Aalto University
School of Science

UNIVERSITY OF HELSINKI

SeCo

HELDIG
Helsinki Centre for Digital Humanities

HELSINKI INSTITUTE FOR
SOCIAL SCIENCES
AND HUMANITIES

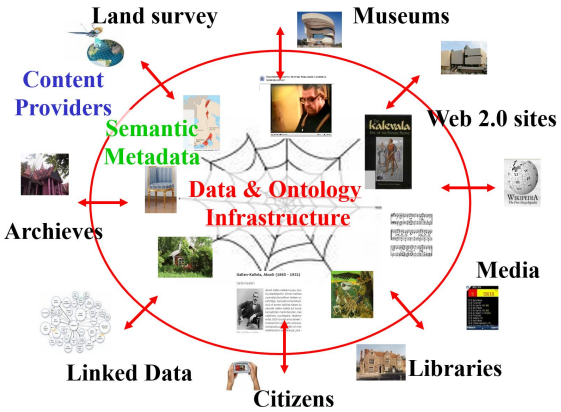
Publishing Sampo data (dump), data service (LDF.fi), and portal on the Web

Jouni Tuominen

Sampo Model for 1) Human Users and 2) Machines

LINKED DATA SERVICE

Findable, **A**ccessible, **I**nteoperable, **R**e-usable
FAIR data!



Linked Data Finland
Living Laboratory Data Service for the Semantic Web

https://ldf.fi/

SPARQL
{REST API}

APPLICATIONS & TOOLS

Use Cases

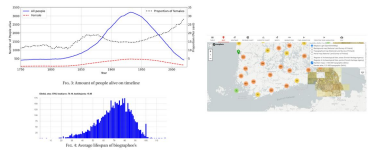
Tools

Sampo Portals & Applications

Sampos:
https://seco.cs.aalto.fi/applications/sampo/

Sampo-UI, etc.

Digital Humanities Research



Google Colab, Jupyter, YASGUI, etc.

References

Eero Hyvönen: **How to Create a National Cross-domain Ontology and Linked Data Infrastructure and Use It on the Semantic Web**. *Semantic Web*, 2024. DOI: 10.3233/SW-243468.

Eero Hyvönen: **Digital Humanities on the Semantic Web: Sampo Model and Portal Series**. *Semantic Web*, vol. 14, no. 4, pp. 729-744, 2023.

Esko Ikkala, Eero Hyvönen, Heikki Rantala and Mikko Koho: **Sampo-UI: A Full Stack JavaScript Framework for Developing Semantic Portal User Interfaces**. *Semantic Web*, vol. 13, no. 1, 69-84, 2022.

Sampo data dumps (static files)

- **Zenodo**
  - Static RDF files
  - Versioning, digital object identifier (DOI)
- **Also other formats (e.g., CSV, XML) can be provided**
  - E.g. ParliamentSampo: CSV & XML (Parla-CLARIN, ParlaMint)

Linked Data Finland Living Lab http://ldf.fi

Home
Project
Datasets
Search Data
Schemas
Services
Policies
Documentation
Validation
Linked Data Science
Applications
Your Data?
Linked Data School

Linked Data Finland

Living Laboratory Data Service for the Semantic Web

This site is the Living Laboratory of the FIN-CLARIAH research infrastructure initiative, conducted by the Semantic Computing Research Group at Aalto University in collaboration with University of Helsinki and a large consortium of other Finnish universities and organizations.

Our goal is to make life easier for both Linked Data publishers and consumers on the Web. We base our work on the Linked Data principles and stack of standards founded on the semantic data model RDF, standardized query language SPARQL, and live HTTP URIs for identifiers.

5-star Linked Data

The baseline of our work is the 5-star Linked Data model, proposed 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.

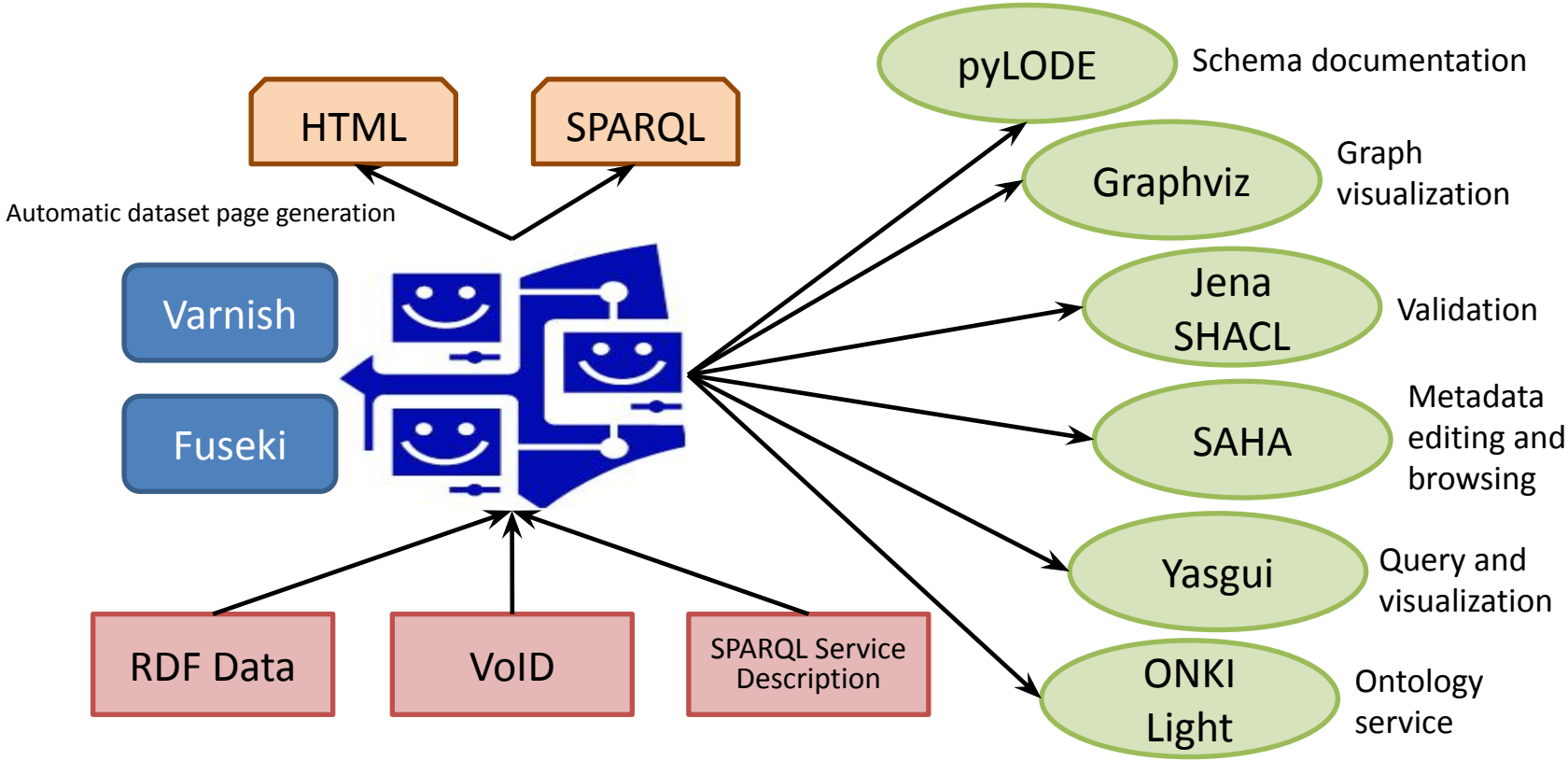
8-star Linked Data

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 the underlying data models used, how well the data actually conforms to the models, as well as trustworthiness of the data. To these ends, we extend the model with three additional stars to encourage data publishers to make their data more re-usable:

★★★★★★ Describe explicitly and publish the schemas used in the dataset (unless the schemas are already available elsewhere).
★★★★★★★ Explain the quality of the dataset against the schemas used in it, so that the user can tell whether the data quality matches her needs.
★★★★★★★★ Give explanations when the data is factually correct with respect to the real world and when possibly not.

Eero Hyvönen and Jouni Tuominen: 8-star Linked Open Data Model: Extending the 5-star Model for Better Reuse, Quality, and Trust of Data. P*osters, Demos, Workshops, and Tutorials of the 20th International Conference on Semantic Systems (SEMANTiCS 2024)*, vol. 3759, CEUR Workshop Proceedings, September, 2024. https://ceur-ws.org/Vol-3759/paper4.pdf

LDF.fi – generic Linked Data publication platform



LDF.fi:
Linked Data browsing for humans and machines (try!)

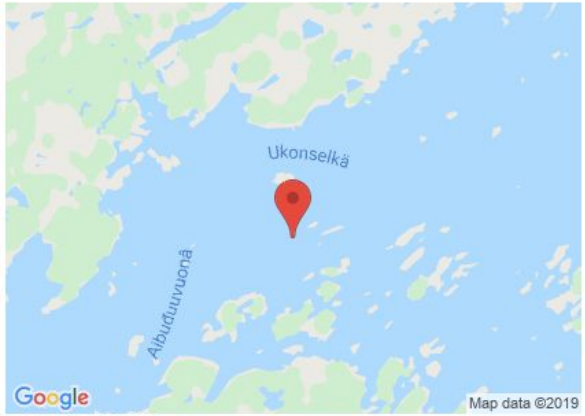
```
<http://ldf.fi/kotus-names-archive/Q5601617>
        a       <http://ldf.fi/schema/kotus-names-archive/place_type_41> ;
        <http://ldf.fi/schema/kotus-names-archive/collection>
                "1-kokoelma" ;
        <http://ldf.fi/schema/kotus-names-archive/collector>
                "Portti, Reija" ;
        <http://ldf.fi/schema/kotus-names-archive/name_type>
                "paikannimi" ;
        <http://ldf.fi/schema/kotus-names-archive/parish>
                "Inari" ;
        <http://ldf.fi/schema/kotus-names-archive/positioning_accuracy>
                "Paikka sijaitsee keruukartalla tämän karttaruudun alueella." ;
        <http://ldf.fi/schema/kotus-names-archive/stamp_date>
                1990 ;
        <http://www.w3.org/2002/07/owl#sameAs>
                <https://nimiarkisto.fi/wiki/Q5601617> ;
        <http://www.w3.org/2003/01/geo/wgs84_pos#lat>
                68.936487 ;
        <http://www.w3.org/2003/01/geo/wgs84_pos#long>
                27.282297 ;
        <http://www.w3.org/2004/02/skos/core#prefLabel>
                "Ukko" .
```

SAHA3 | kotus-names-archive - search

http://ldf.fi/kotus-names-archive/Q5601617

saari: Ukko

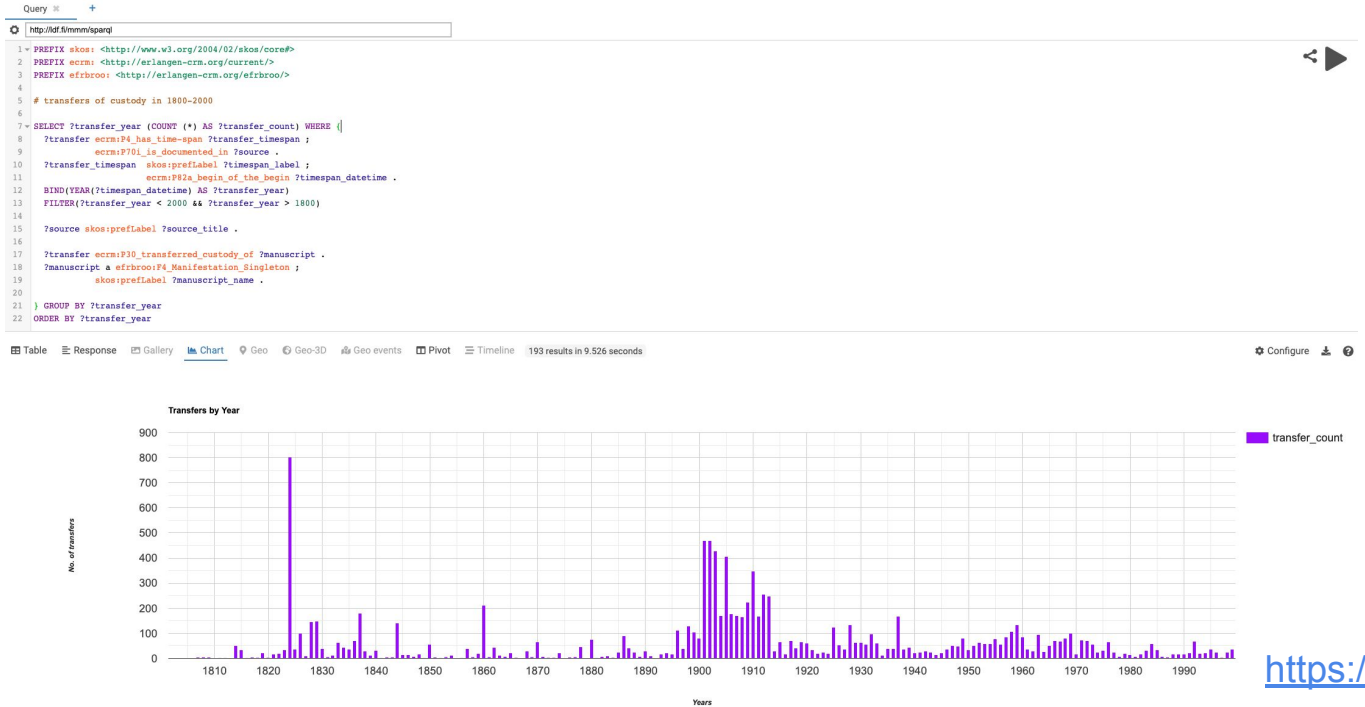
[edit]



68.936487, 27.282297

Collection	1-kokoelma
Collection parish (in ca 1938)	Inari
Collector	Portti, Reija
http://ldf.fi/schema/kotus-names-archive/positioning_accuracy	Paikka sijaitsee keruukartalla tämän karttaruudun alueella.
Latitude	68.936487
Longitude	27.282297
Name	Ukko
Name type	paikannimi
Same as	https://nimiarkisto.fi/wiki/Q5601617
Stamp date	1990
Type	saari

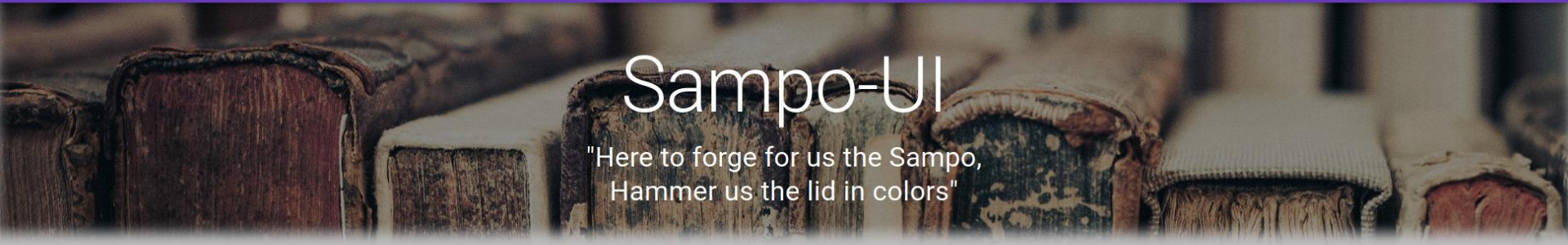
LDF.fi: SPARQL endpoints for querying data for analysis, visualization, mash-up apps (CSV, JSON, XML) (try!)



https://yasgui.triply.cc

LDF.fi:
Download data graphs (Turtle, JSON-LD, RDF/XML, …)
via SPARQL HTTP Graph Store Protocol (try!)

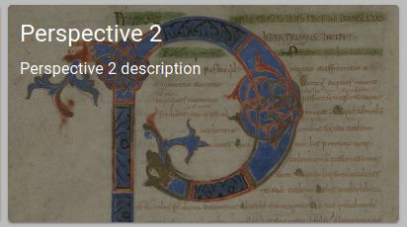
```
<http://ldf.fi/findsampo/finds/km_39503-2>
        a       <http://ldf.fi/schema/findsampo/core/Find> ;
        <http://ldf.fi/schema/findsampo/core/earliest_period>
                <http://ldf.fi/findsampo/periods/p17> ;
        <http://ldf.fi/schema/findsampo/core/facet_object_type>
                <http://ldf.fi/findsampo/object_types/facet_terms/p20> ,
<http://ldf.fi/findsampo/object_types/facet_terms/p58> ;
        <http://ldf.fi/schema/findsampo/core/find_site_coordinates>
                <http://ldf.fi/findsampo/find_sites/find_site_of_39503-2> ;
        <http://ldf.fi/schema/findsampo/core/found_in>
                <http://ldf.fi/findsampo/places/kanta-haeme> ,
<http://ldf.fi/findsampo/places/haemeenlinna> ;
        <http://ldf.fi/schema/findsampo/core/found_in_municipality>
                <http://ldf.fi/findsampo/places/haemeenlinna> ;
        <http://ldf.fi/schema/findsampo/core/found_in_province>
                <http://ldf.fi/findsampo/places/kanta-haeme> ;
        <http://ldf.fi/schema/findsampo/core/has_creation_time_span>
                [ <http://erlangen-crm.org/current/P82a_begin_of_the_begin>
                        "-0500-01-01"^^<http://www.w3.org/2001/XMLSchema#date> ;
                  <http://erlangen-crm.org/current/P82b_end_of_the_end>
                        "1300-12-31"^^<http://www.w3.org/2001/XMLSchema#date>
                ] ;
        <http://ldf.fi/schema/findsampo/core/identifier>
                "KM39503:2" ;
        <http://ldf.fi/schema/findsampo/core/latest_period>
                <http://ldf.fi/findsampo/periods/p17> ;
        <http://ldf.fi/schema/findsampo/core/length>
                255.0 ;
        <http://ldf.fi/schema/findsampo/core/material>
                <http://ldf.fi/findsampo/materials/p11> ;
        <http://ldf.fi/schema/findsampo/core/max_thickness>
                35.0 ;
        <http://ldf.fi/schema/findsampo/core/object_type>
                <http://ldf.fi/findsampo/object_types/keihaeaenkaerjet> ;
        <http://ldf.fi/schema/findsampo/core/period>
                <http://ldf.fi/findsampo/periods/p17> ;
        <http://ldf.fi/schema/findsampo/core/weight>
                505.7 ;
        <http://ldf.fi/schema/findsampo/core/width>
                69.0 ;
        <http://ldf.fi/schema/findsampo/extended/ltk/amount>
                "1" ;
```



Sampo-UI

"Here to forge for us the Sampo,
Hammer us the lid in colors"

Select a perspective to search and browse the knowledge graph:



Sampo infrastructure

- **All software components are run as Docker containers**
  - Portable services, can be run on a local machine/server
- **Triple stores (Fuseki) run on virtual machines**
- **Web applications (Varnish for routing + cache, www.ldf.fi, Sampo portals) run on OpenShift container cloud**

- **Part of the DARIAH-FI research infrastructure program**
- **Technical infrastructure provided by CSC – IT Center for Science (by Ministry of Education and Culture)**

DARIAH-FI
DARIAH-EU
Suomen Akatemia
Finlands Akademi
Research Council of Finland
Funded by the European Union
NextGenerationEU

Thank You!