



Linked Data Finland: visio, missio ja toteutus

Eero Hyvönen
professori

Semantic Computing Research Group (SeCo)
Aalto-yliopisto ja Helsingin yliopisto

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

Sisältö

- Tausta
- Visio
- Missio
- Toteutus
- Jatkonäkymät

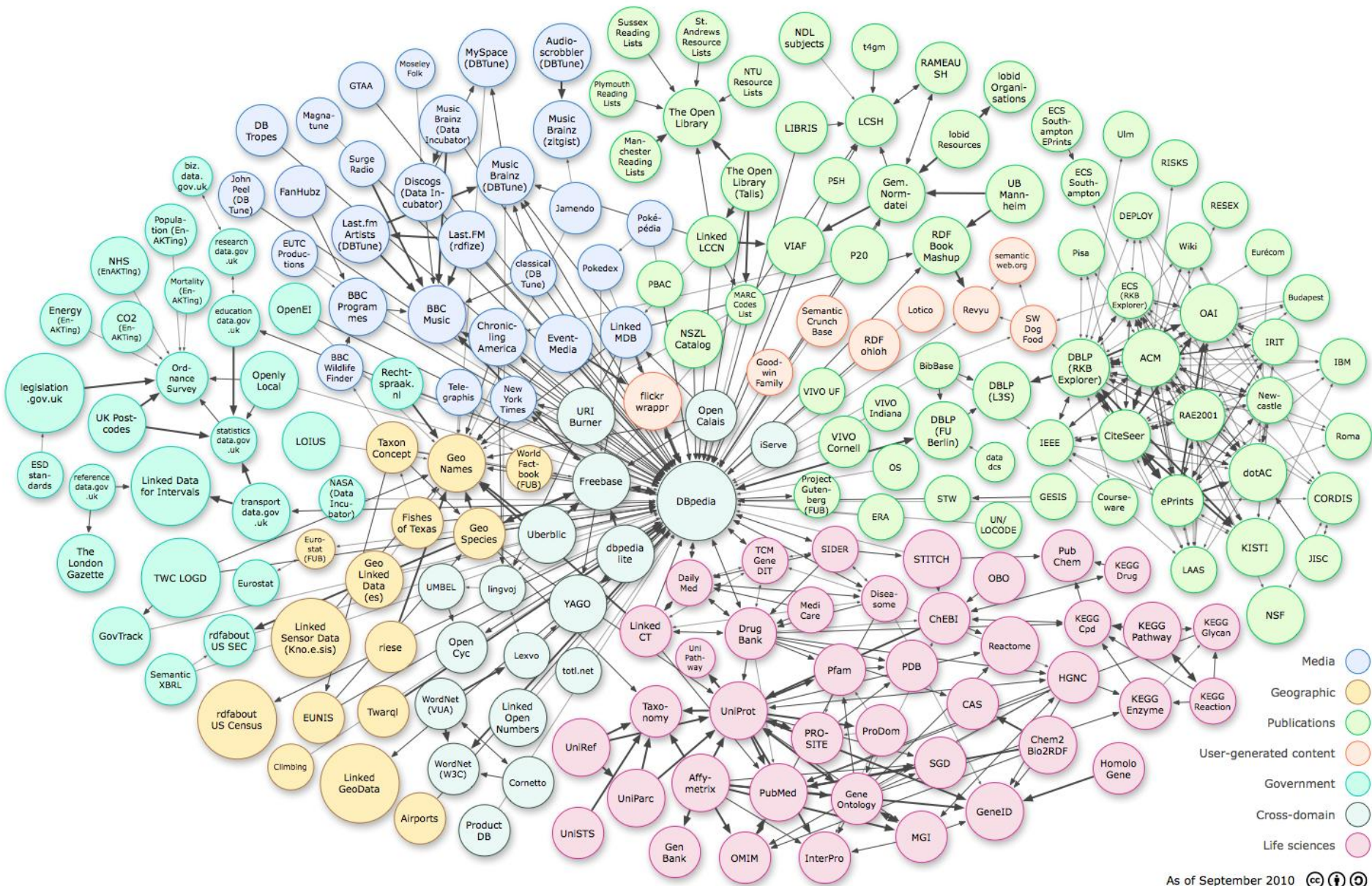
Linked Data Finland: Hankkeen taustalla

- Kansainvälinen kuohunta ja trendit
 - Open Data, Linked Open Data, Web of Data, Big Data
- Kansallinen **tahtotila**: Data avoimeksi!
 - Julkisen sektorin Open Data –hankkeet
 - » UK, USA, Australia, ...
 - Suomessakin useita hankkeita ja ohjelmia
 - » FinnONTO, HRI, ML, IL, KIDE, TTA, YLE, VM, ...
- Kansallinen **toivetila**: Sovelluksia, liiketoimintaa!
 - Datastako perusta uuden Nokian syntymiselle?
 - Esim. Digilen Data to Intelligence –ohjelma
 - Apps4Finland ja lukuisia avoimen datan yhteisöjä
- **Mitä tutkimuksen saralla pitäisi tehdä?**

- **Hyvä uutinen:** tiedon avaaminen
 - Vaikka tässäkin tietysti on pidettävä järki mukana
- **Huono uutinen:** tietoa ei avata parhaalla tavalla
 - Avoimen tiedon käyttö sovelluksissa on hankalaa
 - Tietoa julkaistaan ”siiloissa”: data ei yhdisty toisiin tietojulkaisuihin
- **Visio**
 - Tietoa kannattaa avata yhteisesti sovitulla tavalla
 - » W3C:n standardit ja
 - » Linked Data –periaatteet
 - Kansallisen tason yhteistyötä tarvitaan

Linked Open Data Cloud:

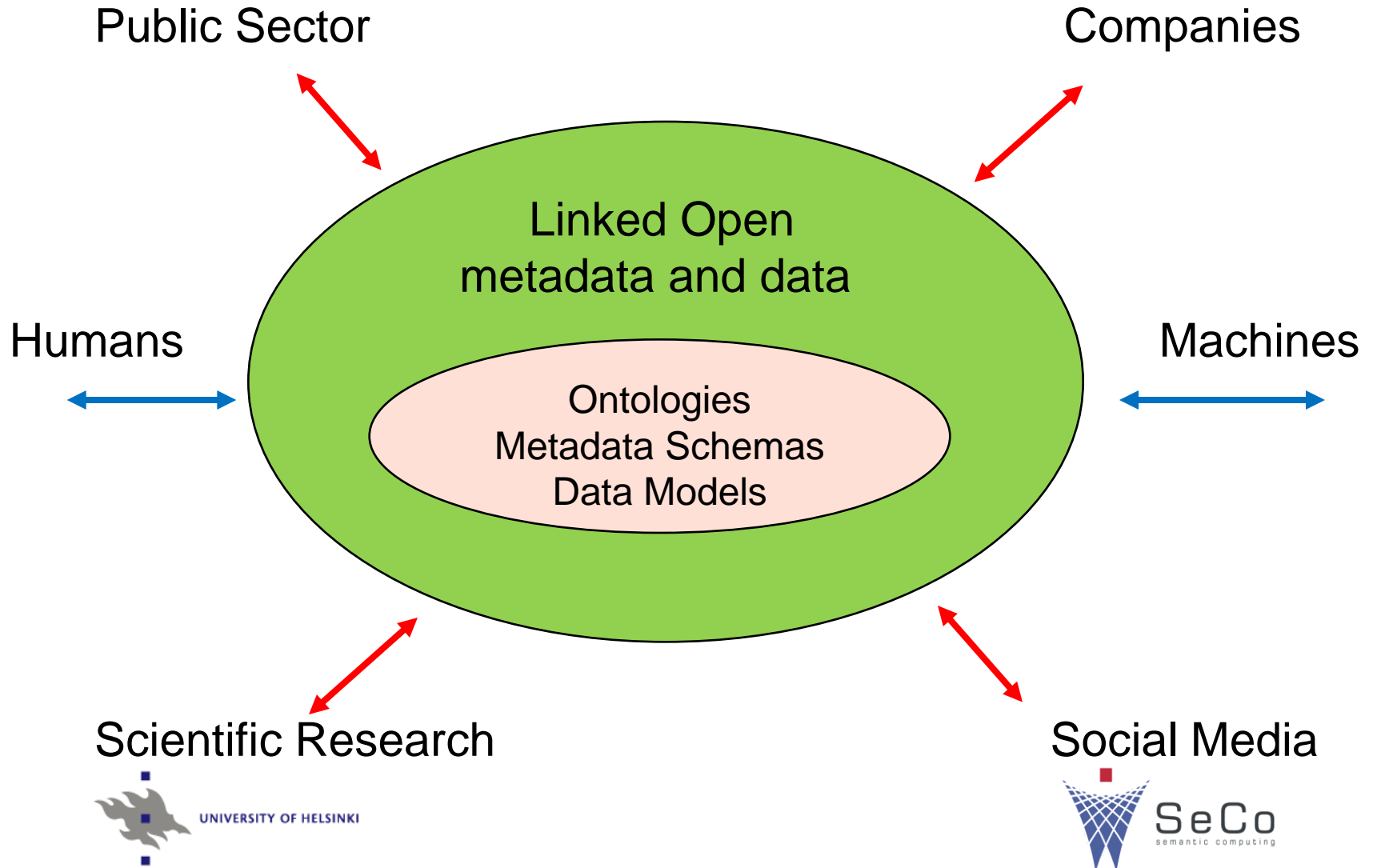
<http://linkeddata.org>



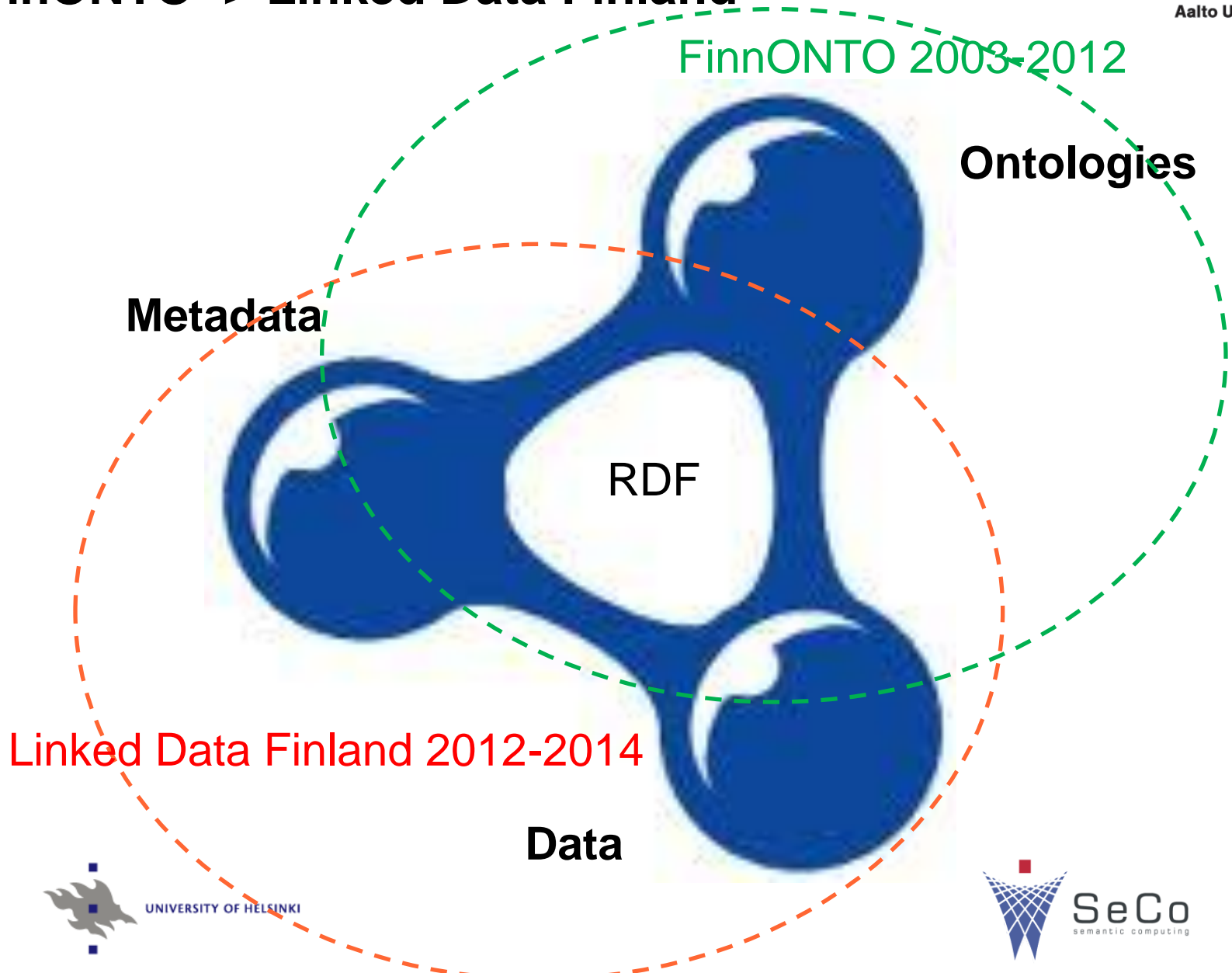
Missio

- Kerätään yhteen laaja kansallinen konsortio
 - Tässä voitiin hyödyntää FinnONTO-hankkeen (2003-2012) piirissä syntynyttä yhteisöä
- Mukana sekä julkinen sektori että yritykset
 - Julkinen sektori yleensä tuottaa avointa tietoa
 - Yrityksen hyödyntävät sitä sovelluksissa
- Hankkeen ”ydinosassa” alasta riippumaton työ
 - tutkitaan yhteisiä asioita ja
 - kehitetään yhteisiä ratkaisuja
- Alakohtaisissa ”työpajoissa”
 - Pilotoidaan ratkaisuja case-tutkimuksessa

Linked (Open) Data Ecosystem



FinnONTO -> Linked Data Finland

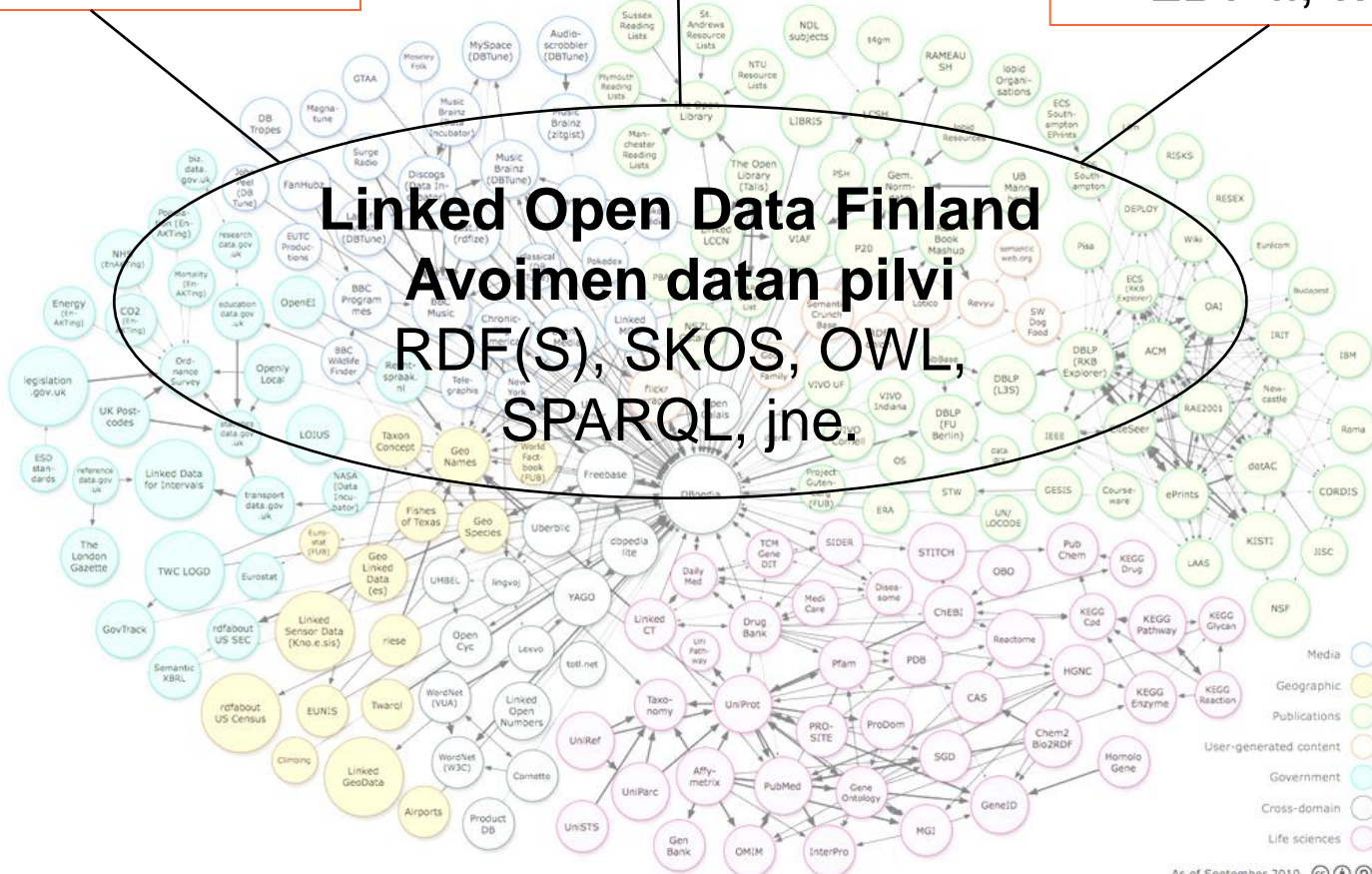


Linked Open Data Finland: Malli pähkinänkuoressa

Sovellukset

Ontologiapalvelut
ONKI, Finto, ...

Datapalvelut
LDF.fi, ...



Toteutus

Linked Data Finland –hankeyhteisö (hankkeen käynnistyessä 2012)

Johtoryhmän kokoonpano (3.2.2012)		
Yritys / organisaatio	Jäsen	Varajäsen
1 Citynomadi Oy	Merja Taipaleenmäki	
2 Connexor Oy	Pasi Tapanainen	Mirka Tapanainen
3 CSC - Tieteen tietotekniikan keskus Oy	Pirjo-Leena Forsström	
4 Edita Publishing Oy	Jari Linhala	Päivi Helander
5 Lean Development Oy	Mikko Mäkelä	
6 Sanastokeskus TSK	Katri Seppälä	
7 Sanoma News Oy	Esa Mäkinen	
8 Suomalaisen Kirjallisuuden Seura SKS	Lauri Harvilahti	
9 Suomen Asiakastieto Oy	Pertti Vahermaa	
10 Svenska Litteratursällskapet SLS	Karola Söderman	
11 Talentum Media Oy	Stina Wikberg	
12 Tieto Oyj	Lasse Akselin	Heimo Hänninen
13 Yleisradio Oy	Sami Kallinen	
1 Oikeusministeriö	Aki Hietanen	Silja Rekomaa
2 Liikenne- ja viestintäministeriö	Taru Rastas	
3 Valtiovarainministeriö	Jukka Uusitalo	Anne Kauhanen-Simanainen
4 Helsingin kaupunginkirjasto	Erkki Lounasvuori	Matti Sarmela
5 Helsingin kaupunki	Mirjam Heikkinen	Leila Oravisto
6 Suomenlinnan hoitokunta	Maire Mattinen	
Asiantuntijajäsenet		
7 W3C Suomen toimisto	Ossi Nykänen	
Tutkimustahot		
8 Aalto-yliopisto	Eero Hyvönen	
9 Helsingin yliopisto	Mika Wahlroos	
Päärahoittaja		
10 Tekes	Keijo Silventoinen	

+ Kansalliskirjasto tuli mukaan v. 2013 casena ONKI:n tuotteistaminen Finto-palveluksi
- Tämä toteutettiin KK:n erillisessä projektissa yhteistyössä LDF:n kanssa

3 Tutkimusalueet ja -kysymykset

- 3.1 Metadatan kustannustehokas tuotanto
- 3.2 Metadatan laatu ja luotettavuus.....
- 3.3 Metadatatoukkojen yhdistäminen
- 3.4 Demonstraatiot hyötykäytöstä ja kriittinen arvio haasteista

Neljä pilottialuetta

- **Suomalainen lakitieto semanttisena LOD-palveluna**
 - Talentum Oy, Edita Oy, oikeusministeriö, liikenne ja viestintäministeriö, ...
- **Mediayrityksen sisällönhallinta**
 - Media-alan LOD-palvelut, datajournalismi, Semanttinen kansallisbiografia
 - YLE Oy, Helsingin Sanomat, SKS, SLS, Tieto Oyj, ...
- **Yritystietojen semanttinen rikastaminen uutisilla**
 - Asiakastieto Oy, YLE Oy, Helsingin Sanomat, ...
- **Semanttinen palvelukartta**
 - Citynomadi Oy, Helsingin kaupunki, Suomenlinna, ...
- **Lisäksi rinnakkaishanke: Linked Open Aalto (data.aalto.fi)**

Tiivistelmä LDF-hankkeesta

Lehdistötiedote 24.1.

Aalto-yliopisto, perustieteiden korkeakoulu

Kansallinen avoin tietoinfrastruktuuri sai sisällöllisen selkärangan

Aalto-yliopiston ja Helsingin yliopiston vetämän kansallisen Linked Data Finland -hankkeen tuloksena esiteltiin prototyyppi kansallisesta, yhdistetyn avoimen datan (Linked Open Data) tietoinfrastruktuurista Otaniemessä pidetyssä seminaarissa. Järjestelmän tavoitteena on edistää tietoaaineistojen avointa, yhteentoimivaa julkaisemista ja hyödyntämistä sovelluksissa, ja se perustuu WWW:n infrastruktuuria koordinoivan W3C-järjestön uusimpiin semanttisen webin standardeihin ja käytäntöihin.

Infrastruktuurin ytimessä on uudentyyppinen yhdistettyjen tietoaaineistojen julkaisupalvelu "Linked Data Finland", LDF.fi, sekä aineistojen tuottamisessa tarvittavat palvelut, erityisesti kansallinen ontologiapalvelu ONKI ja sen ontologiset sanastot. Keskeiset datajulkaisut, kuten Wikipediasta louhittu DBpedia, sekä niihin liittyvät sanastot muodostavat semanttisen webin "selkärangan", jonka avulla verkon tietojoukot voidaan yhdistää.

"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", toteaa hanketta johtanut professori Eero Hyvönen Aalto-yliopistosta. "Palveluita testattiin kehittämällä useita pilottisovelluksia loppukäyttäjille."

Tutkimushankkeeseen liittyi myös Kansalliskirjaston vetämä erillinen tuotteistamisprojekti, jossa Aallon ONKI-ontologiapalvelun prototyypistä julkaistiin jatkossa budjettivaroin ylläpidettävä Finto-palvelu. ONKI/Finto-järjestelmä perustuu Yleiseen Suomalaiseen Ontologiaan ja siihen liitettyyn 14 muuhun suomalaiseen ontologiaan, jotka muodostavat yhdistetyn datan pilven; kyseessä on ensimmäinen kansallisen tason monialainen ratkaisu maailmassa. Finto korvaa jatkossa tuotantokäytön osalta Aalto-yliopiston ONKI-prototyypin, jolla on ollut n. 14 000 käyttäjää kuukaudessa.

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 6000 elämäkertaa verkkona, yli miljoona lintuhavaintoa Tiira.fi-palvelusta, Semanttinen Kalevala, museoiden kokoelmia, säädataa Ilmatieteen laitokselta ja Aallon omia aineistoja (Linked Open Aalto). Myös ONKI-palvelun uusin versio sanastoineen perustuu LDF.fi-palveluun.

Linked Data Finland -hanketta rahoittivat Tekes ja yhteensä 20 yritystä ja julkista organisaatiota.

Jatkonäkymät

- Tekes-rahoitteinen osuus päättyi 31.12.2013
- Julkisrahoitteinen osuus päättyy 31.3.2014
- Linked Data Finland –työ on kuitenkin jatkuu
 - Määrä riippuu työlle saavasta rahoituksesta
 - Haasteena on tutkimustyön jääminen ”rahoitukselliseen mustaan aukkoon” Suomessa:
 - » Tekes rahoittaa yritysveoista tutkimusta, mutta tässä julkisen sektorin panos on tärkeä
 - » Suomen Akatemia rahoittaa (puhtaasti) akateemista työtä, mutta tässä on kyseessä soveltava tutkimus



- **Jos Sinulla on tutkimuksellinen ongelma, johon Linked Data voisi olla ratkaisu, ota yhteyttä!**



A!

Kiitos!

Aika	Esiintyjä	Esityksen aihe ja lisätietoja (muutokset mahdollisia)
09:00-9:15	Eero Hyvönen Aalto ja Helsingin yliopisto	Linked Data Finland: visio, missio ja toteutus
09:15-11:00	Datan tuotanto ja julkaiseminen	
	Ontologiat ja metatietomallit	
	Matias Frosterus Aalto ja Kansalliskirjasto	KOKO-ontologiapilvi ja sen ylläpito
	Jouni Tuominen Aalto	ONKIn instanssiontologiat: toimijat, paikat, tapahtumat, biologian taksonomiat
	Osma Suominen Aalto ja Kansalliskirjasto	Julkisen hallinnon palvelujen tietomalli ja ryhmittely verkkopalveluissa (JHS 183)
	Esko Ikkala Aalto	Suomen historia ontologiana – CIDOC CRM -perustainen kulttuuridatan harmonisointi
	Automaattinen annotointi	
	Eetu Mäkelä Aalto	Kieliteknologiset apuvälineet: SeCo Lexical Analysis -palvelu Automaattinen annotointi ARPA-palvelulla
	Datan laatu	
	Miika Alonen Aalto	Datan analysointi ja validointi: Vocab.at-palvelu
	Linked Data Finland -julkaisualusta	
	Eero Hyvönen Aalto ja Helsingin yliopisto	7 tähden yhdistetyn tiedon julkaisualusta LDF.fi
11:00-12:00	Lounastauko (omakustanteinen)	

12:00-13:45	Pilottisovelluksia	
	Miika Alonen Aalto	Semanttinen Kansallisbiografia: 6000 elämäkertaa semanttisena verkkona
	Matias Frosterus Aalto ja Kansalliskirjasto	Yritysdatan rikastaminen uutisaineistoilla: Case Asiakastieto
	Eetu Mäkelä Aalto	Kulttuurisampo, Matkailusampo ja Kirjasampo: 10 vuotta semanttista sammon taontaa
	Mikko Koho Helsingin yliopisto ja CSC Oy	Lintu- ja säähavainnot yhdistettynä tietona: Case Tutkimuksen tietoaaineistot
	Tomi Kauppinen Aalto	Linked Open Aalto: Yliopiston tiedot avoimesti ja yhdistetysti hyötykäyttöön
	Eero Hyvönen Aalto ja Helsingin yliopisto	Case World War I as Linked Data – yhteistyötä yli rajojen
	Jouni Tuominen Aalto	Semanttinen Finlex: lainsäädäntö ja oikeustapaukset avoimena yhdistettynä tietona
13:45-14:15	Kahvitauko	
14:15-16:00	Hyödyntäminen ja yhteiskunta	
	Ossi Nykänen Suomen W3C toimisto ja TTY	W3C:n suositukset Linked Data -julkaisuissa
	Anne Kauhanen-Simanainen Valtiovarainministeriö	Avoimen tiedon ohjelma 2013-2015
	Taru Rastas Liikenne- ja viestintäministeriö	Liikenne- ja viestintäministeriö ja avoin yhdistetty tieto
	Aki Hietanen Oikeusministeriö	Oikeusministeriö ja avoin yhdistetty tieto
	Juha Haataja Opetus- ja kulttuuriministeriö	OKM:n Avoimen tieteen tiekartta
	Osma Suominen ja Matias Frosterus Kansalliskirjasto	ONKI-ontologiapalvelun tuotteistaminen Kansalliskirjaston Finto-palveluksi
	Mikael Hindsberg YLE Oy	Yleisradio yhdistetyn avoimen tiedon hyödyntäjänä
16:00	Seminaarin päätös	