

Linked Data Finland

Yhdistetty avoin tieto Suomessa



Miten data kannattaa avata verkossa?

Aalto-yliopiston Semanttisen laskennan tutkimusryhmä SeCo kehittää ja ylläpitää prototyyppiä kansallisesta avoimen linkitetyn datan palvelusta Linked Data Finland¹. LDF.fi vastaa kysymykseen miten data kannattaa avata, jotta sitä voidaan rikastaa verkon muilla tiedoilla ja hyödyntää mahdollisimman helposti. Tavoitteena on edistää tietoaineistojen yhteentoimivuutta, julkaisemista, uudelleenkäyttöä ja sovelluskehitystä. Tutkimustyötä on sen eri vaiheissa rahoittanut Tekes ja n. 50 kotimaista organisaatiota ja yritystä kansallisessa FinnONTO-hankesarjassa (2003-2012)², sitä jatkaneessa Linked Data Finland -hankkeessa (2013-) sekä useissa pienemmissä hankkeissa.

Kansallinen verkon tietoinfrastruktuuri

LDF-palvelun taustalla on FinnONTO-hankkeen visio avoimesta, kansallisesta tietoinfrastruktuurista verkossa osana kansainvälistä webin kehitystä. Ideana on tuottaa ja julkaista avointa tietoa yhteisöllisesti yhteentoimivalla tavalla siten, että kertaalleen tuotettua tietoa ei tarvitse tuottaa moneen kertaan eri organisaatioiden toimesta ja että yhden julkaisijan tietoa voidaan rikastaa automaattisesti toisten julkaisijoiden tiedoilla. Tiedon tuottamista, julkaisemista ja uudelleenkäyttöä tuetaan verkkopalveluiden avulla. Tämä on teknisesti mahdollista WWW:n infrastruktuuria koordinoivan W3C-järjestön semanttisten Linked Data -standardien ja käytäntöjen avulla.



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.

[Home](#)

[Project](#)

[Datasets](#)

[Schemas](#)

[Services](#)

[Policies](#)

[Documentation](#)

[Validation](#)

[Applications](#)

[Your Data?](#)

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.

¹ Palvelu on käytettävissä osoitteessa <http://www.ldf.fi>

² Ks. Tieteessä tapahtuu -lehden artikkeli <http://ojs.tsv.fi/index.php/tt/article/view/41559>

Linkitetyn datan ja ontologioiden palvelu

Infrastruktuurin ytimessä on uudentyypinen yhdistettyjen tietoaisteiden ”7 tähden” julkaisualusta sekä aineistojen tuottamisessa ja linkittämisessä tarvittavat palvelut, kuten kansallinen ontologiapalvelu ONKI ja Kansalliskirjaston ONKI:sta tuotteistama Finto, niitä varten kehitetyt ontologiset sanastot, kieliteknologiset verkkopalvelut ja muut työkalut datan tuotantoon, analyysiin ja visualisointiin. Hankkeen vision mukaan keskeiset datajulkaisut, kuten linkitetty KOKO-ontologiapilvi, Wikipediasta louhittu DBpedia, paikka-, toimija- ja tapahtumaontologiat sekä niihin liittyvät sanastot muodostavat semanttisen webin ”selkärangan”, jonka avulla verkon tietojoukot voidaan yhdistää samaan tapaan kuin kansainvälisessä Linked Open Data -pilvessä.

Tavoitteena sovelluskehittäjien tukeminen

Yhdistetyn avoimen tiedon data- ja sanastopalvelut on tarkoitettu tietotyön ammattilaisille ja ohjelmistojen kehittäjille, jotta netin käyttäjille, niin kansalaisille, tutkijoille, kuin toisille verkkopalveluillekin, voidaan luoda sisällöllisesti rikkaampia ja älykkäämpiä sovelluksia. Palveluita on testattu ja testataan kehittämällä useita pilottisovelluksia loppukäyttäjille. Tällä hetkellä työn alla on mm. ”Sotasampo”, hankekokonaisuus, jonka tavoitteena on avata ja linkittää toisiinsa Suomen itsenäisyyden 100-vuotisjuhliin liittyen toisen maailmansodan aineistoja hieman vastaavaan tapaan kuin SeCo-tutkimusryhmän aiemmissa ”Sampo”-perheen prototyypeissä Kulttuurisampo, Matkailusampo ja Kirjasampo. Kansallisessa historiallisten paikkojen sanastopalvelu -hankkeessa luodaan yhteistä verkkopalvelua muistiorganisaatioille paikatiedosta ja historiallisista kartoista. Kansainvälistä Linked Data -alueen yhteistyötä ja tutkijavaihtoa on meneillään mm. Stanfordin, Coloradon, Oxfordin ja Bremenin yliopistojen kanssa liittyen mm. valistuksen ajan filosofien kirjeenvaihtoon ja tieteellisen tutkimusdatan julkaisemiseen (Linked Science).

Kotimaisia linkitetyn datan julkaisuja

Linked Data Finland -palvelussa on sovelluskäyttöön julkaistu mm. ensimmäistä kertaa Suomen lainsäädäntö avoimena datana (oikeusministeriön Finlex-tietokannasta), yleisten kirjastojen Kirjasampo-data, suomenkielisen Wikipedian semanttinen verkko (DBpedia) datapalveluna, Suomalaisen Kirjallisuuden Seuran Kansallisbiografian yli 6000 elämäkertaa verkkona, yli miljoona lintuhavaintoa Tiira.fi-palvelusta, Semanttinen Kalevala, museoiden kokoelmia, säädätää Ilmatieteen laitokselta ja Aallon omia avoimia data-aineistoja (Linked Open Aalto -palvelu³). Myös ONKI-palvelun uusin versio sanastoineen perustuu LDF.fi-palveluun. LDF-teknologiaan perustuvilla sovelluksilla on Suomessa kymmeniä tuhansia loppukäyttäjiä kuukaudessa.

Lisätietoja

Linked Data Finland -portaali:

<http://www.ldf.fi>

Avoin Suomi 2014 menestystarina:

<https://www.youtube.com/watch?v=tLjk-4w-SSA>

Linked Data Finland -hanke:

<http://www.seco.tkk.fi/projects/ldf>

FInnONTO-hanke:

<http://www.seco.tkk.fi/projects/finnonto>

SeCo-tutkimusryhmä:

<http://www.seco.tkk.fi>

Prof. Eero Hyvönen

eero.hyvonen@aalto.fi

Puh. 050 3841618

Semanttisen laskennan tutkimusryhmä SeCo

Aalto-yliopisto, perustieteiden korkeakoulu



³ <http://data.aalto.fi/>