

## Dataviittaamisen tiekartta tutkijalle

Heidi Laine

*Helsingin yliopisto*

heidi.laine@helsinki.fi

<https://orcid.org/0000-0002-6658-0664>

Susanna Nykyri

*Tampereen teknillisen yliopiston kirjasto*

susanna.nykyri@tut.fi

<https://orcid.org/0000-0002-5018-5176>

Asiasanat: dataviittaaminen; data; datanhallinta; meritoituminen; bibliometriikka; tieteellinen julkaisu-  
toiminta; Tracing Data - Data Citation Roadmap for Finland

Opetus- ja kulttuuriministeriön Avoin tiede ja tutkimus (ATT) -hanke tilasi ke-  
vällä 2017 Data-asiain kansalliskomitealta selvityksen tutkimusdataan viittaamisesta. Selvityksen tuloksena julkaistiin huhtikuussa 2018 dataviittaustiekartta suomalaiselle tiedeyhteisölle (Finnish Committee for Research Data, 2018). Tie-  
kartan tavoitteena on yhdenmukaistaa tutkimusdataan viittaamisen käytäntöjä, parantaa tutkimusdatan viitattavuutta sekä lisätä viittaussmääriä.

Tampereen teknillisen yliopiston kirjasto, Data-asiain kansalliskomitea ja Suomen yliopistojen rehtorineuvosto UNIFI ry järjestivät 18.5. Tampereella Tutkimusdataan viittaaminen: tiekartasta käytäntöön -seminaarin dataviittaustiekartan toimeenpanosta.<sup>1</sup>

Tässä kirjoituksessa käymme läpi tutkimusdataan viittaamisen periaatteita ja käytäntöjä tiekartan ja seminaarin pohjalta. Tavoitteenamme on tehdä aihepiiriä tutuksi ja ymmärrettäväksi erityisesti tutkijoille.

---

1 Tilaisuuden materiaalit löytyvät osoitteesta <https://www.fcrd.fi/events/tutkimusdataan-viittaminen-tieka/>

## Dataan viittaaminen tukee avointa tiedettä

Dataan viittaamisen yhteys datan uudelleen käyttöön on tehnyt siitä lupaavan välineen tieteen avoimuuden edistämiseksi. Avoin tiede on liike, joka edistää tutkimustuotteiden nykyistä laajempaa saatavuutta ja käytettävyyttä. Tutkimusdata on yhdessä tutkimusjulkaisujen kanssa merkittävin tutkimustuote.

Tutkimusdataan viittaamisen tapojen vakiintumisen ja yleistymisen uskotaan yleisesti lisäävän vastuullisen aineistohallinnan houkuttelevuutta tutkijoille. Vastuulliseen hallintaan kuuluu datan kuvaileminen tavalla, joka tekee sen ymmärrettäväksi muillekin kuin sen tuottamiseen osallistuneille tutkijoille, sekä datan säilyvyyden turvaaminen luotettavilla tallennusratkaisuilla.

Kirjallisuusviittauksiin perustuvat bibliometriset indikaattorit ovat nostaneet vertaisarvioitua tutkimusjulkaisua tieteellisen meritoitumisen ensisijaiseksi välineeksi. Datasta huolehtimiseen ei tällä hetkellä ole vastaavia kannustimia. Lisääntyessään dataviittaukset mahdollistavat datametriikoiden kehittämisen, jotka puolestaan helpottavat tulevaisuudessa tutkimusdatatuotteiden huomioimista tutkijoiden ja tutkimuksen arvioinnissa.

## Tutkijat avainasemassa

Dataviittaustierkartassa tunnistetaan dataviittauskäytäntöjen kannalta viisi tärkeää sidosryhmää: tutkijat (researchers), datan säilyttäjät (data repositories), tutkimusinstituutit (research institutions), julkaisijat (publishers) sekä päättäjät (policy makers). Jokaisella sidosryhmällä on tärkeä roolinsa tutkimusdataan viittaamisen ekosysteemissä ja siksi tiekartan suositusten yhteydessä nimetään aina sidosryhmä, jonka vastuulla kyseisen suosituksen toteuttaminen on.

Tutkijoiden rooli ekosysteemissä on ratkaiseva. Ilman tutkijoita järjestelmä on moottori ilman polttoainetta. Tutkijoiden vastuu dataviittaus ekosysteemin toiminnassa voidaan jakaa kahteen pääluokkaan: oman datan viitattavuudesta huolehtimiseen sekä muiden dataan viittaamiseen vastuullisesti ja asianmukaisesti.

Datan säilyttäjät, tutkimusinstituutit, päättäjät ja julkaisijat synnyttävät linjauksillaan ja palveluillaan dataviittausinfrastruktuurin. Infrastruktuuri rakentuu muun muassa päättäjien tarjoamasta rahoituksesta, tutkimusinstituutioiden ja julkaisijoiden datapolitiikoista, joissa määritetään datasta huolehtimisen tavat, sekä datan säilyttäjien tarjoamista pysyvästä tunnisteista ja datan kuvailusivusta (eräänlainen datakokonaisuuden profiilisivu).

## Dataviitteen merkitys ja sisältö

Dataan viittaamiseen pätevät samat tutkimuseettiset periaatteet kuin kirjallisuuteen viittaamiseen.

Tunnustuksen antaminen toisten tekemälle tutkimustyölle on yksi tutkijan ammattietiikan tärkeimpiä periaatteita. Muiden työn hyödyntäminen ilman viittaamista on hyvän tieteellisen käytännön rikkomista, vakavimmillaan tutkimusvilppiä. Tutkimuseettisen neuvottelukunnan antamassa hyvää tieteellistä käytäntöä koskevassa ohjeistuksessa todetaan muiden työn tunnustamisesta seuraavasti:

Tutkijat ottavat muiden tutkijoiden työn ja saavutukset asianmukaisella tavalla huomioon niin, että he kunnioittavat muiden tutkijoiden tekemää työtä ja viittaavat heidän julkaisuihinsa asianmukaisella tavalla ja antavat heidän saavutuksilleen niille kuuluvan arvon ja merkityksen omassa tutkimuksessaan ja sen tuloksia julkaistessaan. (TENK, 2013)

Dataviittaustiekartta sisältää suosituksen siitä, mitä tietoja dataviitteeseen tulee sisällyttää. Tärkein tietoelementti on pysyvä tunniste, joko DOI- tai URN-muotoisena. Pysyvä tunniste johtaa oikean data-aineiston äärelle, vaikka sen tallennussijainti muuttuisi vuosien varrella. Pysyvä tunniste myös tekee viittauksesta koneluettavan eli tietokoneelle ymmärrettävän. Viittausten koneluettavuus on edellytys datametriikoiden vaatimalle automatisoidulle tiedonkeruulle.

Pysyvä tunniste voisi periaatteessa yksin riittää viittaukseksi, sillä se toimii linkkinä data-aineiston kuvailusivulle, johon on mahdollista sijoittaa lähes ääretön määrä metatietoja. Ei ole kuitenkaan kohtuullista edellyttää, että vaikkapa tutkimusartikkelin taustalla olevasta näytöstä kiinnostunut klikkaa jokaisen dataviitteen auki saadakseen edes pintapuolisen käsityksen viitatuista aineistoista. Jotta dataviittaus olisi myös ihmisluettava, tarvitaan pysyvän tunnisteen lisäksi muitakin tietoelementtejä: tekijätiedot, aineiston jakelijan, eli data-arkiston johon se on talletettu, aineiston tuottamisen ajankohdan sekä aineiston nimen. Näiden välttämättömien tietoelementtien lisäksi tiekartassa suositellaan mahdollisuuksien mukaan liittämään viittaukseen myös tekijän/tekijöiden ORCID-tunnukset, lisenssistatus, tieto mahdollisesta aineiston saatavuutta rajoittavasta embargo-ajasta, aineiston tiedostomuoto ja versionumero.

## FAIR-tutkimusdata on helposti viitattavaa

Tutkimusdataa koskevat FAIR-periaatteet (Wilkinson ym., 2016) ovat hyödyllinen käsitteellinen työkalu avaamaan tutkimusdataa koskevaa hyvää tieteellistä käytäntöä ja avoimen tieteen mukaista aineistonhallintaa. FAIR on lyhenne, joka tulee englanninkielen sanoista findable (löydettävä), accessible (saatavissa oleva), interoperable (yhteentoimiva) ja reusable (uudelleen käytettävä). Periaatteet on otettu osaksi hyvän tieteellisen käytännön määritelmää esimerkiksi eurooppalaisten akatemioiden organisaation ALLEAN ohjeistuksessa, jota sovelletaan kaikessa Horisontti 2020 -rahoitusta saaneessa tutkimuksessa (ALLEA, 2017).

FAIR-periaatteiden mukaan hallinnoituun tutkimusdataan viittaaminen on helppoa. Periaatteet toteutuvat, kun aineistonhallinnassa on huomioitu laadukkaasti metadatan tuottaminen, eli datan kuvaileminen, sekä tallentaminen luotettavaan data-arkistoon, jossa data saa pysyvän tunnisteen ja kuvailusivun. Luotettavia ja FAIR-periaatteiden mukaisia palveluja tarjoavat Suomessa esimerkiksi Tietoarkisto ja Tieteen tietotekniikan keskuksen CSC:n Fairdata-palvelukokonaisuus.<sup>2</sup>

## Viitattava data edellyttää datanhallintataitoja tutkimusprosessin alusta lähtien

Ennen kuin FAIR-periaatteet toteutuvat ja data on viitattavissa, on täytynyt tapahtua jo paljon. Kaikki perustuu sille, että tutkija on hallinnut dataansa asianmukaisella tavalla. Voidaankin todeta, että datanhallintataidot ovat nykyään oleellinen osa tutkijataitoja.

Helpoin tapa huolehtia oman tutkimusdatan viitattavuudesta on laatia huolellinen aineistonhallintasuunnitelma heti tutkimushankkeen alussa ja noudattaa sitä.

Datanhallinta käsittää koko tutkimusprosessin elinkaaren ja siinä otetaan huomioon myös datan mahdollinen uudelleen- ja jatkokäyttö. Datanhallintaa onkin suunniteltava jo samassa yhteydessä, kun laaditaan tutkimussuunnitelma, ja monilla tutkimusrahoittajilla datanhallintasuunnitelma on jo pakollinen osa tutkimusrahoitushakemusta.

Datanhallinnan suunnittelun apuvälineeksi on tarjolla työkalu, DMPTuuli, jonka kautta on pääsy niin rahoittajien, yliopistojen omiin, Tietoarkiston kuin

---

2 Lisätietoa Fairdata-palveluista löytyy osoitteesta <https://www.fairdata.fi/>

kansallisiin ohjeisiin. Tärkeä hyödynnettävä resurssi ovat myös yliopistokirjastojen datanhallinnan asiantuntijat ja laatimansa datanhallintaoppaat.

Aineistohallinta muodostuu viidestä pääteemasta (ks. tarkemmin SA ja DMPTuuli<sup>3</sup>):

1. Aineiston yleiskuvaus ja laatu
2. Aineiston hallintaan liittyvien eettisten periaatteiden ja lainsäädännön noudattaminen
3. Aineiston dokumentointi ja meta- eli kuvailutiedot
4. Tallentaminen ja varmuuskopiointi hankkeen aikana
5. Aineiston avaaminen ja pitkäaikaissäilytys hankkeen päätyttyä

Näitä kaikkia osa-alueita tulee siis pohtia mahdollisimman varhaisessa vaiheessa tutkimusprosessia ja jo tehtyjä ratkaisuja pitäisi tarkastella ja tarpeen mukaan päivittää prosessin edetessä. Datan viitattavuus pohjautuu hyvään aineistohallintatapaan.

## Lisää tieteenalakohtaista keskustelua tarvitaan

Tutkimusdataan viittaaminen: tiekartasta käytäntöön -tapahtuma sisälsi pienryhmätyöskentelyä kolmen dataviittauksiin liittyvän teeman ympärillä, koskien datatekijyyden määrittelyä, yliopistoissa tuotettavien data-aineistojen metatietojen keruuta sekä viittauskäytäntöjen vaikutusta tutkimuspalveluihin.

Keskusteluissa todettiin, että dataan viittaamisen pullonkaulat ovat luonteeltaan toimintatapoihin liittyviä, eivät teknisiä. Esimerkiksi Tietoarkiston prosessit ovat jo nyt linjassa tiekartan suositusten kanssa<sup>4</sup>. Merkittävin pullonkaula on viitattavan datan vähäisyys, sillä suurin osa tutkimusdatasta on pirstaloitunut tutkijoiden ja tutkimusryhmien kovalevyille, muistitikuille ja jopa disketeille (Ala-Kyyny, Korhonen, & Roinila, 2017).

Tekijyys ja omistajuus ovat kaksi erillistä käsitettä, jotka sekoittuvat helposti tutkijoiden mielissä. Kumpaankin liittyy avoimia kysymyksiä, jotka edellyttäisivät sekä organisaatio- että tieteenalakohtaista keskustelua. Samoin granulariteettiin, eli siihen millä tarkkuustasolla dataviittaus tulisi kohdistaa aineistokokonaisuuteen, olisi etsittävä kysymyksiä tieteenalayhteisöjen piiristä. Tieteenalakohtaisen keskustelun haasteena on siloutumisen välttäminen.

3 Ks. Suomen Akatemia (SA, 20.6.2018): Aineistohallintasuunnitelma, osoitteessa <https://www.aka.fi/fi/rahoitus/nain-haet/hakemuksen-liitteet/aineistohallintasuunnitelma/>. DMPTuuli löytyy osoitteesta <https://dmptuuli.fi/>

4 Dataan viittamisesta Tietoarkiston kotisivuilla: <http://www.fsd.uta.fi/fi/aineistot/jatkokaytto/viittaaminen.html>

## Tiekartta liittyy kansainväliseen keskusteluun

Data-asiain kansalliskomitea on kansainvälisen tiedeneuvoston ICSUn datakomitean CODATAN suomalainen jäsenkomitea. Sen kattojärjestönä toimii kotimaisten tiedeakatemioiden kansainvälisiä jäsenyyksiä hoitava yhteistyöelin Suomen Tiedeakatemit.

CODATA asetti vuonna 2012 työryhmän pohtimaan tutkimusdataan viittaamista. Syntyi Out of Cite, Out of Mind -raportti (CODATA-ICSTI, 2013). Raportti toimi pohjana tiedepoliittisen FORCE<sub>11</sub>-yhteisön dataviittauksia koskevalle julkilausumalle (Data Citation Synthesis Group, 2014). Edistääkseen dataviittauskäytäntöjen soveltamista paikallisella tasolla CODATA käynnisti vuonna 2015 dataviittaustyöpajojen sarjan. Syksyllä 2016 Data-asiain kansalliskomitea järjesti sarjan pohjoismaisen dataviittaustyöpajan ATT-hankkeen järjestämän Nordic Open Science Forum -tapahtuman yhteydessä Helsingissä.<sup>5</sup>

Data-asiain kansalliskomitea asetti ATT-hankkeen tilaamaa dataviittauselävitystä ohjaamaan asiantuntijaryhmän, johon tuli edustajia tiedekirjastoista, yliopistoista sekä data-arkistoista.<sup>6</sup> Näkemysten kirjon laajentamiseksi järjestettiin lisäksi avoin sidosryhmätyöpaja. Tiekartan ensimmäinen versio julkaistiin kaikkien vapaasti kommentoitavaksi syksyllä 2017 ja palautteen pohjalta muokattu lopullinen versio huhtikuussa 2018.

Kansainvälisen yhteentoimivuuden varmistamiseksi tiekartta toteutettiin englanninkielisenä. Siinä hyödynnetään laajasti FORCE<sub>11</sub>-dataviittausperiaatteita. Tiekarttaa voidaan pitää eräänlaisena FORCE<sub>11</sub>-periaatteiden kansallisen ja käytännön tason soveltamisohjeena.

## Viitteet

- Ala-Kyyny, J. P. (2018). Data matkalla tunnustetuksi tutkimustuotokseksi – tiekartta ohjaa tiedeyhteisöä kohti parempia dataviittauskäytäntöjä. *Think Open*, 22.5.2018. <https://blogs.helsinki.fi/thinkopen/data-matkalla-tunnustetuksi-tutkimustuotokseksi-tiekartta-ohjaa-tiedeyhteisoa-kohti-parempia-dataviittauskaytantoja/>
- Ala-Kyyny, J., Korhonen, T., & Roinila, M. (2017). Tutkimusdatan avaamisen esteet: haastattelu-tutkimus Helsingin yliopistossa. *Signum*, 49(4), 25–29. <https://doi.org/10.25033/sig.69198>
- ALLEA. (2017). The european code of conduct for research integrity - revised edition. Berlin: ALLEA – All European Academies. <http://www.alllea.org/wp-content/uploads/2017/04/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017.pdf>

---

5 Videotallelle Nordic CODATA Data Citation Workshop -tapahtumasta löytyy osoitteesta <https://www.youtube.com/watch?v=jw8p5ft44hU>

6 Data-asiain kansalliskomitean kotisivuilla on lisätietoja projektista osoitteessa <https://www.fcrd.fi/data-citation/>. Katso myös Ala-Kyyny (2018) ja Suomen tiedeakatemit (2017).

- CODATA-ICSTI. (2013). Out of cite, out of mind: the current state of practice, policy, and technology for the citation of data. *Data Science Journal*, 12. <https://doi.org/10.2481/dsj.050M13-043>
- Data Citation Synthesis Group. (2014). *Joint declaration of data citation principles*. (M. Martone, toim.), FORCE11. San Diego: FORCE11. <https://doi.org/10.25490/a97f-egy>
- Finnish Committee for Research Data. (2018). *Tracing data: data citation roadmap for Finland*. (H. Laine, toim.). Helsinki: Finnish Committee for Research Data. <http://urn.fi/URN:NBN:fi-fe201804106446>
- Suomen tiedeakatemit. (2017). Datan jäljillä – miksi dataa tulisi julkaista ja miksi dataan tulisi viitata? Haastattelussa Heidi Laine Data-asiain kansalliskomiteasta. *Suomen Tiedeakatemit*. 10.4.2017. <http://www.academies.fi/datan-jaljilla-miksi-dataa-tulisi-julkaista-ja-miksi-dataan-tulisi-viitata-haastattelussa-heidi-laine-data-asiain-kansalliskomiteasta/>
- TENK. (2013). Hyvä tieteellinen käytäntö ja sen loukkausepäilyjen käsitteleminen Suomessa: tutkimuseettisen neuvottelukunnan ohje 2012. Helsinki: Tutkimuseettinen neuvottelukunta. <http://www.tenk.fi/fi/hyva-tieteellinen-kaytanta>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*. <https://doi.org/10.1038/sdata.2016.18>