

Linked Open Data - innovaatio jaetulle avoimelle tiedolle

Eero Hyvönen

Aalto-yliopisto ja Helsingin yliopisto

Webin kehitys osoittaa, että tiedon arvo ei välttämättä vähene sitä avoimesti ja maksuttomasti jakamalla, vaan voi päinvastoin kasvaa ja mahdollistaa uudentyyppisten innovaatioiden ja liiketoiminnan syntymistä. Uusi lupaava kehityssuunta on tiedon semanttinen yhdistäminen ja rikastaminen Linked Data -periaatteiden mukaisesti. Tuloksena syntyy uudentyyppinen web-sivujen sisältöön perustuva käsitteiden ja sisältöjen semanttinen verkko, Web of Data.

Avoim tieto - Open Data

Tietoyhteiskunnan keskeisin tuote *tieto* poikkeaa perinteisistä teollisuuden tuotteista siinä, että siitä voidaan sekä valmistaa kopioita (monistaa) että jakaa näitä (sähköisesti) lähes ilmaiseksi. Teoriassa tietoyhteiskunnan kansalainen voisi siis elää tiedon osalta loputtomassa yltäkylläisyydessä ja me kaikki olla tiedon miljonäärejä. Mutta kuka kustantaisi tietotuotteen ensimmäisen kopion valmistamisen, jos kaikki muut saavat sen ilmaiseksi käyttöönsä? Ja eikö tiedon arvo vähene sitä levitettäessä sitä mukaa, kun yhä harvemmalla on tarve ostaa tieto itselleen?

Avoimen tiedon (Open Data) idea haastaa tämän perinteisen ajattelutavan. Ideana on, että tiedon arvo ei vähene vaan päinvastoin kasvaa sitä avoimesti jakamalla, jos tietoa tuotetaan ja sitä voidaan hyödyntää yhteisöllisesti. Webin kehitys ja Wikipedian, YouTuben ja Facebookin kaltaiset sovellukset ovat osoittaneet, mikä voima yhteisöllisessä avoimessa tiedontuotannossa voi piillä. Sopivasti suunnattu ilmaisuus ja avoimuus tarjoavat myös uusia mahdollisuuksia liiketoiminnallisiin innovaatioihin, kuten vaikkapa Googlen eri palvelut ovat osoittaneet.

Julkisen tiedon tuotannon kustannamme me veronmaksajat. Ensimmäisen kopion tuottamisen ja tiedon jakamisen ei siksi pitäisi olla mikään ongelma, jos tiedon avulla voidaan tuottaa kansalaisille hyödylliseksi koettuja palveluita. Julkisen tiedon maksuttomaan avaamiseen sitä tarvitseville tahoille on useita syitä.

- Kansalaisilla on oltava demokraattinen oikeus julkisin varoin tuotettuun tietoon. Tämä on jo pidempään ollut yleisperiaatteena mm. Yhdysvalloissa. Miksei Suomessakin?
- Paikallinen tietoon liittyvä eduntavoittelu saattaa estää tiedon hyödyntämisen laajemmassa yhteydessä. Esimerkiksi kartta- ja paikkatiedon osalta helposti saatavilla olevat aineistot kuten Google Maps ovat käynnistäneet todellisen sovellusten ja innovaatioiden aallon. Suomessa julkiset organisaatioita kuitenkin suorastaan edellytetään mm. maksuperustelailta sulkemaan ja tekemään taloudellista tulosta aineistoillaan. Yhden organisaation pieni etu tukahduttaa helposti innovaatioita ja kehitystä laajemmassa yhteydessä valtiotasolla.
- Yhdistämällä oma aineisto muiden toimijoiden aineistoihin, jokaisen tiedontuottajan tiedon arvo kasvaa yhteisöllisesti. Wikipedian artikkelit erillisinä eivät ole kovin arvokkaita, vaan arvo syntyy näiden muodostamasta kokonaisuudesta.
- Yhteiset avoimet pelisäännöt tiedon julkaisemisessa, yhdistämisessä ja jakamisessa edistävät tietojärjestelmien ja palveluiden yhteentoimivuutta, joka on keskeinen este mm. julkisen sektorien eri toimijoiden yhteistyössä.

- Tietoja ja ratkaisuja jakamalla voidaan tehostaa työnjako, eliminoida tarpeetonta päällekkäistä työtä ja lisätä näin työn tuottavuutta. Missä määrin Suomen 342 kuntaa ovat esimerkiksi kysyneet tekemään tietojärjestelmiään yhteistyössä? Asiakkaat ja heidän tarpeensa eri kunnissa ovat paljolti samanlaisia.
- Tietoa avoimesti julkaisemalla mahdollistetaan ennalta arvaamattomia innovaatioita. Joku voi keksiä ja toteuttaa tiedolle yllättäviä sovelluksia ja hyödyntämistapoja, jos vain tieto on helposti saatavilla. Esimerkkinä mainittakoon vaikkapa asunnonhakupalvelu Lontoossa, jossa voidaan etsiä keskustasta tietyllä ajallisella etäisyydellä olevia kohteita julkisen liikenteen avointen aikataulutietojen avulla.

Ei ole liioittelua sanoa, että avoin tieto on tietoyhteiskunnan toiminnan keskeisimpiä tukipilareita.

Yhdistetty tieto -- Linked Data

Avoimen tiedon idea ei ota kantaa siihen, miten tieto julkaistaan. Tämä voidaan tehdä teknisesti esimerkiksi Excel-taulukkona, tietokantakopiona tai toiminnallisesti palvelurajapintojen kautta. Haasteeksi jää kysymys siitä, miten tiedot oikeastaan kannattaisi jakaa, jotta ne voitaisiin vaivattomasti ottaa käyttöön eri organisaatioissa, ja miten eri tahojen tiedot voidaan yhdistää sisällöllisesti eli semanttisesti toisiinsa. Ratkaisumallin tähän tarjoaa yhdistetyn tiedon (Linked Data) idea ja kansainvälinen Linked Data -liike. Sen ideana on esittää kaikki data ja niissä käytetyt sanastot eli ontologiat yhteismitallisesti semanttisena verkkona RDF-muodossa, joka XML:n jälkeen seuraava standardi webin standardipinossa.

Idean tunnetuimpana veturina on toiminut webin kehitystä koordinoivan W3C-järjestön vetäjä ja koko "webin isä" Sir Tim Berners-Lee. Maailmanlaajuisen yhteisöllisen työn seurauksena verkossa on nyt vapaasti saatavilla miljardien tietoresurssien ja näitä yhdistävien linkkien semanttinen verkko, Web of Data. Se sisältää kymmenittäin toisiinsa sillattuja laajoja aineistoja, kuten koneellisesti louhitut Wikipedian erikieliset versiot (DBPedia), Geonames-palvelun miljoonat paikkatiedot, EU:n Eurostat-tilastot, US Census Data väestötiedot ja kirjastojen viitetietokantoja kuten Ruotsin kansalliskirjaston LIBRIS. (Ks. kuva.)

Avointen tietoa-aineistojen ja niiden yhdistämisen perustalle on alkanut nopeasti syntyä uusia innovatiivisia toimintamalleja ja käytännön sovelluksia, joita ollaan ottamassa käyttöön tietoyhteiskunnassa ja julkishallinnossa eri puolilla maailmaa. Tammikuussa Iso-Britanniassa avattiin hallituksen toimesta julkisen sektorin avoimen tiedon portaali <http://data.gov.uk>, jonka taustavaikuttajina ovat olleet pääministeri Gordon Brown ja Tim Berners-Lee. Portaalin kautta löytyy sovelluksia vaikkapa lähellä olevien koulutusmahdollisuuksien tai hoitokotien löytämiseen tai liikennetietojen seuraamiseen. Erityisen tunnettu sovellus on BBC:n kotisivujen verkkopalvelut, joissa organisaation eri tahojen mediasisältöjen yhdistämiseen käytetään Wikipedian semanttisen webin muunnosta DBPediaa. Myös Barack Obaman hallituksella on käynnissä kansallinen avoimen tiedon ohjelma, samoin Australiassa.

Suomessa aihepiiri on ollut esillä kansallisessa semanttisen webin FinnONTO-hankkeessa (2003-2012), jossa on jo kehitetty useita hajautettuun semanttiseen sisällöntuotantoon perustuvia yhteisöllisiä portaalreja, kuten MuseoSuomi.fi, TerveSuomi.fi ja Kulttuurisampo.fi, sekä kansallisen semanttisen webin sisältöinfrastruktuurin prototyyppi ONKI.fi ontologiapalveluineen. Valtiotason tietoaarkkitehtuurit -hankkeessa Valtasa (2009-) semanttisen yhteentoimivuuden asiaa ollaan viemässä eteenpäin valtiotasolla, ja tema on ollut esillä EVA:n tietoyhteiskuntaraportissa Nykyaikaa etsimässä (2009), viestintäministeriön Data-oppaassa julkisen tiedon avaamiseen (2010) sekä opetusministeriön Tutkimuksen tietoa-aineistot -selvityksessä (2010). Aiheeseen liittyvää uutta lainsäädäntöä ollaan parhaillaan valmistelemassa.

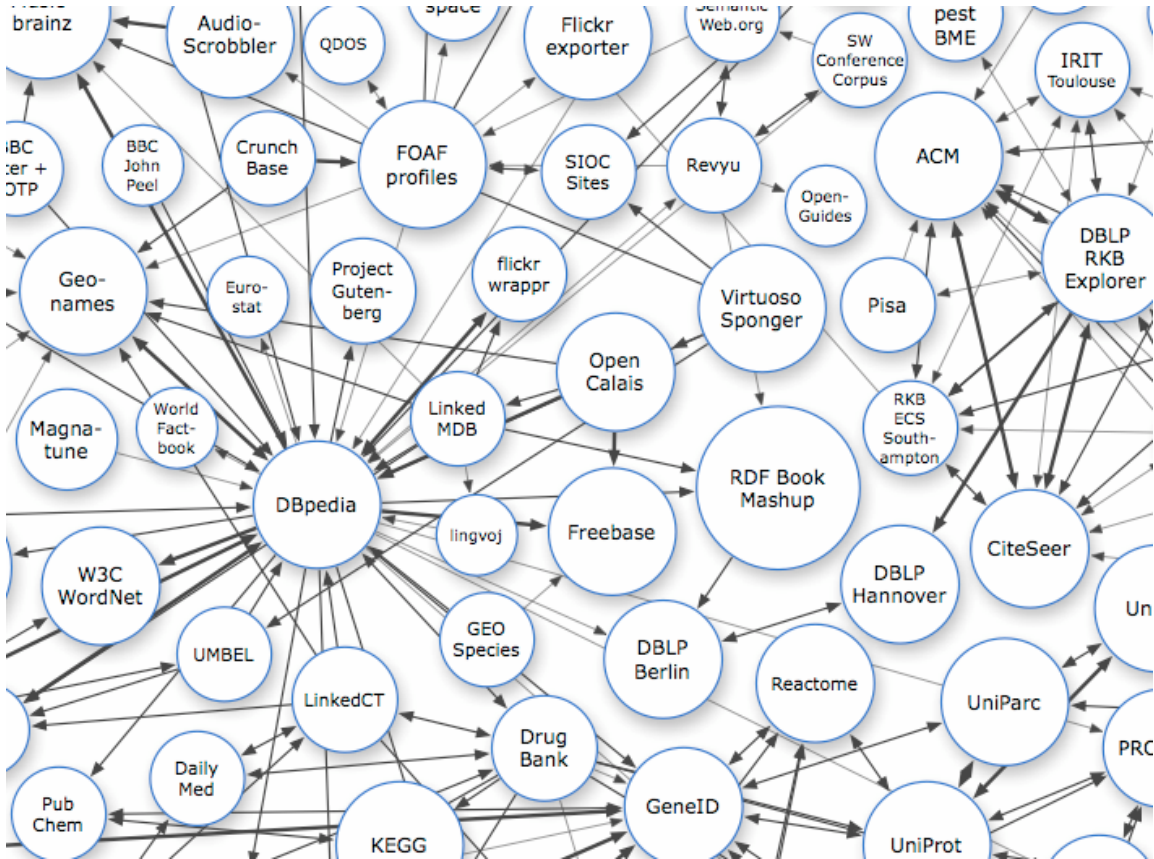
Linkitetyn avoimen tiedon tulevaisuus näyttää ainakin tässä vaiheessa valoisalta.

Lisätietoja

Linked Data -yhteisö: <http://linkeddata.org>

W3C:n Semantic Web -ohjelma: <http://www.w3.org/standards/semanticweb/>

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



Kuva. Kansainvälinen Linked Open Data -pilvi sisältää kymmenittäin toisiinsa sillattuja avoimia tietoaaineistoja, joissa on miljardeja tietoresursseja ja niiden välisiä yhteyksiä (RDF triple). Vasemmalla keskellä Wikipedian semanttinen verkko DBpedia ja siihen sillatut aineistot. (Lähde: <http://linkeddata.org/>)



Kirjoittaja on professori ja tutkimusjohtaja FinnONTO-hanketta vetävässä Aalto-yliopiston ja Helsingin yliopiston semanttisen laskennan tutkimusryhmässä SeCo (<http://www.seco.tkk.fi/>).