

Sampo Systems Infrastructure Data Services & Portals

Aalto University and University of Helsinki
Helsinki Centre for Digital Humanities (HELDIG)

SEMANTIC WEB INFRASTRUCTURE

The Aalto work (WP 3.2 Data Ingestion) is related to maintaining and developing further the national **Linked Open Data (LOD) Infrastructure for Digital Humanities (DH) in Finland** [1]. Its components are depicted below.

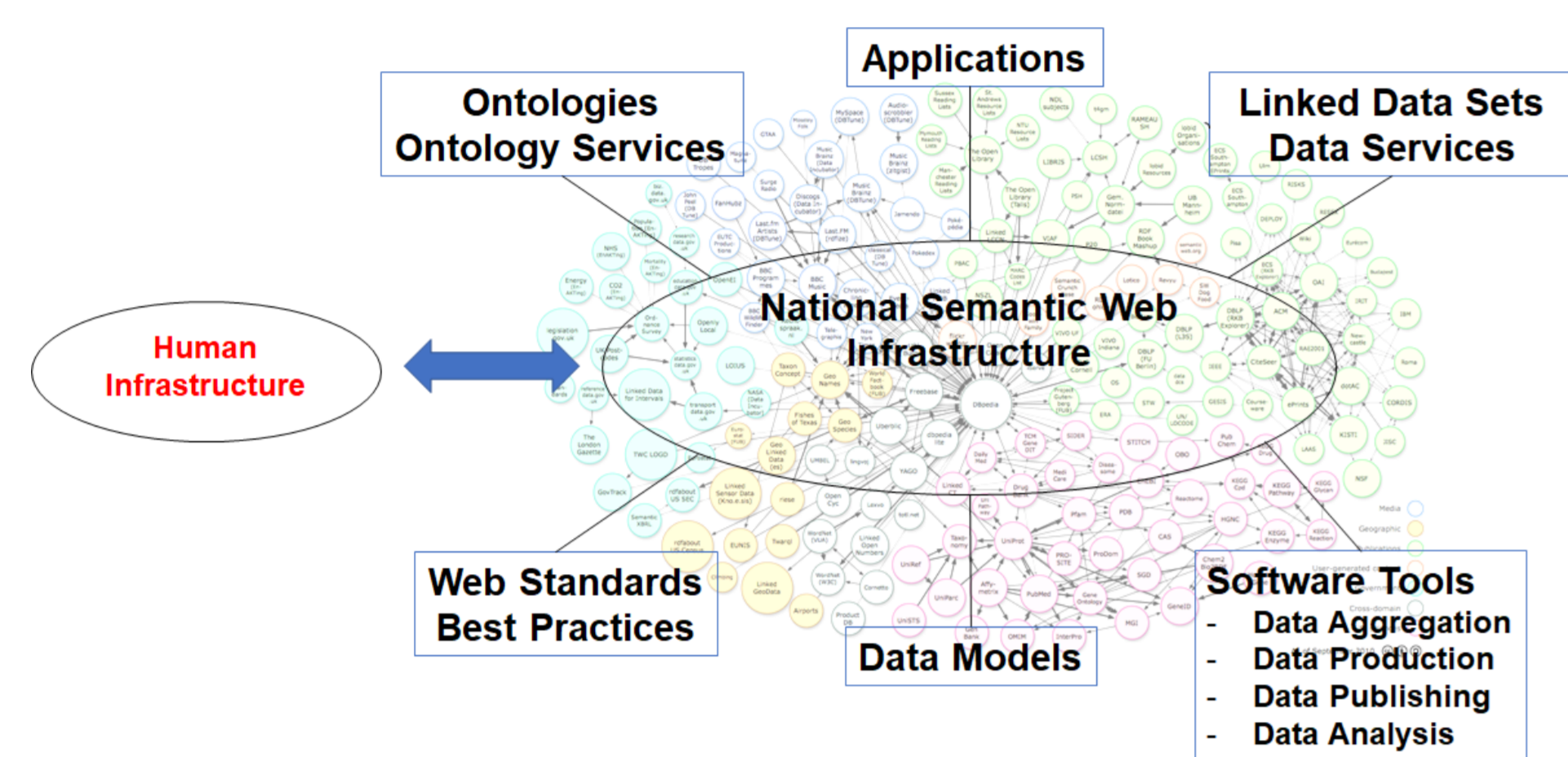


Figure 1: Components of the national semantic web infrastructure

The work falls into several R&D areas in order to create LOD and web services for DH research:

Speech2Text. Tooling for creating textual time-stamped semantic descriptions of video and audio recordings, e.g., in the WarMementoSampo system.
Image2Text. OCR services for, e.g., the historical plenary session minutes of the Parliament of Finland in the ParliamentSampo system.

Text2Knowledge. Extracting and linking entities, keywords, classifications, and linked data from unstructured Finnish texts in, e.g., the BiographySampo system.

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

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

WEB SERVICES

The infrastructure components are built and served as data and web services:

Ontology services (ONKI.fi, light.onki.fi) Extending the Finto.fi services of the National Library. This work comprises ontologies for, e.g., historical persons, places, events, times, occupations, and person 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 LOD 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 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 tools and services for extracting linked data from texts.

User interface tools (Sampo-UI) Making application development and using portals as easy as possible.

Learning materials Providing the Digital Humanities community with online educational materials on using linked data.

SAMPO LINKED OPEN DATA SERVICES & PORTALS

We develop and maintain the Sampo series of some 20 linked open data services and semantic portals in use in Finland [2], including, e.g.:

NameSampo *Toponomastic research* (main data owners: Finnish Institute of Languages, National Survey)

BiographySampo *Biography and prosopography* (main data owners: Finnish Literature Society (SKS) and others)

WarSampo, WarVictimSampo 1914–1922, and WarMementoSampo *Military history* (main data owners: National Archives, Defence Forces, Tammenlehvän Perinneliitto ry, and others)

AcademySampo *Academic history* (main data owners: University of Helsinki Archives, National Archives, and Wikidata)

FindSampo, CoinSampo and PASampo *Archaeology, citizen science* (main data owners: Finnish Heritage Agency, National Museum, and British Museum (UK))

Mapping Manuscript Migrations Sampo *Manuscript studies* (main data owners: Oxford University (UK), Schoenberg Institute (US), and IRHT (Paris))

LetterSampo and Kirjesampo 1809-1917 *Epistolary studies* (main data owners: Huygens Institute (NL), Berlin-Brandenburg Academy of Sciences (D), Oxford University (UK), Finnish archives, libraries, and museums)

LawSampo *Legal informatics* (main data owners: Ministry of Justice and Edita Publishing Ltd)

ParliamentSampo *Parliamentary studies* (main data owners: Parliament of Finland, Finnish Literature Society, and Wikidata)

Operasampo *Historical music performances* (main data owner: Sibelius Academy and Wikidata)

ConfermentSampo *Academic history* (Main data owner: University of Helsinki, Finna.fi, and Wikidata)

ArtSampo *Art history* (Main data owner: National Gallery and other art museums)

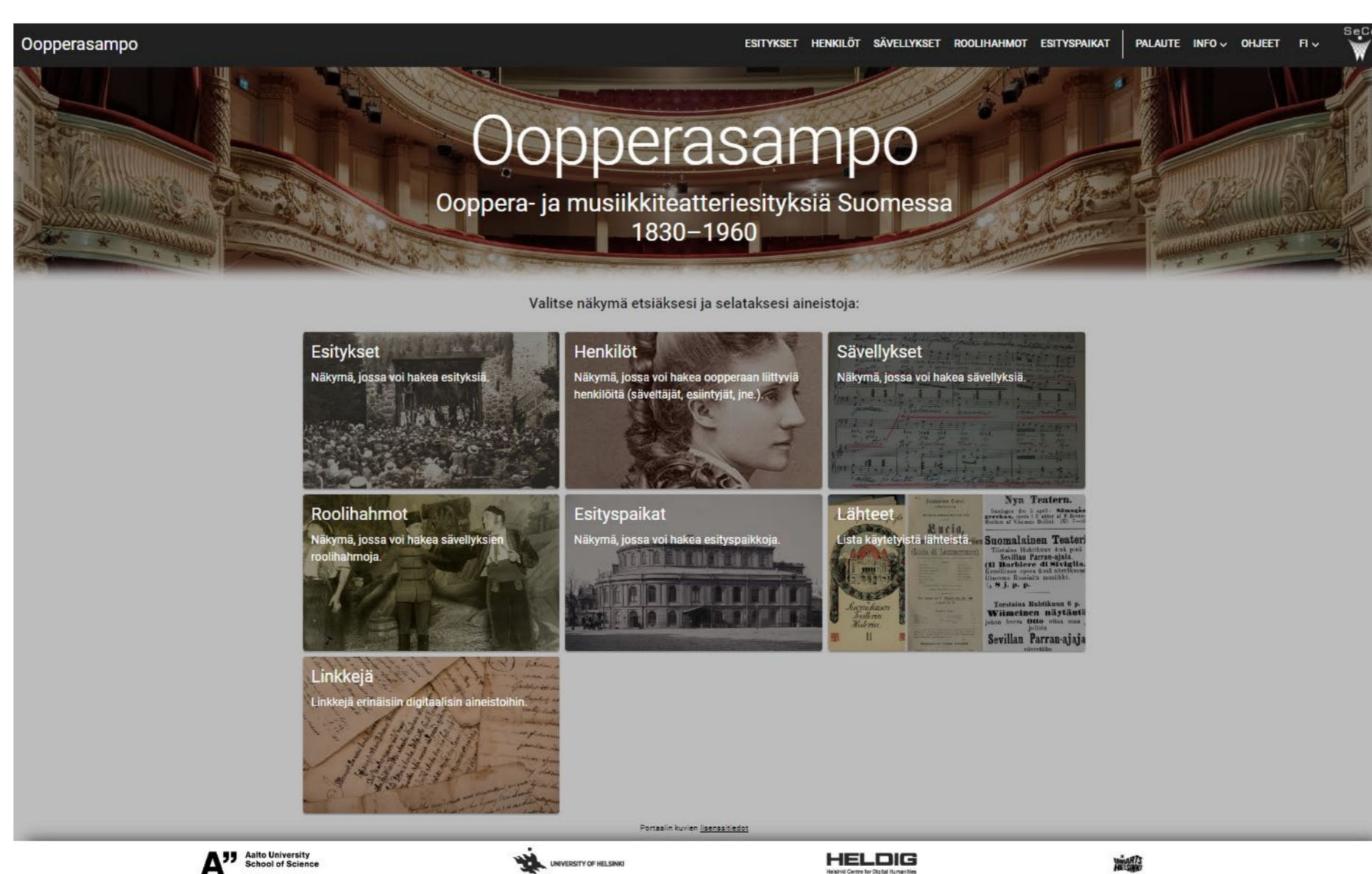


Figure 2: OperaSampo portal landing page with seven views into the underlying LOD service at Linked Data Finland LDF.fi

SAMPO-UI

Sampo-UI is a framework available in Github for creating Sampo portals on top of SPARQL data services. Sampo-UI “standardizes” user interfaces to help both end users and portal developers [3,4]. The framework includes a semantic faceted search engine integrated seamlessly with data analytic tools for DH research. Sampo-UI has been used by external researchers and companies, too.

SAMPOSAMPO

Data alignment services are being developed for mapping mutually related Sampo datasets onto each other and external datasets and services, such as Finna, Wikidata, etc. On top of the service, a sampo portal covering all Sampos, a.k.a., “SampoSampo”, is being developed.

MORE INFO

Aalto FIN-CLARIAH/DARIAH-FI project homepage:

<https://seco.cs.aalto.fi/projects/fin-clariah/>

Homepage of all Sampo systems, including links, videos, and publications:

<https://seco.cs.aalto.fi/applications/sampo/>

Overview of the Finnish semantic web infra:

[1] 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.

Overview of the Sampo model and systems:

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

Articles about the Sampo-UI framework:

[3] 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, pp. 69-84, 2022.

[4] Heikki Rantala, Annastiina Ahola, Esko Ikkala, and Eero Hyvönen: How to create easily a data analytic semantic portal on top of a SPARQL endpoint: introducing the configurable Sampo-UI framework. *VOILA! 2023 Visualization and Interaction for Ontologies, Linked Data and Knowledge Graphs 2023*, CEUR Workshop Proceedings, Vol. 3508, 2023.

Over 500 publications of the Semantic Computing Research Group are available online:

<https://seco.cs.aalto.fi/publications/>