# DARIAH-EU

## CLARIN
Common Language Resources and
Technology Infrastructure

# FIN-CLARIAH

Eero Hyvönen
Aalto & HELDIG
Mikko Koho
Aalto
Jouni Tuominen
Aalto & HSSH & HELDIG
Minna Tamper
Aalto & HELDIG

# WP 4.2 Linked Open Data Services
https://seco.cs.aalto.fi/projects/fin-clariah/

## What is FIN-CLARIAH?

FIN-CLARIAH (2022–) is the premier Finnish digital research infrastructure for Social Sciences and Humanities (SSH) comprising two components,

1. FIN-CLARIN (Finnish dimension of the Pan-European CLARIN infrastructure) and
2. DARIAH-FI (Finnish collaborations with the pan-European DARIAH infrastructure).

## Our Mission

The Aalto work in FIN-CLARIAH focuses on maintaining and developing the Linked Open Data Infrastructure for Digital Humanities in Finland (LODI4DH) in collaboration with the University of Helsinki (HELDIG, Faculty of Arts) and other partners within the DARIAH-FI part of FIN-CLARIAH. This includes also work on language infrastructures for spinning the Semantic Web and collaborations with FIN-CLARIN and CLARIN-EU. Our work is part of the cooperative partnership agreement between Aalto and DARIAH-EU.

## Components

Figure 1 shows elements that are needed in a national Semantic Web infrastructure according to our experiences. The infrastructure is based on domain agnostic W3C Web Standards and Best Practices of publishing Linked Data.

Our work provides the research community and the larger audience, including memory institutions, citizens, media, and companies with following services:

1. Ontology services
2. Linked open data services
3. Tools for extracting knowledge from texts
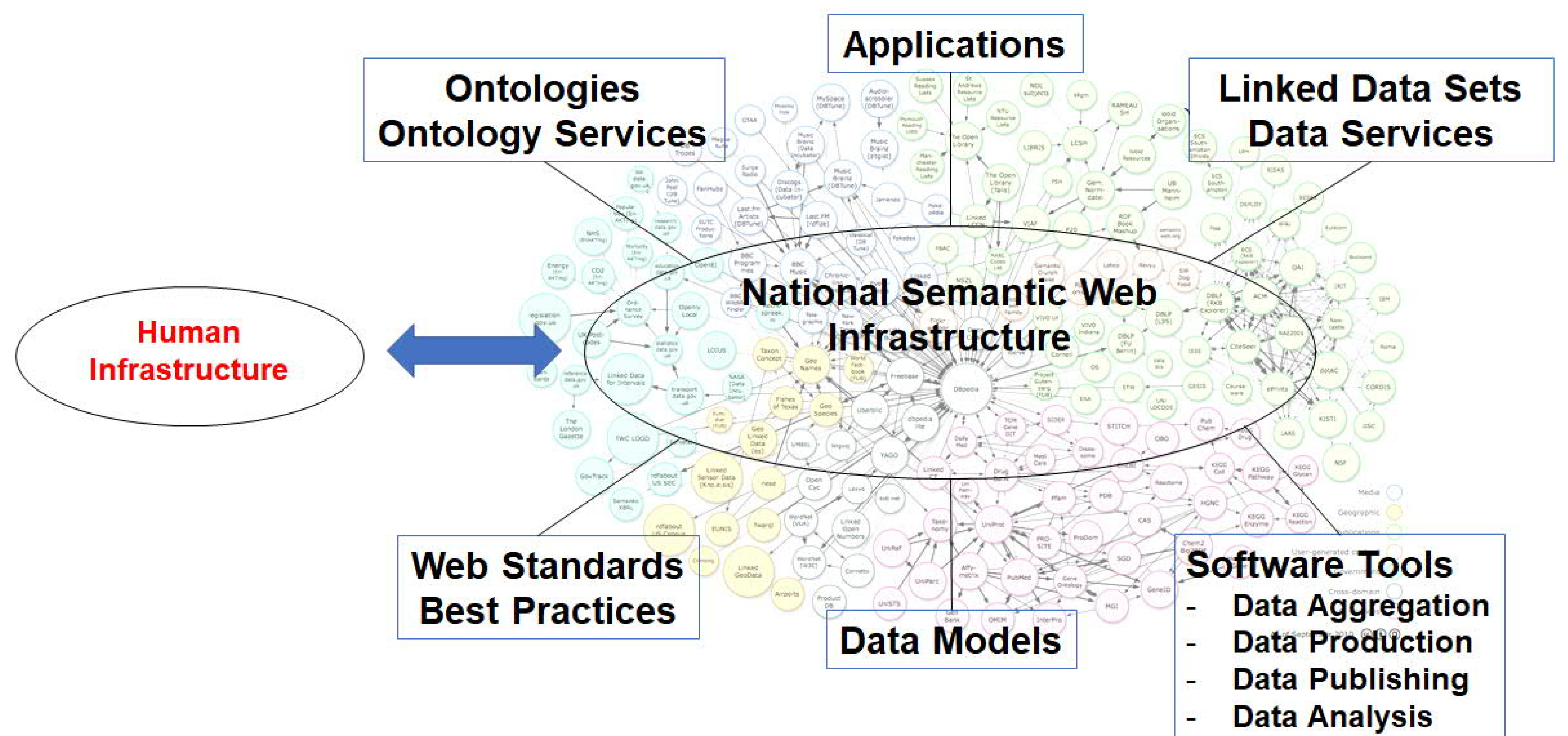4. Sampo portals for Digital Humanities



Figure 1: Components of National Linked Open Data Infrastructure

## Supporting Infrastructure

Our work in FIN-CLARIAH falls into several areas that need to be covered in order to create data and services for DH research:

Speech2Text. Tooling for creating textual time-stamped representations of videos and audio recordings. Here the goal is, e.g., to facilitate preservation of intangible cultural heritage and easy access to it.

Image2Text. OCR services developed, e.g., for the historical minutes of the Parliament of Finland in the ParliamentSampo system.

Text2Knowledge. Finnish language toolkit & web services for linked data knowledge extraction from unstructured Finnish texts, including named entity recognition and linking, automatic keyword annotation, relation extraction, and semantic labeling.

Knowledge2DataAnalysis. Reusable tooling for Digital Humanities on top of a linked data service and SPARQL endpoint, as used in various Sampo systems.

DataAnalysis2AI. Tooling for knowledge discovery and computational creativity. Here the machine is seen as an intelligent agent searching itself for interesting patterns in knowledge graphs, solving problems, and even explaining the results to the human user.

Infrastructure components to be maintained and built in our part of the FIN-CLARIAH initiative include:

ONKI ontology services (ONKI.fi) for history, extending the Finto.fi services of the National Library. This work comprises ontologies for historical persons, places, events, times, occupations, and names.

Historical map services (Hipla.fi). Here historical maps can be aligned with contemporary ones and used as layers in applications, based on the MapWarper tool and linked data for storing related metadata.

Linked Data Finland (LDF.fi). This platform is used for publishing linked data as services using the standards and best practices of W3C. Our focus here is on using the "7-star" model, extending the classic Tim Berners-Lee's 5-star model, for better reusability and quality of linked datasets.

Natural language processing toolkit and services for extracting linked data.

Learning materials Providing the DH community with more educational online material on using linked data, such as developing the Linked Data School Linda.

Maintaining the Sampo Series of Linked Open Data services and semantic portals in use in Finland and the Sampo-UI framework for developing Sampo applications.