

Miten avoin yhdistetty tieto saadaan verkkoon: kansallinen Linked Data Finland -hanke-ehdotus

Eero Hyvönen
professori, tutkimusjohtaja
Semantic Computing Research Group (SeCo)
Aalto-yliopisto ja Helsingin yliopisto

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

Outline

- Background: Legacy of FinnONTO
- Question: What next?
- Answer: Linked Data Finland!

Background: FinnONTO

- **FinnONTO**

- 2003-2004
- 2004-2005
- 2005-2006
- 2006-2007

14 funding organizations
16 funding organizations
30 funding organizations
37 funding organizations

- **FinnONTO 2.0**

- 2008-2010

38 funding organizations



- 2010-2011
 - » 35 funding org.
 - 1.52 MEUR

FinnONTO spin-off-project
Semantic UBICOM-services
17 funding org., 2010-2011
0,55MEUR

FinnONTO – Some Results

- Major demonstrator systems
 - MuseumFinland: www.museosuomi.fi
 - HealthFinland: www.terveysuomi.fi
 - CultureSampo: www.kulttuurisampo.fi
 - National Ontology Library Service ONKI: www.onki.fi
- Major awards
 - Semantic Web Challenge Award, Japan, 2004
 - Prime Minister’s Innovation Acknowledgement, Finland, 2004
 - Semantic Web Challenge Award; Germany, 2008
 - Apps4Finland – Doing Good with Data, Finland, 2010
 - (World Summit Award (WSA) Mobile, Abu Dhabi, 2010)
- Academic
 - 100+ of refereed publications www.seco.tkk.f/publications
 - 4 PhD works 2010
 - Lots of MSc thesis

FinnONTO - Finnish Semantic Web Infrastructure

- **Finnish Semantic Web community**
 - tens of FinnONTO-organizations
- **National level ontology infrastructure**
 - KOKO = YSO + 15 vertical interlinked ontologies
 - International and other ontologies
- **National ontology services**
 - ONKI Ontology Service
 - Semantic portal -based services: Kulttuurisampo
- **Tools SAHA, HAKO etc.**
- **Knowhow and educated IT professionals**

How to make use of it?

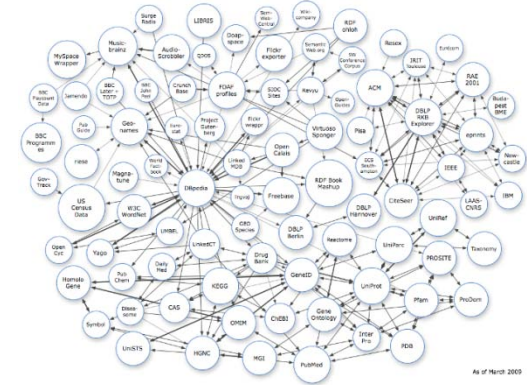
Linked Data Finland – the Vision and Plan

Avoin tieto – Open Data

- Kansalaisilla on oikeus avoimeen tietoon
- Avoimelle tiedolle syntyy hyötysovelluksia
- Tiedon arvo kasvaa sitä jakamalla
- Tuotannon tehokkuus kasvaa tietoa jakamalla
- Avoimuus parantaa yhteistyötä ja työnjakoa
- Avoimuus edistää yhteentoimivuutta

Tiedon avaaminen ei riitä!

- Open Data ei ota kantaa julkaisutapaan
 - Tietotekninen näkökulma puuttuu
- Open Data ei ota kantaa yhteistyöhön
 - Web 2.0 -näkökulma puuttuu
- Open Data avaa tiedon vain ihmiskäyttäjälle
 - Semanttisen webin näkökulma puuttuu



Linked Data on moderni ratkaisumalli tietojoukkojen julkaisemiselle ja hyödyntämiselle

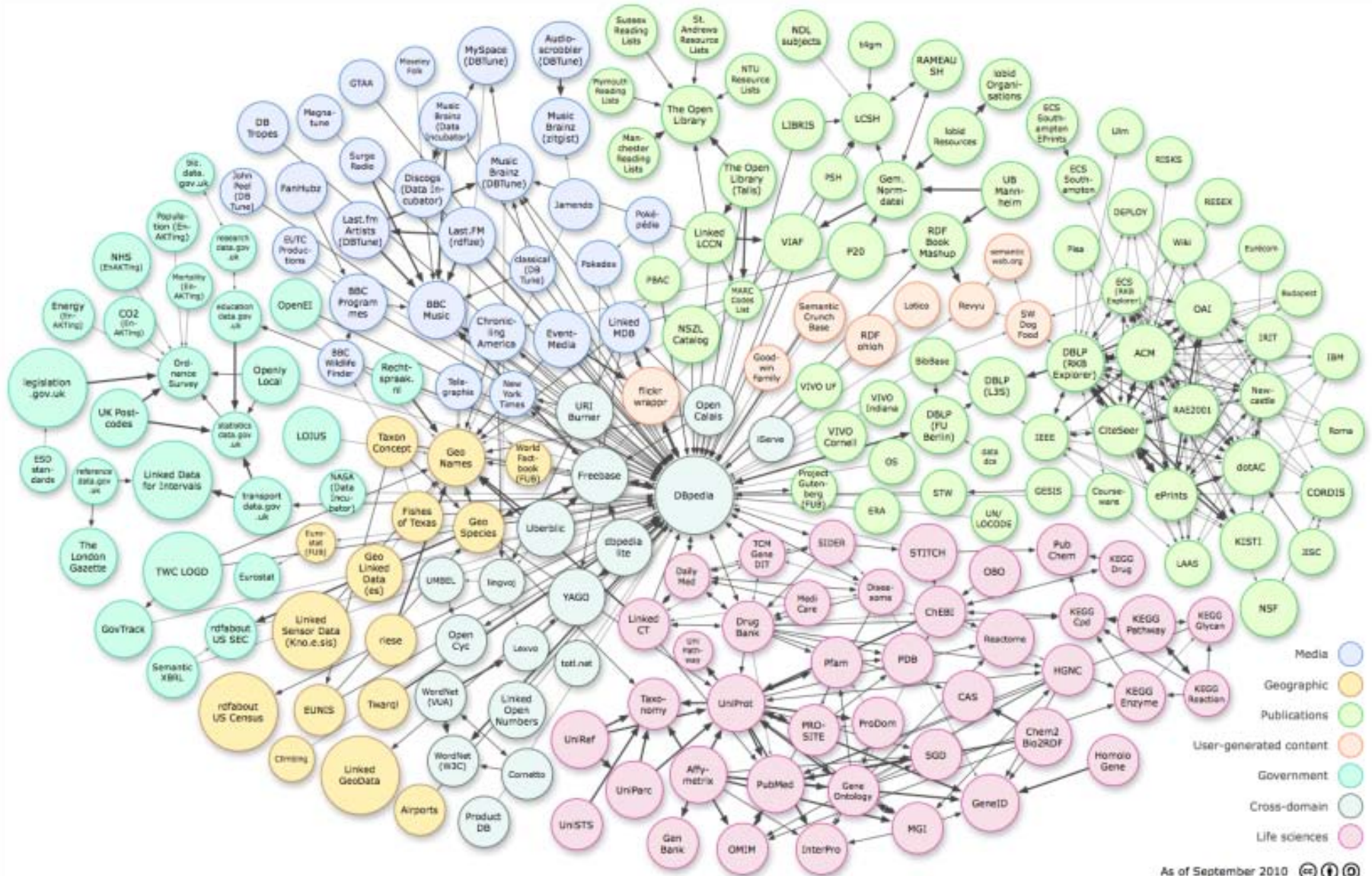
- Yksinkertainen ja käytännöllinen tapa julkaista tietojoukkoja
- Olennaista tietojoukkojen yhdistäminen ja yhteentoimivuus
- Perustuu W3C:n semanttisen webin standardeihin
- Aktiivinen tutkimus- ja kehitysyhteisö <http://linkeddata.org>
 - W3C:n tukema

Linked Data: Idea

- Nykyinen web verkko joka linkittää sivuja: **Web of Pages**
 - Voidaan hakea/selailla perinteisellä selaimelle
- Linked Data linkittää tietoa: **Web of Data**
 - Voidaan hakea/selailla myös semanttisella sovelluksella
- Uusi semanttinen web koostuu *molemmista* verkoista
 - Web of Pages (ihmiselle)
 - Web of Data (koneelle ja ihmiselle)

Linked Data + Open Data = Linked Open Data

Linked Open Data Cloud 2011



LOD Cloud Statistics

Domain	Data Sets	Triples	Percent	RDF Links	Percent
Cross-domain	20	1,999,085,950	7.42	29,105,638	7.36
Geographic	16	5,904,980,833	21.93	16,589,086	4.19
Government	25	11,613,525,437	43.12	17,658,869	4.46
Media	26	2,453,898,811	9.11	50,374,304	12.74
Libraries	67	2,237,435,732	8.31	77,951,898	19.71
Life sciences	42	2,664,119,184	9.89	200,417,873	50.67
User Content	7	57,463,756	0.21	3,402,228	0.86
	203	26,930,509,703		395,499,896	

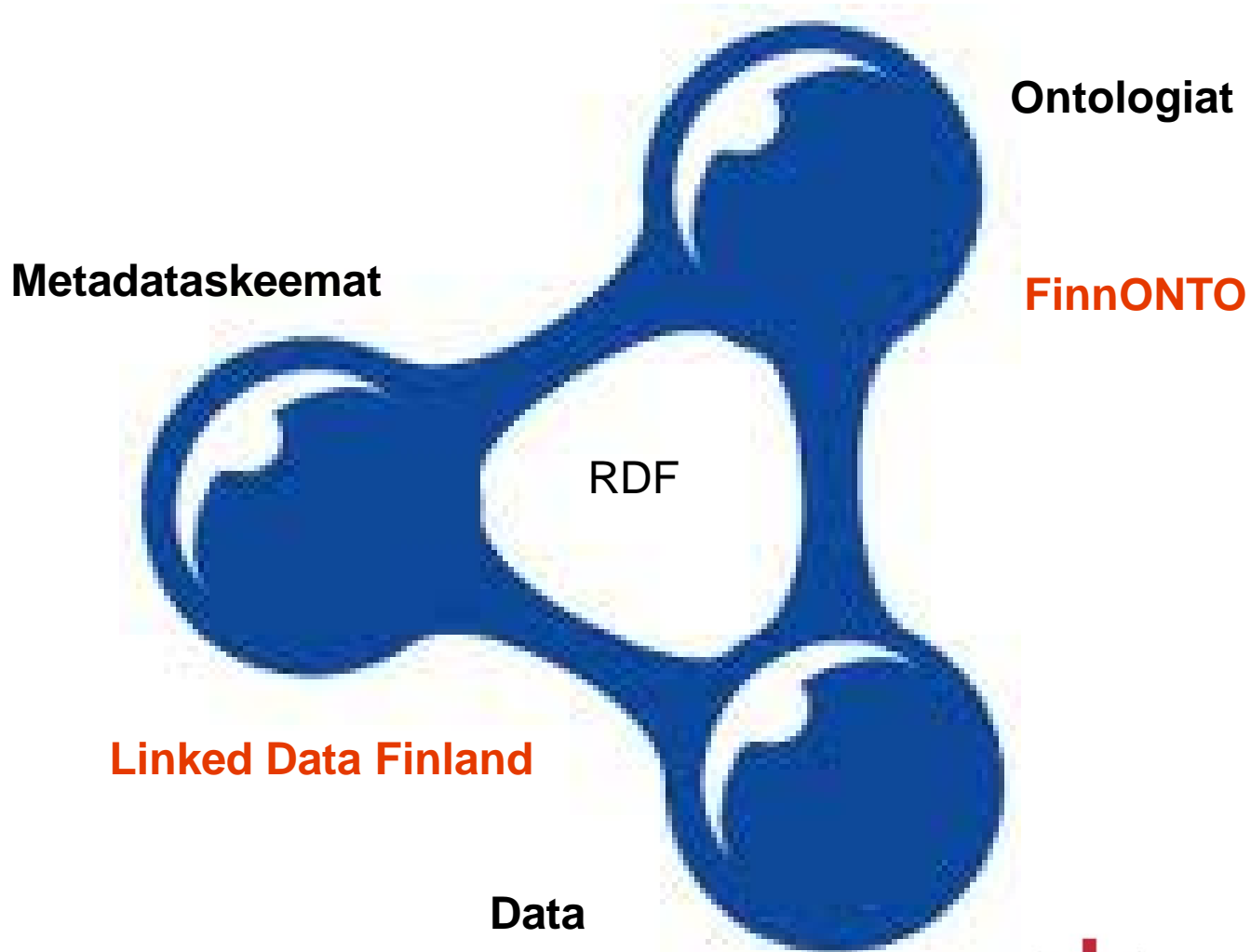
Table 3.1: Number of data sets, amount of triples, and amount of RDF links per topical domain

(Heath, Bizer, 2011)

Linked Open Data esimerkkisovellus: BBC

The screenshot shows the BBC Music website interface. At the top, there's a navigation bar with the BBC logo, a search bar, and links for 'Explore the BBC', 'Sign in', and 'Register'. Below this is a 'Music BETA' header with a 'QUICK FIND' search bar. The main content area features a large banner for 'January's Best Releases' with a text box stating: 'Starting 2010 with a flourish Mike Diver runs down the best of the new year's releases, featuring albums from Four Tet, Charlotte Gainsbourg, Laura Veirs and These New Puritans.' Below the banner are three smaller album covers: 'JANUARY'S BEST RELEASES', 'MOSCA - MARY ANNE HOBBS' GUEST MIX', and 'LATIN MUSIC USA - FRIDAY 9PM, BBC FOUR'. A 'BROWSE BY GENRE' section lists various music categories like 'Classic Pop & Rock', 'Classical', 'Country', etc. The 'Most Played Artists On The BBC' section for '8-14 FEBRUARY 2010' displays a horizontal list of artists with their play counts: Owl City (1), Tinie Tempah (2), Hot Chip (3), Paolo Nutini (4), Alicia Keys (5), Muse (6), Iyaz (7), and Mumford & Sons (8). A note at the bottom states: 'Information displayed about artists played on BBC programmes is incomplete at present. Find out more about this artist play count information.'

Web of Data: pyhä kolminaisuus



Linked Data Finland -hankeyhteisö



	Yritys / organisaatio
1	Connexor Oy
2	CSC Tieteen tietotekniikan keskus Oy
3	Eläketurvakeskus
4	FloApps Oy
5	Helsingin kaupunginkirjasto
6	Helsingin kaupunki
7	Helsingin Sanomat
8	Hgin yliopiston kirjasto
9	Kansalliskirjasto
10	Kansaneläkelaitos
11	Kotimaisten kielten tutkimuskeskus
12	Lean Technologies Oy
13	Liikenne- ja viestintäministeriö
14	Luonnontieteellinen keskusmuseo
15	Maanmittauslaitos
16	Metsähallitus
17	Metsäntutkimuslaitos
18	Museovirasto
19	Mysema Oy
20	Sanastokeskus TSK
21	Suomalaisen Kirjallisuuden Seura SKS
22	Suomen ympäristökeskus SYKE
23	Svenska litteratursällskapet SLS
24	Terveyden ja hyvinvoinnin laitos THL
25	Tieto Oyj
26	Valtiovarainministeriö / ValtIT
27	Yleisradio Oy
28	Asiantuntija jäsenet
29	W3C Suomen toimisto
30	Tampereen yliopisto, vakuutustiede
	Tutkimustahot
31	Aalto-yliopisto
32	Helsingin yliopiston tietojenkäsittelytieteen laitos



Tavoite

- Kansallisten tietoaineistojen julkaiseminen Linked (Open) Datana
 - Yhteisen open source -teknologian kehittäminen
 - FinnONTO:n perintöä hyödyntäen
 - Pilotointi käytännössä: aineistot ja sovellukset
- Laaja kansallinen FinnONTO-tyyppinen hanke
 - Kesto 2012-2013
 - Hakemus käsittelyssä
 - Jos rahoituspäätös kielteinen, yritetään uudelleen

Case-sovellusalueita

- Yhdistetty paikkatieto (mm. MML)
 - Linked Library Data (mm. yleiset ja tiet.kirjastot)
 - Bio- ja ympäristöalan yhdistetty tieto (mm. LTKM, SYKE, Metsähallitus)
 - Julkishallinnon tietorekisterit (mm. ValtIT)
 - Vakuutus- ja eläketieto (mm. Kela ja Eläketurvakeskus)
 - Media-alan avoin tieto (mm. YLE ja Helsingin Sanomat)
-
- Lisätietoja: eero.hyvonen@tkk.fi

Kiitos

Kysymyksiä?

Loppukeskustelu

Sana on vapaa