

# Työmaana tutkimusdatan avoimuus

Sami Borg

*Pidin marraskuussa 2013 esitelmän otsikon aiheesta tieteellisen kirjastoseuran seminaarissa ”Tutkijapalvelut näkyviksi – tutkimusaineistot ja kirjastojen rooli”. Tämä kirjoitus ei ole esityksen toisinto, vaan tavoitteeni on pikemminkin luoda tilannekuvaa tutkimusdatan avaamisesta Suomessa ja alan haasteista.*

**P**ääteesini pelkistyy parin melko ilmeisen syyseuraus -suhteen skenaarioksi. Tutkimusdatan avaamisella on jo nyt niin vahva kansainvälinen ja kansallinen poliittinen tuki, että tämä pakottaa tutkimusrahoittajat ja -organisaatiot luomaan tai tarkistamaan datan avaamista tukevia menettelytapaohteita. Tällaisten datapolitiikkojen tehokas toteuttaminen edellyttää lähitulevaisuudessa riittävän laadukkaita tietopalveluja ja teknisiä ratkaisuja, joiden kehittämisen ja koordinoinnin tarve on suuri. Datan avaamista ja tutkimusdatan jatkokäyttöä tukevien tietopalvelujen järjestäminen tulee olemaan keskeinen haaste kaikille alan potentiaalisille toimijoille, myös tieteellisille kirjastoille. Alan koulutustarve on ilmeinen.

## Poliittinen tuki

Datan, ja sen osana tutkimusdatan, avaamisen tarpeellisuudesta ei internetaikana enää oikeastaan kannata kiistellä. Keinoista ja avaamisen yksityiskohdista pitää keskustella, mutta itse päämäärällä on nykyisin jo niin laaja ja vahva poliittinen tuki, että avata pitää (vrt. Wilhelmsson ja Keinonen 2013). Datan avoin saatavuus vahvistaa yleisesti yhteiskuntien päätöksenteon tietopohjaa. Lisäksi tuottavuutta halutaan lisätä tehostamalla datan keruuseen liittyviä julkisia investointeja ja luomalla mahdollisuuksia uudelle, dataperustaiselle liiketoiminnalle. Nämä yleiset taustamotiivit ovat löytyneet esimerkiksi sähköiseen muotoon tallennetun julkisrahoitteisen datan avaamisen suosituksista jo pitkään (esim. OECD 2007, ks. Borg ja Kuula 2007) ja nyt da-

tan avaaminen sisältyy myös maamme hallitusohjelmaan.

Suorimmin datan avaamispyrkimys näkyy Kataisen hallituksen viestintäpoliittisissa tavoitteissa. Hallitusohjelma (VNK 2011) toteaa: ”Julkisin varoin tuotettuja tietovarantoja avataan kansalaisten ja yritysten käyttöön. Tavoitteena on julkisen sektorin hallinnoimien digitaalisten tietoaaineistojen saattaminen helposti uudelleenkäytettävässä muodossa tietoverkkojen kautta kansalaisten, yritysten ja yhteisöjen, viranomaisten, tutkimuksen ja koulutuksen hyödynnettäväksi.”

Lisäksi hallitusohjelma toteaa viestintäpoliittisten linjausten yhteydessä, että hallituksen tulee varmistaa ”...julkisten tietojärjestelmien yhteentoimivuus käyttämällä avoimia rajapintoja ja standardeja. Avoimeen lähdekoodiin perustuvien ratkaisujen käyttöönottoa edistetään julkisen hallinnon kokonaisarkkitehtuurin puitteissa ja kustannushyötyanalyysin pohjalta.”

Erityisesti tutkimusdatan avaamisen kannalta on niin ikään oleellista, että tiedettä koskeva lyhyt hallitusohjelmakirjaus mainitsee ensimmäisenä asiana kansallisen tutkimusinfrastruktuuripolitiikan: ”Edistetään kansallista infrastruktuuripolitiikkaa”. Tutkimusinfrastruktuurien kehittämislä on tiedeyhteisössä keskeinen merkitys tutkimusdatan avoimelle saatavuudelle. Euroopan Unionissa tämän suuntauksen vahvistaminen on ollut käynnissä jo kymmenisen vuotta tutkimusinfrastruktuurien strategiafoorumien eli ESFRI:n puitteissa. Ilman tätä prosessia Suomessakaan olisi tuskin edetty asiassa niin rivakasti, kuten nyt on teh-

ty. 2000-luvun loppupuolella laaditun tutkimusinfrastruktuurien kansallisen tiekartan päivitys julkaistaan aivan lähiaikoina.

EU näyttää tutkimusjulkaisujen ja -datan avaamisessa myös esimerkkiä. Se edellyttää uudessa tutkimuksen puiteohjelmassaan Horisontti 2020 perusteellista avoimuutta sekä rahoitettavien hankkeiden tuottamilta julkaisuilta, että niihin liittyvältä tutkimusdataalta. Asiaa edistetään muun muassa pilottihankkein sekä olemassa olevien avaamisinfrastruktuurien avulla.

## Dataa on jo avattu

Suomi on onneksemme ollut hereillä tutkimusdatan avaamisessa jo kohtuullisen varhain. Useilla tieteenaloilla on jo kauan osallistuttu hankkeisiin, joiden tutkimusdata on ainakin alan sisäisesti avoimessa käytössä. Suomen Akatemia on muutamien vuosien ajan edellyttänyt kaikilla tieteenaloilla aineistonhallintasuunnitelman laatimista rahoitushakemuksen yhteyteen.

Keskitettyjä tuki-infrastruktuurejakin on perustettu. Yhteiskuntatieteellinen tietoarkisto ([www.fsd.uta.fi](http://www.fsd.uta.fi)) perustettiin Suomen Akatemian aloitteesta 1990-luvun loppupuolella tukemaan yhteiskuntatieteiden ja niiden lähitieteiden piirissä kertyvien sähköisten tutkimusaineistojen pitkäaikaissäilytystä ja avaamista jatkokäyttöön. Yksikkö on arkistoinut noin 1400 tutkimusaineistoa, pääosin eri alojen kyselytutkimusten havaintoaineistoja. Tietovarannossa on pari sataa kvalitatiivista tutkimusaineistoa, joista useimmat sisältävät haastatteluista purettua ja dokumentoitua tekstiä.

Internetin kehittymisen myötä useat keskeiset yhteiskunnalliset tiedontuottajat ryhtyivät 2000-luvulta alkaen kehittämään verkossa toimivia avoimia tietokantoja, jotka mahdollistavat pääsyn yleiseen käyttöön soveltuviin tietoihin. Tutkijoiden kannalta ongelma on kuitenkin siinä, että avattujen tietojen taustalla olevat perusaineistot eivät useinkaan ole avoimesti tutkimuksen saatavilla.

Opetus- ja kulttuuriministeriön hallinnonalalla

2000-luku toi mukanaan lukuisia Open Access –toimenpiteitä. Tutkimusdatan avaamisen kannalta olennaisin on CSC:n koordinoima Tutkimuksen tietoaaineistot –hanke, joka alkoi ensin selvitushankkeena vuosina 2009–2010 (Tieto käyttöön 2011) ja jatkui sitten laajana yhteistyöhankkeena vuosina 2011–2013. Vuodesta 2014 alkaen TTA-hankkeen toiminnot ja palvelut siirtyvät osaksi CSC:n palveluja (ks. [www.tdata.fi](http://www.tdata.fi)).

TTA:n myötä käyttöön tulevat palvelut ovat jo nyky muodossaan laaja-alaisia kattaen kehittyviä ratkaisuja tutkimusdatan säilytyspalveluille (IDA), metat datapalveluille (KATA), hyödyntämispalveluille (AVAA) sekä aineistojen käyttövaltuutukselle (REMS). Lisäksi TTA tulee tuottamaan Kansallinen digitaalinen kirjasto -hankkeen kanssa tutkimusdatan pitkäaikaissäilytyspalvelun, joka otettaneen käyttöön vuonna 2015.

TTA-palvelut ovat varteenotettava vaihtoehto sellaisten tieteenalojen tutkimusdatalle, joilla ei ole toistaiseksi käytettävään tietoaarkiston kaltaisia ala- tai aineistotyyppikohtaisia palveluja. Lisäksi ne tarjoavat käyttökelpoista teknistä infrastruktuuria olemassa oleville alakohdaisille toimijoille. Tietoaarkiston palvelut sopivat muun muassa kysely- ja haastatteluaineistoihin, jotka ovat tyypillisiä ihmistieteissä ja eri yhteiskuntasektoreilla.

## Tutkimusdatan avaaminen vaatii osaamista

Tutkimusdatan avaamista kannattaa tarkastella tutkimusaineistojen elinkaaren näkökulmasta. Datan keränneen tutkijan tai tutkijaryhmän perustyon jälkeen pitkän elinkaaren tukeen sisältyy työvaiheita, joita tutkimushankkeet ja niiden tutkijat eivät yleensä ole tottuneet itse suunnittelemaan ja toteuttamaan.

Jotta ensikäyttäjät voivat saada avoimessa jatkokäytössä olevista tutkimusaineistoistaan heille kuuluvan viittausmeriitin, tutkimusdata pitää julkaista. Julkaisemista varten data pitää tarkistaa ja dokumentoida. Lisäksi on tarvittaessa sovittava, millä ehdoilla aineistoa voi käyttää uu-

siin käyttötarkoituksiin.

Dokumentointi ei käy käden käänteessä. Pitää kuvailla yksityiskohtaisesti tutkimusdatan tietosisällöt, datan kerännyt tutkimushanke ja tiedonkeruuseen liittyvät muut olennaiset seikat. Dokumentointi vaatii standardeja sekä taitoja ja resursseja niiden soveltamiseen. Lisäksi tarvitaan tekninen infrastruktuuri tukemaan tiedon löytymistä ja datan jakamista verkossa.


Tehtävälistan pituus puhuu sen puolesta, että tutkijoiden omaan toimeliaisuuteen datan avaamisessa ei kannata luottaa liikaa. Sanon tämän 15 vuoden data-arkistokokemuksella. Kuka tai mikä tahoo tutkimusdatatietopalveluja tarjoaakaan, kaikissa tapauksissa tarvitaan osaavaa datapalveluammattilaista, joka voi auttaa aineiston tuottajia datan dokumentoinnissa, arkistoinnissa ja julkaisemisessa, sekä tukea tutkijoita, opettajia ja opiskelijoita tutkimusdatan jatkokäytössä. Työ on vaativaa ja vie aikaa, enkä siksi pidä lainkaan pahana sitä, että tutkijat haluavat ja heidän pitää saada keskittyä tutkimukseen.

## Kirjastojen rooli?

Dataan avaamiseen ja jatkokäyttöön liittyviä tukipalveluja kannattaa hajauttaa silloin kun se on perusteltua. Toisaalta niitä kannattaa myös perustelluista syistä keskittää. Kirjastoille soveltuvia tehtäviä on jo nykyisellään runsaasti. Kirjastoasiantuntija voi opastaa opiskelijoita ja tutkijoita esimerkiksi tutkimusdatan löytämiseen ja käyttömahdollisuuksiin sekä keskitettyjen tutkimusdatapalvelujen käyttöön.

Tutkimusdatan elinkaarta ja jatkokäyttöä selvästi laaja-alaisemmin tukeva profiili on jo eri asia. Olisi varmaankin hyvä kerryttää kokemuksia erilaisista toimintamalleista. Suomesta löytyy jo nyt ala- tai aineistotyyppikohtaisesti manda-toituja valtakunnallisia ratkaisuja (tietoarkisto), teknistä infrastruktuuria laaja-alaisesti tarjoavia kansallisia ratkaisuja (TTA) sekä orastavia yliopistokohtaisia tukiratkaisuja (esimerkiksi Jyväskylän yliopistossa suunnitelma Dataverse-alustan hyödyntämisestä).

Toistaiseksi suomalaisilla yliopistokirjastoilla ei tietääkseni ole ollut tutkimusdatapalvelujen tuottamisessa niin merkittävää asemaa, että siihen liittyisi laajaa vastuuta myös itse tutkimusdatan säilyttämisestä, arkistoinnista ja jatkokäytön tuesta. Mitenkään mahdoton ajatus ei ole, sillä esimerkiksi Pohjois-Amerikassa yliopistokohtaiset, kirjastotietopalveluihin linkittyvät datapalvelut ovat yleisiä. Sielläkään ne eivät tosin kovin usein kata datan elinkaaren hallintaa laajasti, vaan ne toimivat lähinnä paikallisina tukipisteinä ja ”käyttöliittyminä” keskitettyihin data-arkistoihin ja data-keskuksiin.

Mikä tieteellisten kirjastojen rooli tulee tulevaisuudessa asiassa olemaan, pidän varsin todennäköisenä, että myös kirjastot tarvitsevat jatkossa yhä useampia datapalveluammattilaisia, joiden tulee kouluttautua tutkimusdatan hallintapalvelujen vaatimiin erityistaitoihin. 

## Lähteet

Borg, Sami ja Kuula, Arja (2007): Julkisrahoitteen tutkimusdatan avoin saatavuus ja elinkaari. Valmisteluraportti OECD:n datasuosituksen toimeenpanomahdollisuuksista Suomessa. Yhteiskuntatieteellisen tietoarkiston julkaisuja 6. Tampere: Yhteiskuntatieteellinen tietoarkisto.

OECD (2007). OECD Principles and Guidelines for Access to Research Data from Public Funding. <http://www.oecd.org/dataoecd/9/61/38500813.pdf>

Tieto käyttöön (2011): Tieto käyttöön. Tiekartta tutkimuksen sähköisten tietoaineistojen hyödyntämiseksi. Opetus- ja kulttuuriministeriön julkaisuja 2011:4. Helsinki: OKM.

Wilhelmsson, Thomas ja Keinonen, Juhani (2013): Tiedon avoimuus mullistaa myös yliopistomaailman. Vieraskynäkirjoitus Helsingin Sanomissa 23.11.2013.

VNK (2011): Pääministeri Jyrki Kataisen hallituksen ohjelma. Helsinki: Valtioneuvoston kanslia.

## Tietoa kirjoittajasta

*Sami Borg*  
johtaja, Yhteiskuntatieteellinen tietoarkisto FSD  
Email. [Sami.borg@uta.fi](mailto:Sami.borg@uta.fi)