

RDA luo maailmanlaajuisia standardeja avoimelle tutkimusdataalle

Näkökulmat Julkaistu 29.08.2014 klo 13:24 Kirjoittaja Anni Jakobsson, CSC CC 4.0 BY

Maailmassa luodaan valtavia datamääriä joka hetki. Edellisen lauseen lukemisen aikana maailmassa tuotettiin yli 28 miljoonaa tavua, eli noin 6000 DVD-levyn verran dataa[1]. Data muuttaa myös tutkimusmaailmaa ja tieteen tekemisen tapoja.

"Olemme tieteellisen datan kanssa nyt samassa tilanteessa kuin Internet oli 1980-luvulla: tietokoneita haluttiin yhdistää keskenään, mutta ei tiedetty millä tavalla: mikä on oikea teknologia, prosessi ja politiikka tietokoneiden yhdistämiseksi. Tutkimuspuolella ja datassa on aivan sama tilanne tällä hetkellä: on olemassa valtavia tietovarantoja ympäri maailmaa, mutta millä lailla data saadaan liikkumaan rajojen yli teknisesti ja poliittisesti", kertoo johtaja **Leif Laaksonen** CSC:ltä.



”Olemme RDA:n ja tutkimusdatan kanssa paljolti samassa tilanteessa kuin IETF ja Internet 1980-luvulla.” – Leif Laaksonen, CSC

Globaalia avointa dataa

Näitä kysymyksiä ratkomaan luotiin vuonna 2012 Research Data Alliance (RDA), jonka päämääränä on edistää tieteellisen datan avoimuutta ja jaettavuutta maailmanlaajuisesti eri tieteenalojen kesken.

Hanketta rahoittavat Euroopan komissio, Yhdysvaltojen National Science Foundation (NSF) ja National Institute of Standards and Technology (NIST) sekä Australian tiedeministeriö. RDA Europe -projekti on hankkeen eurooppalainen taho, jota Laaksonen johtaa ja koordinoi CSC:llä.

"IETF (Internet Engineering Task Force) oli ratkaiseva instituutio Internetin kanssa - sovittiin käytäntöjä ja tehtiin standardeja globaalisti. Tuloksena syntyi TCP/IP-standardiin perustuvia ratkaisuja, jotka yhdistivät tietokoneet toisiinsa tarjoten monet nyt jokapäiväisesti käytössä olevat verkkopalvelut maailmanlaajuisesti. Internet syntyi tutkimusmaailman ajatuksista. Olemme RDA:n ja tutkimusdatan kanssa paljolti samassa tilanteessa kuin IETF ja Internet 1980-luvulla", Laaksonen kertoo.

Mitä kaikkea tutkimusdata oikeastaan käsittää?

"Tutkimusdata on kaikkea sitä dataa, mitä tutkija tarvitsee työssään, Laaksonen määrittelee. Kaikki asiat tutkimusmaailmassa ovat globaaleja, harva voi enää tehdä kansallisesti mitään tutkimusta", Laaksonen jatkaa.

Konkreettisia vastauksia

Tällä hetkellä RDA-hankkeessa on mukana lähes 2000 jäsentä 83 maasta. Jäsenet ovat joko asiantuntevia yksittäisiä henkilöitä tai organisaatioita. Mukana on myös suuria yrityksiä kuten Microsoft. Suomesta on mukana 37 jäsentä, mikä on Pohjoismaiden korkein osallistujamäärä.

RDA toimii eri aihealueiden ympärille koottujen työryhmien kautta. Tällä hetkellä RDA:ssa on lähes 50 työryhmää, joiden aiheet ulottuvat laidasta laitaan: RDA:ssa on esimerkiksi metadatastandardeja pohtiva ryhmä ja maanviljelyn tuottamaa dataa käsittelevä ryhmä. Ryhmät keskustelevat konkreettisista ongelmista RDA:n verkkosivuilla, ja kokoontuvat puolivuositain pidettävissä kokouksissa. Työryhmien työskentelyaika on rajattu maksimissaan puoleentoista vuoteen.

Ensimmäisiä konkreettisia tuloksia työryhmiltä odotetaan syyskuussa Amsterdammassa pidettävästä neljännessä RDA Plenary -kokouksesta. Kuka tahansa voi liittyä mukaan RDA:n toimintaan verkkosivujen kautta.

"Internetin syntymisen jälkeen on luotu valtavasti uudenlaisia työpaikkoja. Jos datan suhteen voidaan päästä edes lähelle sitä, voidaan luoda aivan uudenlaista taloutta ja työpaikkoja, jotka perustuvat datan avoimeen liikkuvuuteen", Laaksonen sanoo.

RDA tekee yhteistyötä myös muiden standardointia tekevien organisaatioiden ja hankkeiden kuten W3C:n (World Wide Web Consortium) ja CODATAN (Committee on Data for Science and Technology) kanssa.

Dataa eri tieteenaloilta

Avoimen datan avulla on mahdollista tehdä aivan uudenlaista tiedettä. Tutkija voi päästä käsiksi toisella puolella maailmaa olevaan tietoaaineistoon helposti, ja esimerkiksi yhdistellä tätä aivan toisen tieteenalan data-aineistoihin. Tavoitteena on aidosti avoin datan liikkuvuus maiden ja tieteenalojen kesken.

"Esimerkiksi mitä mekanismeja voidaan luoda, että eurooppalaisen fyysikon data, aasialaisen ympäristötieteilijän aineisto ja yhdysvaltalaisen kielitieteilijän data voisivat keskustella keskenään, Laaksonen kuvailee. Tulevaisuudessa tietoaaineistot tulevat toimimaan meriittinä tutkijalle samaan tapaan kuin tieteelliset julkaisut nykyään", Laaksonen jatkaa.

Uusi dataan pohjautuva tiede ja talous on mahdollista vain yhdistämällä maailmanlaajuinen tieteellinen datainfrastruktuuri, joka on RDA:n päämääränä.

[1] European Commission. “Commission urges governments to embrace potential of big data.” IP/14/769. 02/07/2014.

Linkkejä:

<https://europe.rd-alliance.org/>

<https://www.rd-alliance.org/groups/working-groups>

<https://www.rd-alliance.org/groups/interest-groups>

<https://www.rd-alliance.org/user/register>