  ?transfer ecrr:P30 ?in_document ?in_document .  
49  ?transfer ecrr:P30 ?in_document ?in_document .  
50  ?transfer ecrr:P30 ?in_document ?in_document .  
51  ?transfer ecrr:P30 ?in_document ?in_document .  
52  ?transfer ecrr:P30 ?in_document ?in_document .  
53  ?transfer ecrr:P30 ?in_document ?in_document .  
54  ?transfer ecrr:P30 ?in_document ?in_document .  
55  ?transfer ecrr:P30 ?in_document ?in_document .  
56  ?transfer ecrr:P30 ?in_document ?in_document .  
57  ?transfer ecrr:P30 ?in_document ?in_document .  
58  ?transfer ecrr:P30 ?in_document ?in_document .  
59  ?transfer ecrr:P30 ?in_document ?in_document .  
60  ?transfer ecrr:P30 ?in_document ?in_document .  
61  ?transfer ecrr:P30 ?in_document ?in_document .  
62  ?transfer ecrr:P30 ?in_document ?in_document .  
63  ?transfer ecrr:P30 ?in_document ?in_document .  
64  ?transfer ecrr:P30 ?in_document ?in_document .  
65  ?transfer ecrr:P30 ?in_document ?in_document .  
66  ?transfer ecrr:P30 ?in_document ?in_document .  
67  ?transfer ecrr:P30 ?in_document ?in_document .  
68  ?transfer ecrr:P30 ?in_document ?in_document .  
69  ?transfer ecrr:P30 ?in_document ?in_document .  
70  ?transfer ecrr:P30 ?in_document ?in_document .  
71  ?transfer ecrr:P30 ?in_document ?in_document .  
72  ?transfer ecrr:P30 ?in_document ?in_document .  
73  ?transfer ecrr:P30 ?in_document ?in_document .  
74  ?transfer ecrr:P30 ?in_document ?in_document .  
75  ?transfer ecrr:P30 ?in_document ?in_document .  
76  ?transfer ecrr:P30 ?in_document ?in_document .  
77  ?transfer ecrr:P30 ?in_document ?in_document .  
78  ?transfer ecrr:P30 ?in_document ?in_document .  
79  ?transfer ecrr:P30 ?in_document ?in_document .  
80  ?transfer ecrr:P30 ?in_document ?in_document .  
81  ?transfer ecrr:P30 ?in_document ?in_document .  
82  ?transfer ecrr:P30 ?in_document ?in_document .  
83  ?transfer ecrr:P30 ?in_document ?in_document .  
84  ?transfer ecrr:P30 ?in_document ?in_document .  
85  ?transfer ecrr:P30 ?in_document ?in_document .  
86  ?transfer ecrr:P30 ?in_document ?in_document .  
87  ?transfer ecrr:P30 ?in_document ?in_document .  
88  ?transfer ecrr:P30 ?in_document ?in_document .  
89  ?transfer ecrr:P30 ?in_document ?in_document .  
90  ?transfer ecrr:P30 ?in_document ?in_document .  
91  ?transfer ecrr:P30 ?in_document ?in_document .  
92  ?transfer ecrr:P30 ?in_document ?in_document .  
93  ?transfer ecrr:P30 ?in_document ?in_document .  
94  ?transfer ecrr:P30 ?in_document ?in_document .  
95  ?transfer ecrr:P30 ?in_document ?in_document .  
96  ?transfer ecrr:P30 ?in_document ?in_document .  
97  ?transfer ecrr:P30 ?in_document ?in_document .  
98  ?transfer ecrr:P30 ?in_document ?in_document .  
99  ?transfer ecrr:P30 ?in_document ?in_document .  
100 ?transfer ecrr:P30 ?in_document ?in_document .
```



Linked Data for DH: what is needed?

- **Ontologies and metadata models for harmonizing data**
- **Tools for data creation and processing**
 - Ontology services
 - Linguistic processing tools
 - Linked Data publication services
- **Analysis and visualization tools**

Next presentation today:
Extracting Knowledge from Finnish Texts

Today's presentations:
Sampo-UI Framework and
10 Sampo portals!