

# DATAVIITTAUKSET - TUTKIMUSEKOSYSTEEMIN KULMAKIVI

Tietoarkisto tiedusteli asiakaskyselysään, kuinka moni viittaa julkaisussaan omaan tutkimusdataansa tai hyödyn-tämäänsä muiden tutkijoiden tuotamaan dataan. Noin 60 prosenttia vastaajista kertoi viittaavansa muiden dataan ja vain alle puolet omaan dataansa. Pidän tulosta hieman yllättävänä. Onhan viittaaminen tieteellisen kulttuurin ytimessä.

Kyselyn otos oli suppea ja tuskin kovin edustava, mutta tulos tukee näkemystä siitä, että dataan viittaaminen ei – edelleenkaan – ole samalla tavalla tieteentekemisen selkärangassa kuin viittaaminen julkaisuun. Näin ei varmastikaan ole siksi, ettei datan tekijälle haluttaisi antaa ansaittua meriittiä. Pikemminkin on niin, ettei dataa tai sen tulkinnan kannalta keskeisiä oheistiedostoja, kuten ohjelmakoodia, aina nähdä vastaavana tutkimustuotoksena kuin vertaisarvioitua artikkelia.

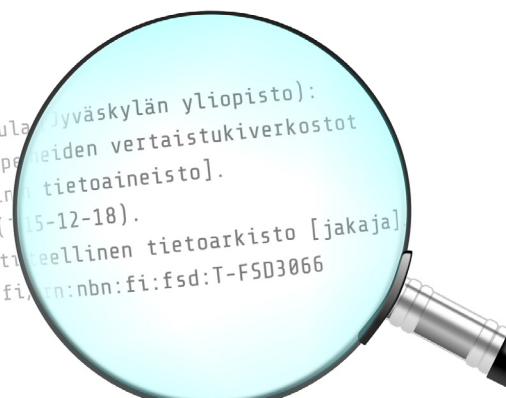
Pyrkimystä muutokseen on. Vastuullisen arvioinnin linjaukset korostavat, että data tulisi nähdä yhtä arvokkaana tuotoksena kuin artikkeli. Tieteelliset lehdet edellyttävät yhä useammin, että datan saatavuudesta on selkeästi kerrottu. Silti datan tuottaja tai jakelija osoittautuu usein vaikeammin viitattavaksi tahoksi kuin artikkelin kirjoittaja tai kirjan kustantaja. Data on abstraktimpaa, siinä ei ole sivunumeroita ja se voi päivittyä. Viit-tausmallien tuki on ollut rajoittunutta dataan viittaamisessa. Lähdeluetteloissa data sijoitetaan usein erilleen muista lähteistä tai itse viittaukset eivät osoita

suoraan dataan, vaan siitä julkaistuun kirjallisuuteen.

Kohta jo vuosikymmenen olemme kuulleet, että tutkimusdatan tulisi olla FAIR. Itseasiassa – koko tieteentekemisen prosessin pitäisi varmistaa, että FAIR on mahdollista. Tämä koskee datan ohella menetelmiä, koodia ja metatietoa. Niinpä FAIR ei enää tarkoita ainoastaan datan löydettävyyttä, saatavuutta, yhteentoimivuutta ja uudelleenkäytettävyyttä, vaan sitä, että tutkimus tapahtuu ympäristössä, jossa FAIR vahvistuu ja samalla arkipäiväistyy.

Dataan viittaaminen on tutkimusekosysteemin yksi kulmakivi tai vähintään vahva peruslohkare, jonka päälle rakentaa. FAIRin ytimessä on aina koneluettavuus ja se, että suuresta joukosta dataa ja sitä dokumentoitavaa metatietoa voidaan ohjelmallisesti jalostaa merkityksellisiä yhteyksiä ja uutta tietoa. Onnistuminen edellyttää, että julkaisun, sen taustalla olevan datan ja tuotosten laatineiden tahojen välille on tarjolla yhteys, joka on yksiselitteinen ja pysyvä.

Nähdäkseni tieteellisten kirjastojen ohjeissa ja tiedonhankinnan ja -hallin-



Nystedt, Ursula (Jyväskylän yliopisto):  
Erityislapsiperheiden vertaistukiverkoston  
2014 [sähköinen tietoaineisto].  
Versio 1.0 (2015-12-18).  
Yhteiskuntatieteellinen tietoarkisto [jakaja]  
<http://urn.fi/urn:nbn:fi:fsd:T-F5D3066>

nan koulutuksissa tuodaan esille, että systemaattinen viittaaminen mihin tahansa lähteeseen on mahdollista. Kirjastoille pysyvä tunniste, useimmiten DOI, on tuttu. Yksinkertaistaen tarvitaan siis vain, että artikkelissa viitataan kone- ja ihmisluettavasti hyödynnettyihin datoihin pysyvää tunnistetta käyttäen. Hyvä dataviite ei kuitenkaan ole aina istunut perinteisten viittausmallien muotoihin. Zoteroon kattava tuki datasetille tyyppinä tuli vasta keväällä 2023.

Pätevässä viitteessä dataan – tai sen tulkitsemiseen vaadittavaan materiaaliin – tulee olla tekijän ja datan nimi, julkaisuvuosi, versio, datan julkaisijan nimi ja pysyvä tunniste. Oletammehan, että datalla on tunniste. Näitä perustietoja on hyvä täydentää merkinnällä, joka kertoo viittauksen kohteen olevan [tutkimusaineisto] tai [lähdekoodi].

Yleisten viittausjärjestelmien uudemmissa versioissa tämä onnistuu, mutta edellyttää viittaajalta soveltamista ja ymmärrystä siitä, mitä datan saatavuus viitteeltä vaatii. Esimerkiksi viitteessä tulee käyttää pysyvää tunnistetta sen sijaan, että kopioidaan datan URL-osoite selaimen osoiteriviltä. Datalähteen avoimuutta kuvattaessa on hyvä tarkentaa, onko viitattu data ”saatavissa” (available) vai ”haettu” (retrieved).

Tietoarkisto antaa aineistoilleen viittausmallin, mutta silti usein kysytään, kuinka mallia soveltaa esimerkiksi oman oppilaitoksen vaatimusten mukaisesti. Neuvomme mielellämme, ja minulla on valmiina esimerkkejä eri viittausjärjestelmillä ilmaistuna. Samalla näen, että vastauksen tulisi pääosin selvittää opintojen myötä. Kirjastot asiantuntijuudellaan, laadukkaalla koulutuksellaan ja edunvalvojina viittauskäytäntöjen ja viitteidenhallintajärjestelmien kehittämisessä vaikuttavat siihen, että tilanne paranee vuosikurssi vuosikurssilta.

Ammoisina aikoina – noin toistakymmentä vuotta sitten – tutkimusdataa kuvattiin kirjastotietokantoihin. Menettely vaati luovuutta, jota FINMARC- tai MARC 21 -formaatti ei mielellään mahdollistanut. Ajatus oli edellä aikaansa: data ja artikkeli löytyisivät samasta käyttöliittymästä. Nyt tämä on mahdollista esimerkiksi Finnaassa, koska sekä tieteellisten kirjastojen että Tietoarkiston metatiedot ovat osa Finnaa. Yhteyttä voi tiivistää siten, että saamme yhteiseen metatietovarantoon vietyä yhä enemmän uniikkeja tunnisteita datoilta, tekijöille ja organisaatioille. Näiden tunnisteiden avulla voidaan luoda verkko, jossa data, artikkeli ja tutkija kohtaavat ja vaikuttavat nykyistä enemmän. ❖

#### TUOMAS J. ALATERÄ

Kirjoittaja on Yhteiskuntatieteellisen tietoarkiston digitaalisen pitkäaikaissäilytyksen, viestinnän ja verkkopalveluiden erityisasiantuntija. Hän on osallistunut useisiin kotimaisen AVOTT-koordinaation tutkimusdatan avoimuuden työryhmiin ja kansainvälisesti toimintunut tutkimusdata-ammattilaisten IASSIST-järjestön puheenjohtajana. Tällä hetkellä hän on CESSDAn dataviittausuudistuslaativan työryhmän jäsen.