Linked data Finland: towards a 7-star service platform for linked datasets

Eero Hyvönen[1,2], Miika Alonen[1], Jouni Tuominen[1], Eetu Mäkelä[1]

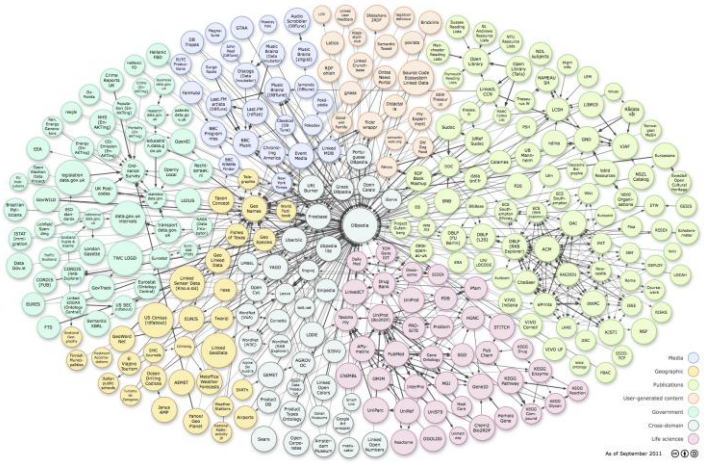
[1]Aalto University, Espoo, Finland

[2]University of Helsinki, Helsinki, Finland

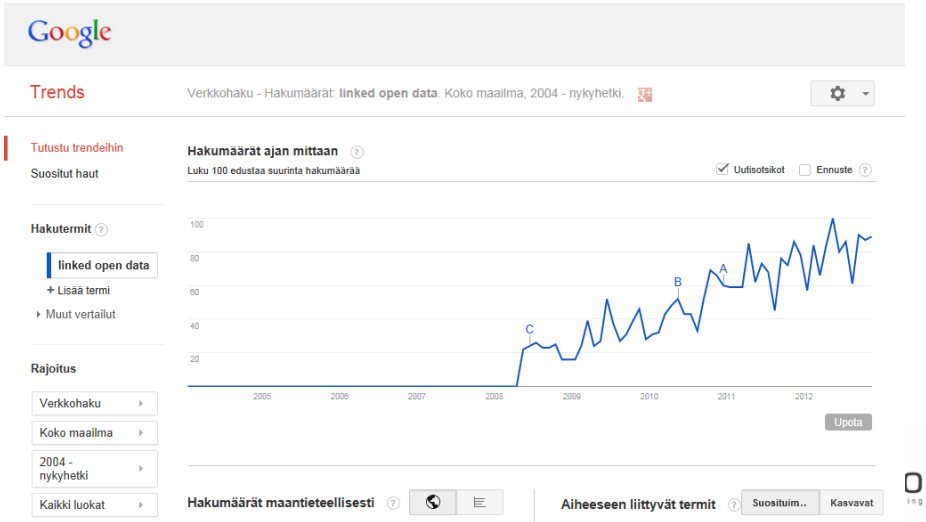
KNOWeSCAPE

Why Linked Data Revolution

- **Interoperability:** harmonizing heterogenenous data
- **Linking:** integrating distributed data
- **Semantic enriching:** intelligent applications
- **Data as services:** support re-use and application development
- **Coordination: W3C Semantic Web standards**



Linked Open Data Cloud



Google Trends: Linked Data

How to Publish Linked Data?
5-star Linked Data Model

Aalto University

HELSINGIN YLIOPISTO

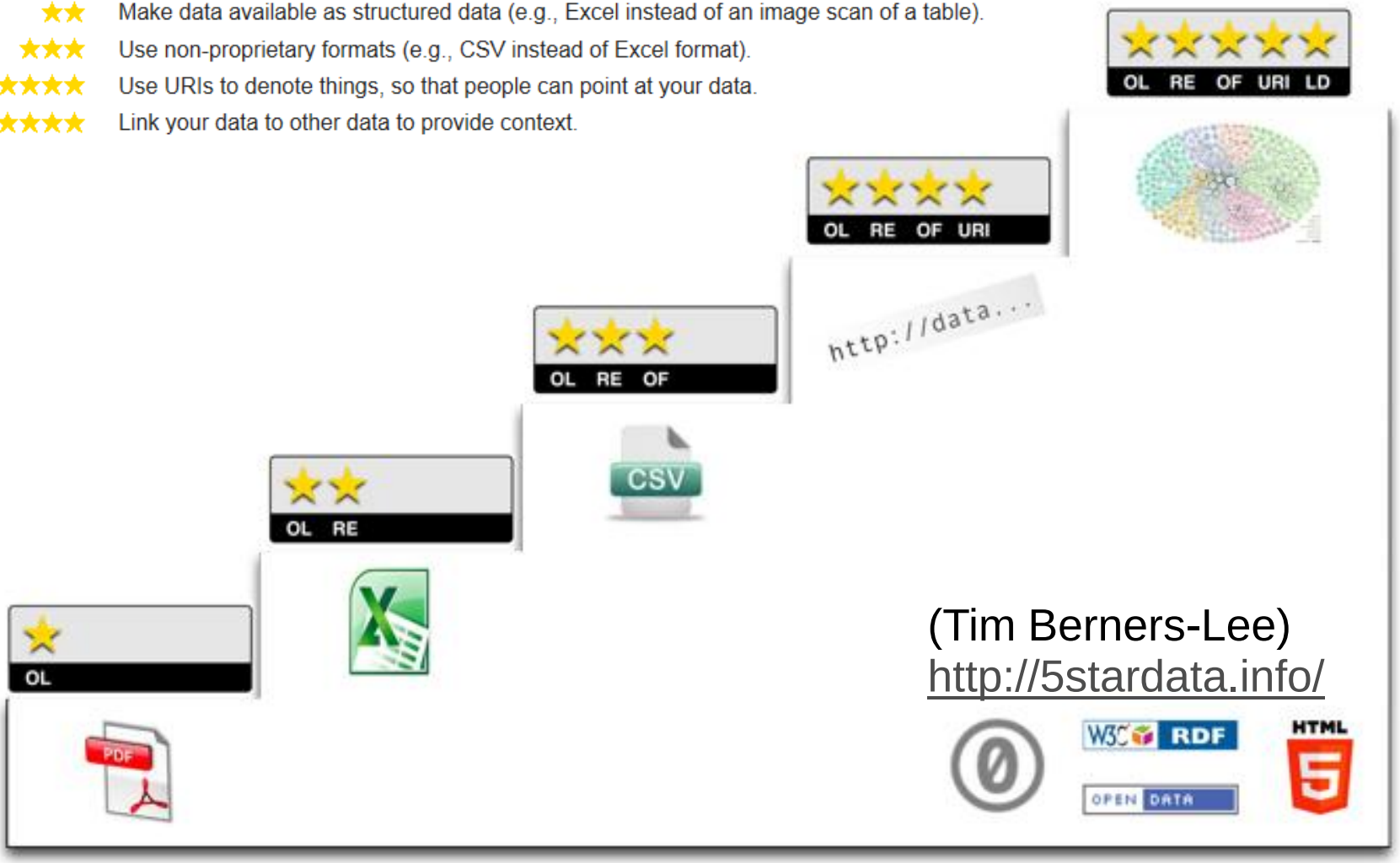
★ Make data available on the Web in whatever format under an open license.

★★ 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.



(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

In addition to these services, we extend the 5-star model into a *7-star model* of our own with two additional requirements:

★★★★★★ Provide your data with a schema so that people can understand and re-use your data easily.

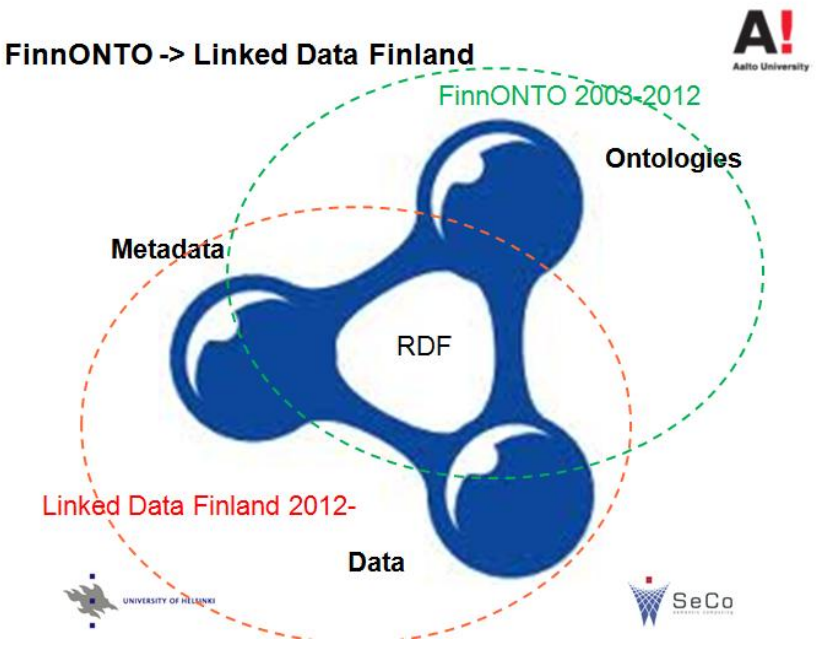
★★★★★★★ Validate your data against the schema so that people can trust the quality of your data.

Linked Data at the LDF.fi portal is published at different quality levels. Our goal is to get any useful data out quickly. The data can be made better by the community and more stars earned later.

Burj Al Arab

Why LDF.fi?

- Living Laborary 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 http://ldf.fi

Aalto University

HELSINGIN YLIOPISTO

- 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 the Aalto University and University of Helsinki. Our goal is to create a prototype of an open platform for publishing Linked Data as services on the web for applications to use. This is a sister platform of our National Ontology Service ONKI platform that focuses on ontologies.

5-star Linked Data

The basis of our work is on the 5-star Linked Data model, proposed originally by Tim Berners-Lee.

- ★ Make data available on the Web in whatever format under an open license.
- ★★ 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.

In a Linked Data service, the data is available for download in RDF with internal and external URI links. The data is also provided via a Linked Data browser that provides access to individual data items in RDF and HTML formats. A SPARQL endpoint is provided for querying and managing data: it constitutes the basis for application development.

7-star Linked Data Service

In addition to these services, we extend the 5-star model into a *7-star model* of our own with two additional requirements:

- ★★★★★★ Provide your data with a schema so that people can understand and re-use your data easily.
- ★★★★★★★ Validate your data against the schema so that people can trust the quality of your data.

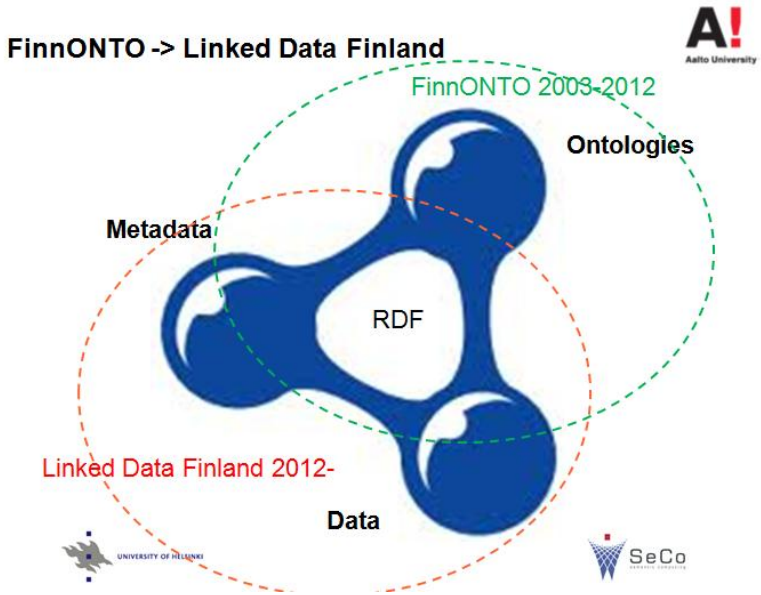
Linked Data at the LDF.fi portal is published at different quality levels. Our goal is to get any useful data out quickly. The data can be made better by the community and more stars earned later.

How to Use the Portal

SeCo semantic computing

Linked Data Finland Project

- ”National” Linked Data Project
  - Research: Linked data creation, enrichment, linking, curation, quality
  - 20 organizations, including e.g. 4 Finnish ministries
- Linked Open Data Living Laboratory
  - ONKI http://onki.fi for ontologies
  - LDF.fi for data and schemas
- Application Pilots
  - Cultural Heritage
  - Law
  - News
  - Travel
  - Linked Science



Datasets

Linked Data Finland

Below you find a list of datasets (services) published by LDF.fi.

Encyclopedic

1. Finnish Wikipedia as Linked Data (DBpedia). First publication of the Finnish DBpedia linked open data via a SPARQL endpoint.

Cultural Heritage

1. MuseumFinland. The data behind the classic online application from 2004 that was rewarded e.g. the international Semantic Web Challenge Award.
2. Semantic Kalevala and Folklore. Arguably the first publication of folklore as Linked Open Data in the world.
3. Data from CultureSampo (Kulttuurisampo) (work in progress). Data from the CultureSampo application that got the Outstanding Paper Award 2011-2012 of the Semantic Web Journal.
4. BookSampo (Kirjasampo). Data behind the famous collaborative Kirjasampo application of the Finnish Public Libaries that contains rich semantic descriptions of virtually all Finnish fiction literature.

History

1. Semantic National Biography. First national history ontology created in collaboration with Finnish historians.
2. World War I Linked Open Data. Result on an international collaboration regarding sharing data about war history.

Law

1. Finnish Law as Linked Open Data (Semantic Finlex). Finnish Law and law cases available as (Linked) Open Data for the first time!

Linked Open Universities

1. Linked Open Aalto. Aalto University's pioneering work in opening its data repositories as Linked Data for applications.

Linked Science

1. Bird Observations of the nation wide Tiira.fi Service. Collaborative citizen science data on the semantic web for applications, such as BirdWatch .
2. Bird Observations of Halias Station with Hanko Weather Data. Exceptional dataset of systematic bird observations for ca. 30 years from the Halias bird observation station in Hanko linked with weather observation data.

Ontologies

1. ONKI Light on SPARQL Ontology Service for thesauri. Large Linked Data collection of national thesauri developed into ontologies by the FinnONTO project 2003-2012.
2. Agent Ontology. First attempt to create a shared national data repository of historical persons and organizations.
3. Finnish Places as Linked Data. First attempt to create a shared national Linked Data repository of historical and contemporary places.
4. Finnish History Events. First attempt to create a shared Linked Data repository of national historical events created in collaboration with Finnish historians.

Aalto University

HELSINGIN YLIOPISTO

SeCo semantic computing

Services

- 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

Example Dataset: Finnish Law as Linked Data

Aalto University

HELSINGIN YLIOPISTO

Semantic Finlex

Linked Data Finland

★★★★★★

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

Semantic Finlex is the first open data publication of Finnish Legislation. Its RDF data has been transformed from the Finlex Data Bank maintained by the Ministry of Justice, and includes also selected cases from the Case Law. For more information, see the project page.

Download

Select link for downloading a graph in the service (i.e., dataset) in Turtle format:

- Finlex Legislation
- Finlex Case-law

Schemas Used

Following schemas are used in the datasets above:

- Court Decisions Schema: documentation, download
- Finlex Legislation Schema: documentation, download

URI Data Services

Give the URI without brackets (<>) and without encoding.

Human Use Case 1: Viewing the RDF Description of a URI

View the RDF description of a URI using the template: `http://ldf.fi/SERVICE/data?uri=URI`

URI: `http://ldf.fi/finlex/laki/p35`
Format: Turtle
☑ View result in browser (accept header = `text/plain`)

SeCo semantic computing

Human Use Case 2: Linked Data Browsing Starting from a URI

Start browsing from a URI using the template `http://ldf.fi/SERVICE/page?uri=URI`

URI: `http://ldf.fi/finlex/laki/p35`

Start

Machine Use Case: Getting the RDF Description of a Resource

If the URI dereferenceable, e.g., minted at the http://ldf.fi domain, just input the URI in a browser or use it in your favourite software. For example: `URI=http://ldf.fi/finlex/laki/p35`. If the URI is described in this LDF service you can use the URI template `http://ldf.fi/SERVICE/data?uri=URI` no matter whether the URI is dereferenceable or not: just input the URI in the form below:

Give URI:

`http://ldf.fi/finlex/laki/p35`

Format: Turtle

Get RDF

SPARQL Endpoint

Queries are represented using the URI template: `http://ldf.fi/SERVICE/sparql?query=QUERY`

Query Test Form

```
# Find properties and their values for a given resource

SELECT * WHERE {
 <http://ldf.fi/finlex/laki/p35> ?property ?value
 }
```

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.

- Finlex Legislation
- Finlex Case-law

SPARQL Service Description

Get the description in RDF. (**Notice**: A service description is not available for all datasets - then empty despription is returned.) LDF uses W3C SPARQL Service Description recommendation for this.

Visualization

Visualizations for this dataset are available here.

License

SeCo semantic computing · Aalto University · University of Helsinki · Tekes

SeCo semantic computing

Documentation

- Schema-based Documentation
  - Classes, properties, domains, ranges, cardinalities
  - Automatic documentation generator can be used
    - Live OWL Documentation Environment LODE http://www.essepuntato.it/lode used at the moment
- Vocabulary Usage
    - We developed a new service: **http://vocab.at**
    - Provides a detailed analysis of vocabularies and their usage in a dataset
    - Points out some quality issues

Table of contents

1. Introduction
2. Vocabularies in use
3. Classes in use
4. Properties in use
5. Dataset linking
6. Assessment
7. Issues
8. References

Applications: BookSampo of Finnish Public Libraries (http://www.kirjasampo.fi)

Example: Booksampo Linked Data
– Finnish Fiction Literature on the Semantic Web

Table 1

Important classes in BookSampo along with their instance counts

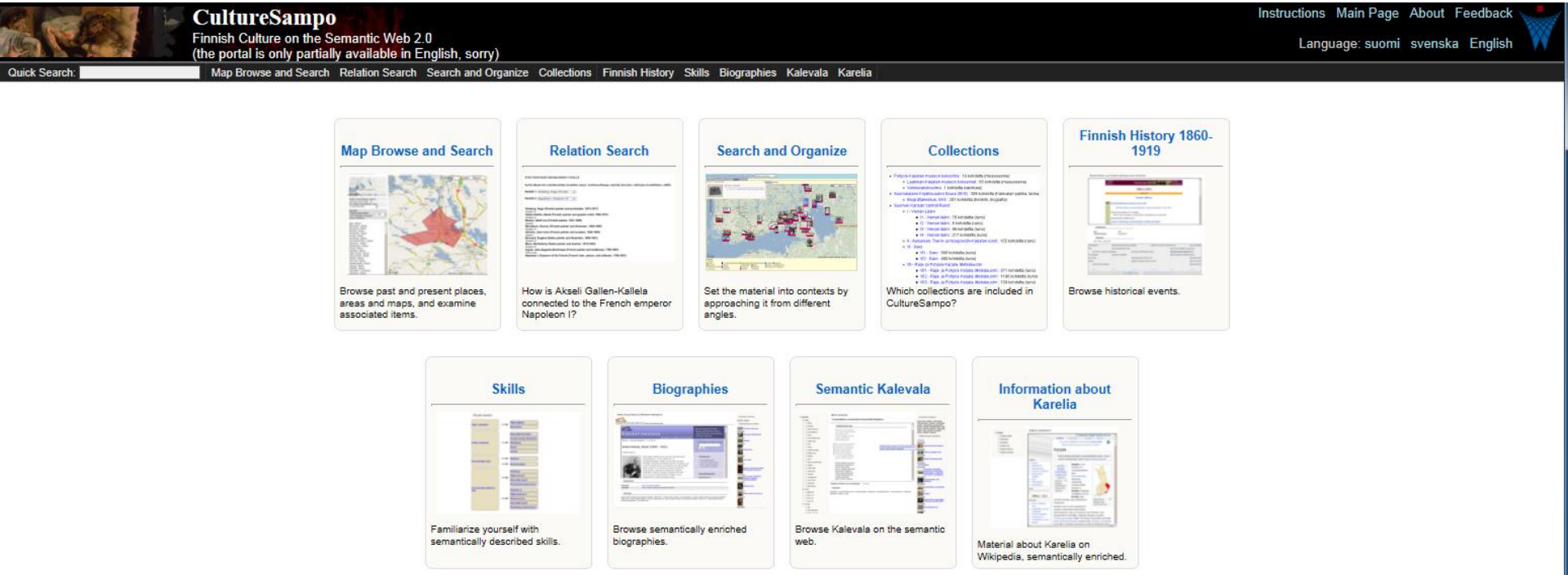
Type	Number of Instances
Literary Works	93,000
Editions	127,000
Book Covers	27,000
Fictional Characters	19,000
Contemporary Reviews	15,000
Weblinks	10,000
Literary Series	2,900
Literary Awards	2,700
Literary Award Series	200
Movies	1,100
People (e.g. Authors)	29,000
Author's Pictures	2,600
Publishers	2,600

Semantic Web, (Mäkelä, Hypen, Hyvönen, 2012)

CultureSampo – Finnish Culture on the Semantic Web 2.0

Aalto University

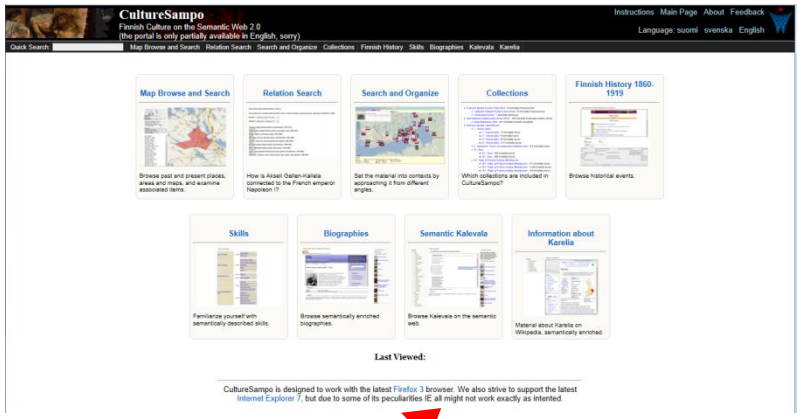
HELSINGIN YLIOPISTO

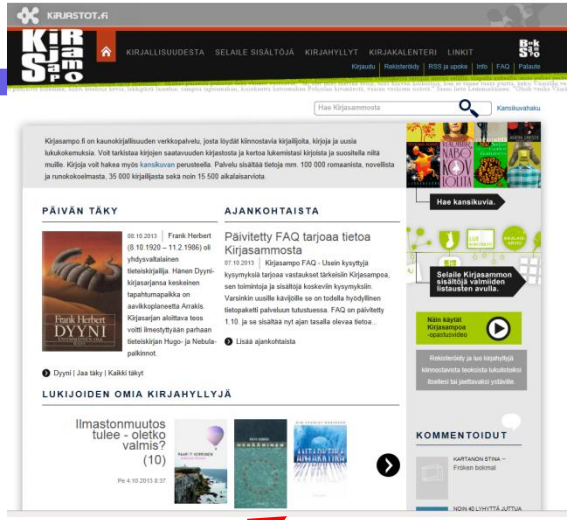


SeCo semantic computing

Using Linked Data in Applications

Aalto University

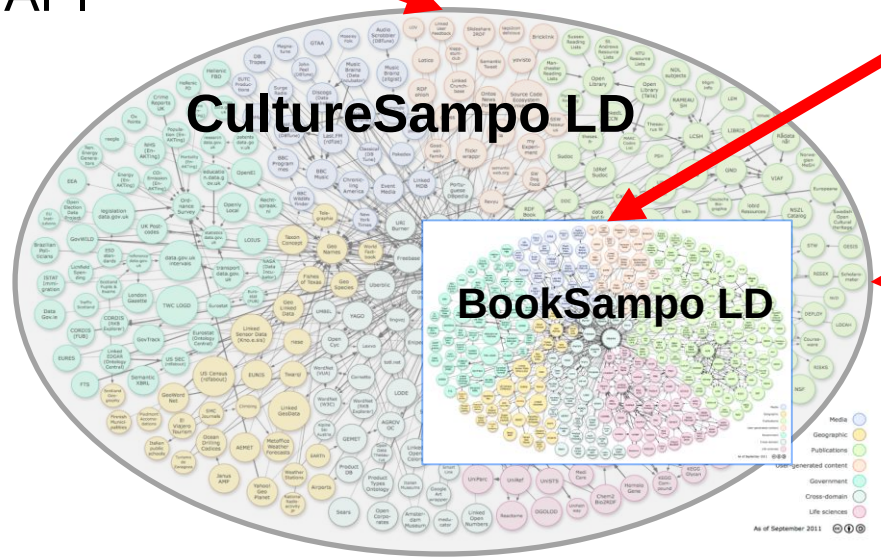




API

API

API



YOUR NEXT APPLICATION

SeCo semantic computing

Application Case: Linked Science

- Tiira.fi bird observations (1,25 million)
  - Bird Life Finland, GBIF
- Halias Bird Observation Station Data (30 years)
  - Finnish Museum of Natural History (0.5 million observations)
  - Finnish Meteorological Institute (100 000 observations)
- Bird Species Ontology (AVIO http://onki.fi/)
- Demonstrators
  - Mobile BirdWatch (Citizen Science)
  - Analysing / visualizing observations (Science)

BirdWatch—Supporting Citizen Scientists for Better Linked Data Quality for Biodiversity Management

Aalto University

HELSINGIN YLIOPISTO

http://www.seco.tkk.fi/applications/birdwatch/

Eero Hyvönen, Miika Alonen, Mikko Koho, and Jouni Tuominen

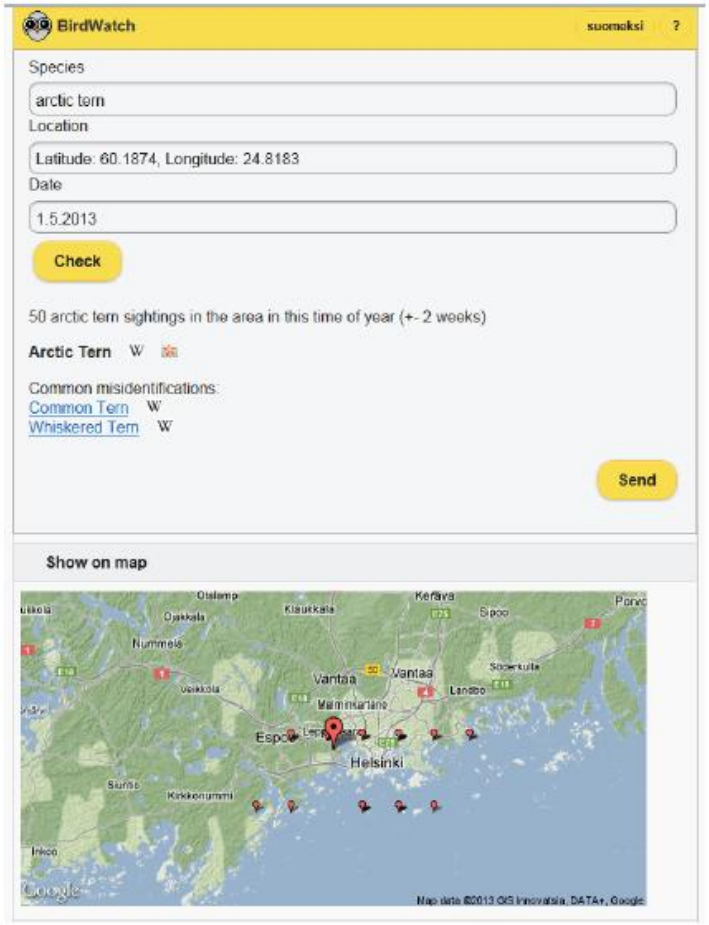


Fig. 1. Screenshot with a map showing the current position and related observations.

Fig. 2. Screenshot with statistical info related to the observations.

SeCo
semantic computing

Linked Open Universities
Linked Open Aalto: http://data.aalto.fi

Aalto University

HELSINGIN YLIOPISTO

Aalto University

- Home
- Mission
- Team
- Contact

Data

- Courses
- People
- Publications
- Places
- Projects
- Organization
- Events

- Visualizations
- Statistics
- SPARQL Endpoint
- Playground
- Licenses

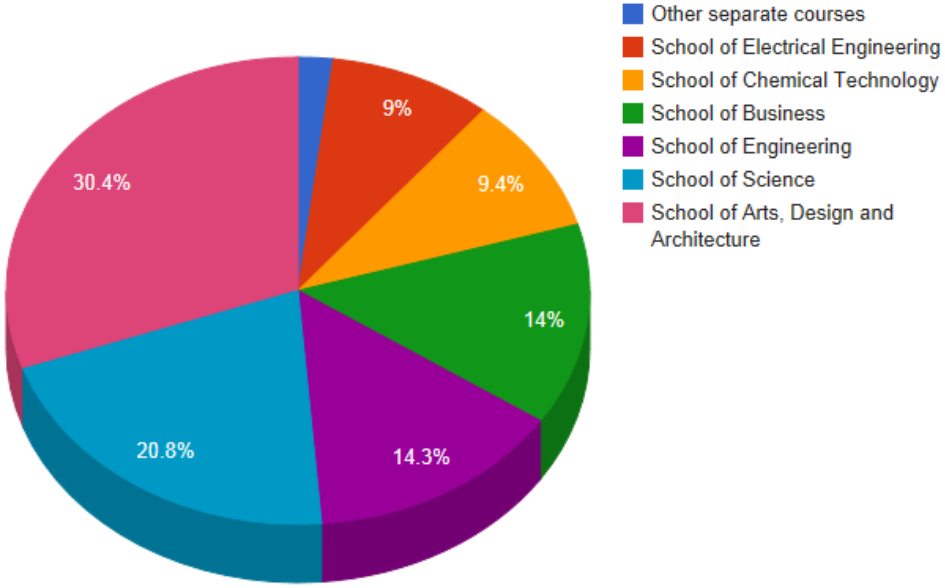
Linked Open Aalto Data Service

Opening Aalto University Content for Applications

Linked Open Aalto Visualizations

These visualizations represent the linked data that has been converted from various legacy sources. Visualizations have been made with the Linked Open Aalto Visualization Playground that can create visualizations on the fly with the Google Chart Editor.

Number of courses offered by schools of Aalto University



Summary

- Linked Open Data: Revolution on the Web
  - Web of Data for Machines and Applications
- Eco System Includes
  - Data Publishers (often Public Organizations)
  - Application Developers (e.g. Companies)
- Support and Tools are Needed
  - For Producing and Publishing Data
  - For Re-using the Data as Services
- LDF.fi is Being Built to Satisfy These Needs
- More Information
  - LDF.fi: http://www.ldf.fi/
  - Semantic Computing Research Group SeCo: http://www.seco.tkk.fi