Aalto University
School of Science

UNIVERSITY OF HELSINKI

SeCo

HELDIG
Helsinki Centre for Digital Humanities

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

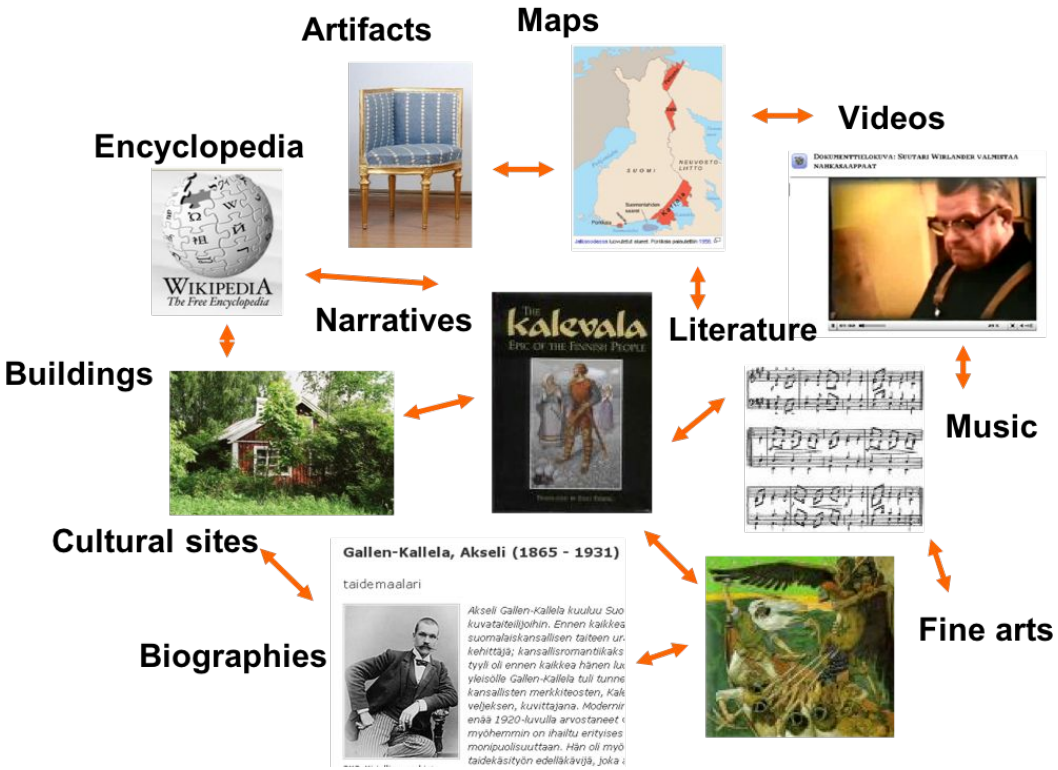
Digital Humanities in Action:
Sampo Model and Portals for Cultural Heritage, 29.10.2020

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

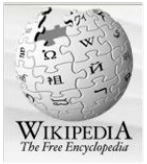


Cultural heritage content production system – different independent organizations

Land survey

Museums

Web 2.0 sites



Archives

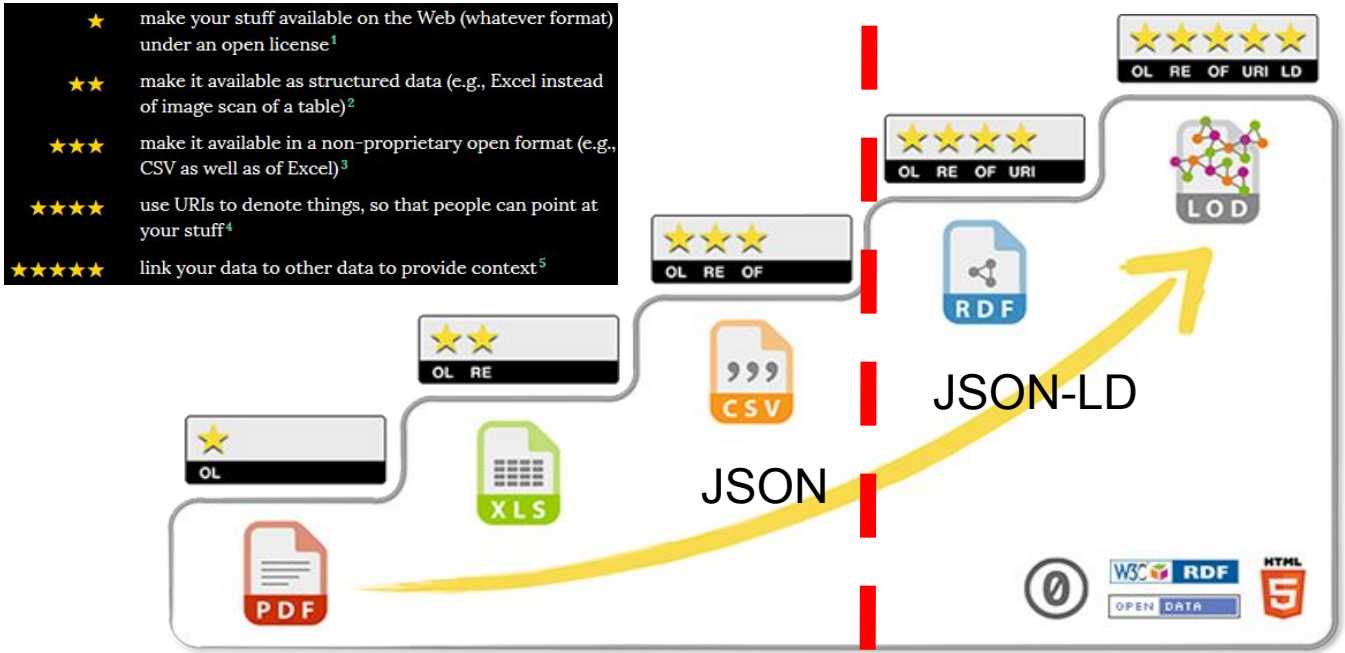
Media

Linked Data

Citizens

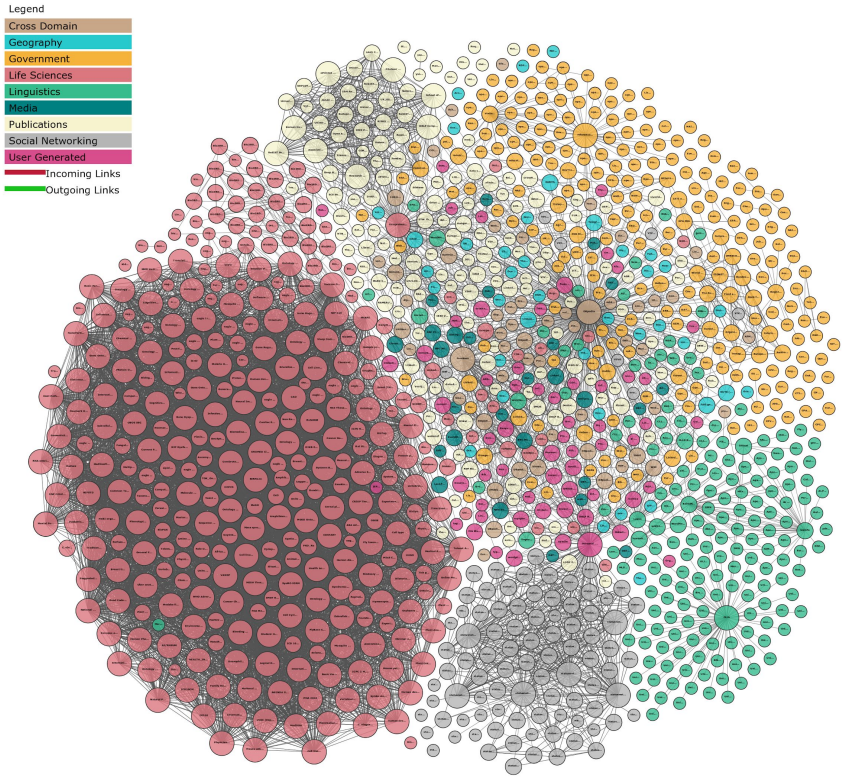
Libraries

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



"Traditional" Open Data | Linked Open Data (Tim Berners-Lee)

http://5stardata.info



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)

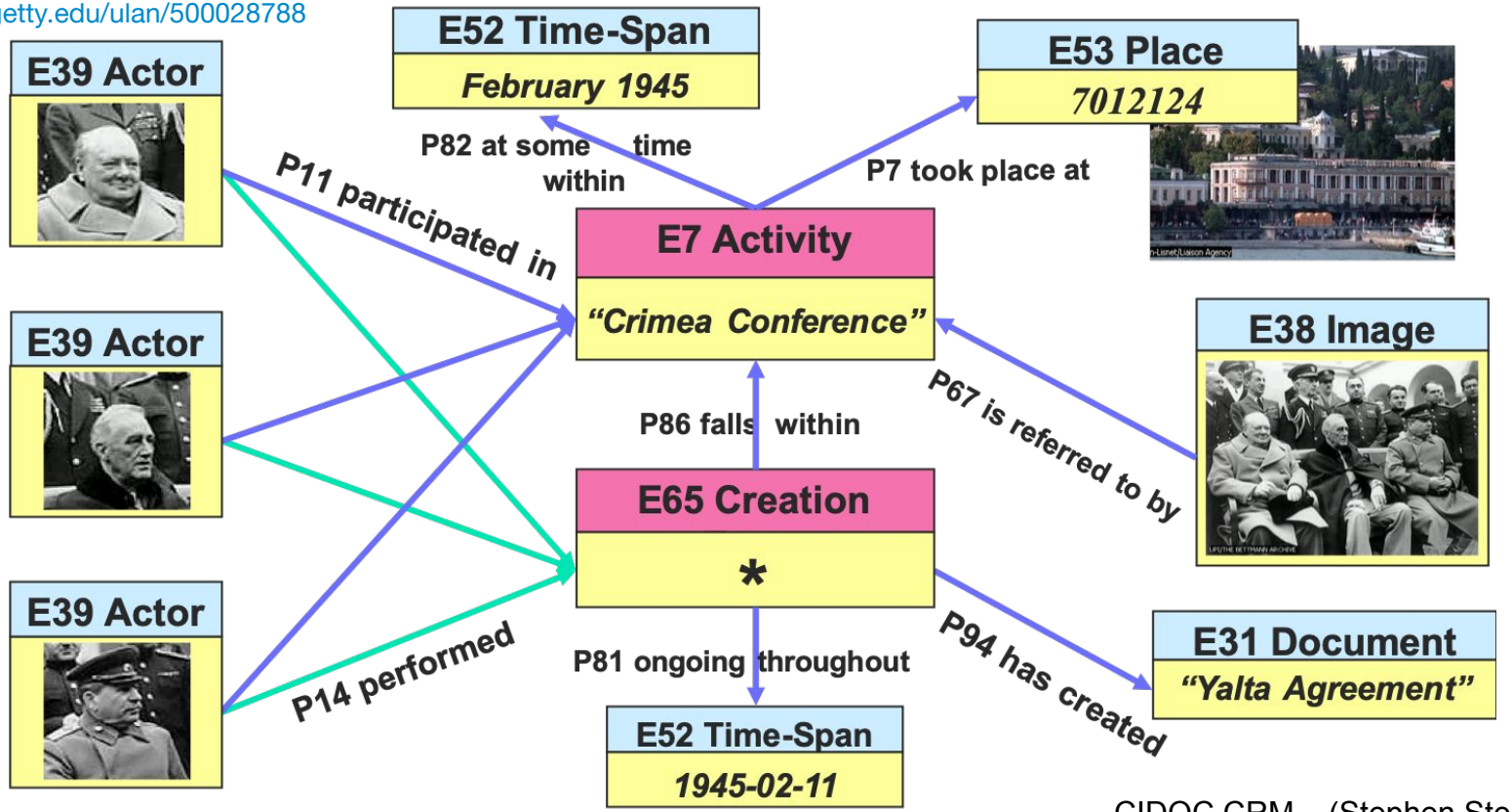
type	resource	prefLabel
"Battle"@en	http://ldf.fi/warsa/events/event_1078	"Defensive battles of Viipuri"@en
"Battle"@en	http://ldf.fi/warsa/events/event_1080	"Viipuri"@en
"Battle"@en	http://ldf.fi/warsa/events/event_1092	"Battle of Vyborg, Pappilanniemi and Kärenmäki sub-blocks"@en
"Battle"@en	http://ldf.fi/warsa/events/event_...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



CIDOC CRM (Stephen Stead, 2008)

Domain ontologies

Actors

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

Pjotr Tšaikovski (fi)
Пётр Ильич Чайко́вский (ru)
Pyotr Ilyich Tchaikovsky (en)
Pjotr Tjajkovskij (sv)
Pjotr Tsjajkovskij (no)
Piotr 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)

Events

SAHA3 | historia - search

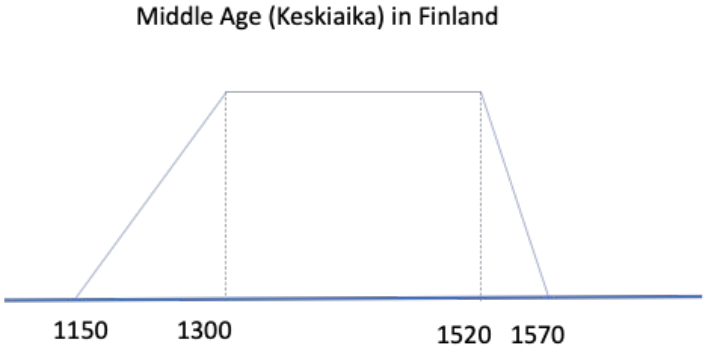
http://agricola.utu.fi/rdf/phe863

historiallinen tapahtuma: YYA-sopimus raukesi

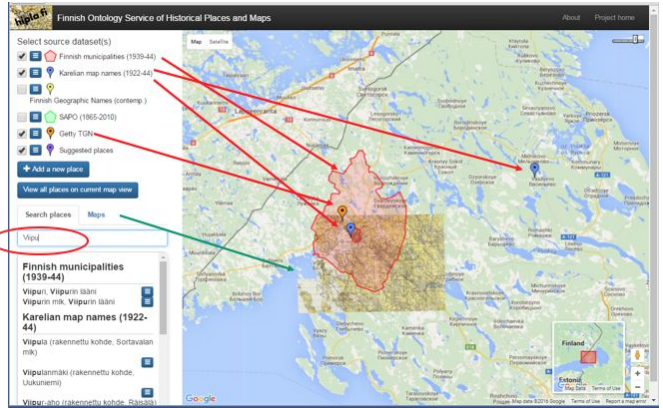
[edit]

nimi	(fi) YYA-sopimus raukesi
tapahtumatyyppi	kahdenkeskiset kansainväliset sopimukset
kuuluu teemaan	poliittinen historia
kuka	Aho, Esko, Burbulis, Gennadi 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

Times and periods

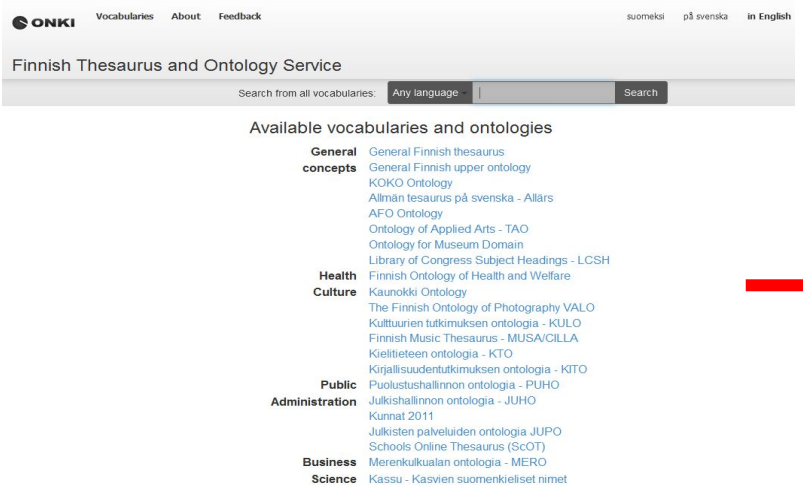


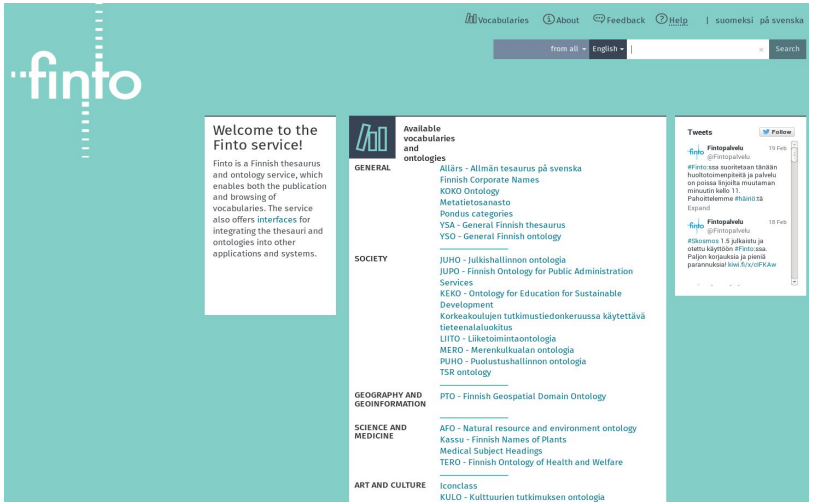
Places



Keyword concepts

Ontology services





Finto / 2019:

280 000 users (returning visits)

1,9 million page loads

32 million requests on programmatic application interface (API)

Linguistic processing tools

SeCo Language Analysis Services

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.

SeCo Aalto University HELDIG Helsinki Centre for Digital Humanities UNIVERSITY OF HELSINKI

Linked Data publication services

Linked Data Finland

Living Laboratory Data Service for the Semantic Web

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

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
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 We
our work on the Linked Data paradigm and stack of standards, which combines an expressive, sema
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 t
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

This added support should come with as little extra work as possible to the data publisher. Our hypot
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.

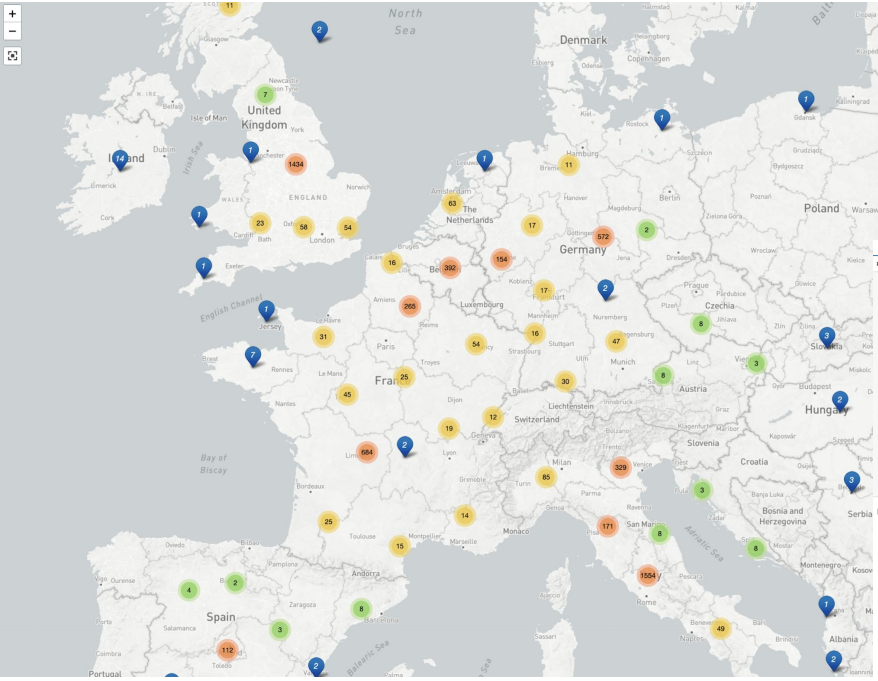
Detailed Dataset Contents

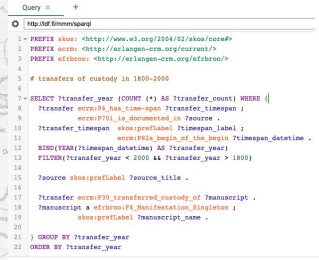
Wartime Events (URI: http://ldf.fi/warsa/events)

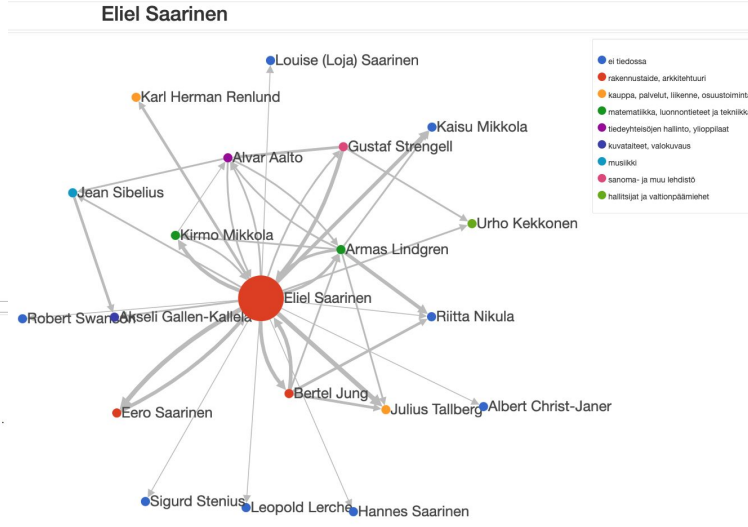
(Browse data / View in Sotasampo.fi / Download)

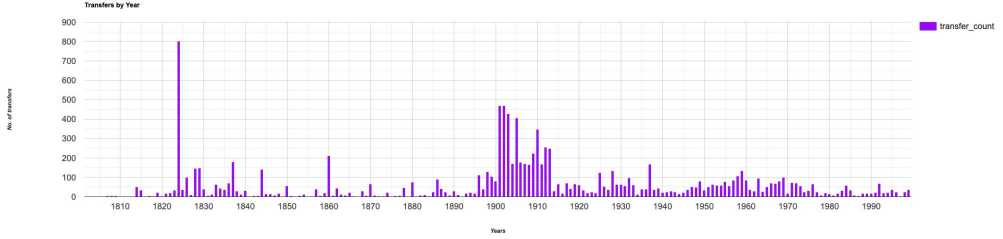
Information on the events of the Winter War and Continuation War were derived from the books Talvisodan pikkujättiläinen, Jatkosodan pikkujättiläinen, and the organization records of the Finnish Defence Force.

Analysis and visualization tools









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!