

<https://helda.helsinki.fi>

---

## Sosiaalitieteiden tiedon avoin saatavuus : Haasteita ja ratkaisuja

Komp-Leukkunen, Kathrin

2022

---

Komp-Leukkunen , K & Lagus , K 2022 , ' Sosiaalitieteiden tiedon avoin saatavuus :  
Haasteita ja ratkaisuja ' , Janus : sosiaalipolitiikan ja sosiaalityön tutkimuksen aikakauslehti ,  
Vuosikerta. 30 , Nro 3 , Sivut 296-301 .

---

<http://hdl.handle.net/10138/350092>

---

*Downloaded from Helda, University of Helsinki institutional repository.*

*This is an electronic reprint of the original article.*

*This reprint may differ from the original in pagination and typographic detail.*

*Please cite the original version.*

# SOSIAALITIEDEIDEN TIEDON AVOIN SAATAVUUS: HAASTEITA JA RATKAISUJA

**Kathrin Komp-Leukkunen:** *apulaisprofessori, Helsingin yliopisto*

**Krista Lagus:** *professori, Helsingin yliopisto*

*kathrin.komp@helsinki.fi; krista.lagus@helsinki.fi*

*Janus* vol. 30 (3) 2022, 296–301

## JOHDANTO

Avoimien tieteiden yleistyminen on yleistyminen. Avoimien tieteiden tarkoittaa, että tieteellinen näkemys jaetaan avoimesti varhain, mieluiten pian sen luomisen jälkeen. Avoimien tieteiden käytäntö on jo saanut jalansijaa julkaisu- ja tiedonvälityksessä, jossa avoimien julkaisujen julkaisu- ja tiedonvälitykseen yleistyvät ja yhä useammat rahoittajat jopa vaativat niitä. Avoimien tieteiden määrittely avoimien pääsyn julkaisuja pidemmälle, sisältäen muun muassa avoimien pääsyn tutkimusdataan ja yhteistyöhön perustuvan tiedon luomisen (Fecher & Friesike 2014). Näillä muilla alueilla avoimien tieteiden periaatteet toteutuvat toistaiseksi harvemmin. Avoimien tieteiden periaatteiden vahvistaminen näillä alueilla on lähivuosien tärkeimpiä tehtäviä tutkimusyhteisössä. Avoimien pääsyn tarjoaminen tutkimustietoihin voi olla haastava tehtävä, varsinkin ensimmäisillä kerroilla. Virheitä ja väärinkäsityksiä voi tapahtua, ja muutokset itse tutkimusprosessissa voivat vaikeuttaa tätä toimintaa. Yhteiskuntatieteiden tutkijat kohtaavat erityisiä haasteita tarjotessaan avoimien pääsyn tietoihinsa. Ensinnäkin, koska he tutkivat ihmisiä, tutkimuksen kohteet ovat usein tunnistettavissa tutkimusaineistosta. Toiseksi, osa henkilöistä suostuu osallistumaan tutkimukseen vain, jos tutkija takaa, että tietoja ei jaeta kenenkään kanssa. Ja kolmanneksi, jotkin tiedot ovat järkeviä vain silloin,

kun niitä tulkitaan niiden keräämisen yhteydessä, kuten ympäristön olosuhteet etnografisissa tutkimuksissa tai haastattelijan ja haastateltavan suhde narratiivisissa haastatteluissa. Siksi yhteiskuntatieteilijät eivät voi noudattaa yhtä standardisoitua lähestymistapaa, kun he asettavat tietojensa saataville avoimien pääsyn kautta. Sen sijaan heidän on löydettävä ratkaisu, joka sopii heidän erityiseen tietoon ja tiedonkeruutapaan.

Tämä puheenvuoro nostaa esiin keskeisiä teemoja joulukuussa 2021 Helsingin yliopistossa järjestetystä “Open access to social science data: challenges and solutions” -verkkotapahtumasta. Tapahtumassa tutkittiin, kuinka yhteiskuntatieteilijät suoriutuvat harkitessaan tietojensa avointa saatavuutta. Siinä käsiteltiin erilaisten aineistojen parissa työskentelevien tutkijoiden kokemuksia, heidän saatavillaan olevaa tukea, ja se sisälsi yhdessä luovan ja ajatuksia jäsentävän osuuden.

## KOKEMUKSIA SOSIAALITIEDEIDEN AVOIMEN SAATAVUUDEN TIEDON TUOTOSTA

Kokemukset tutkimusaineiston avoimesta saatavuudesta vaihtelevat suuresti. Yksi syy kokemusten erilaisuuteen ovat erot tiedoissa. Avoimien pääsyn tarjoamisen menetelmät vaihtelevat riippuen

siitä, onko tietoaaineisto big dataa, kyselytietoa, julkisesti saatavilla olevaa tietoa tai kvalitatiivista aineistoa. Tilaisuudessa neljä erilaisten aineistotyyppien parissa työskentelevää tutkijaa kertoi kokemuksistaan.

Big dataksi eli massadataksi kutsutaan aineistoa, joka on niin suurta ja kompleksista että sen keruuseen, analyysiin ja käsittelyyn tarvitaan poikkeuksellisia menetelmiä. Usein sillä viitataan aineistoihin, jotka ovat syntyneet luonnollisesti ja suoraan digitaalisissa järjestelmissä ja prosesseissa, niiden käytön sivutuotteena, esimerkiksi kun mobiilisolvelluksilla käytetään internetpalveluita (De Mauro et al. 2016). Tällaista aineistoa ovat esimerkiksi suuret internetistä kerätyt keskusteluaineistot, joita voidaan tutkia vaikkapa kirjoittajien välisten kytkentöjen, keskustelun ajallisen rytmittymisen, keskustelunaiheiden jäsentämisen tai käsitteiden löytämisen näkökulmista. Krista Lagus Helsingin Yliopiston Yhteiskuntatieteiden keskukselta (Center for Social Data Science) kertoi kokemuksista tämän tyyppisen Suomi24-aineiston tuomisesta osaksi Kielipankin kaikille tutkijoille avoimia varantoja. Poikkeuksellisesti aineisto päätettiin tuoda Kielipankkiin ennen sen tutkimista itse. Lagus toi esiin, että vaikka tämä altruistinen toiminta rajapinnassa olevilla tutkijoilla näyttäytyi ennalta mahdollisena urariskinä, se ei kuitenkaan käytännössä osoittautunut urariskiksi ydintoimijoille, vaan päinvastoin he kaikki etenivät merkittävästi akateemisilla urillaan. Hän myös nosti esiin yhdessä tunnistetun ja jaetun motivaation ja inspiroitumisen merkityksen ja toisaalta eri organisaatioiden ja toimijoiden välisen yhteistyön ja henkilökohtaisten verkos-

tojen ja kohtaamisten roolin, kun pyritään tuomaan jokin kaupallisen alan omistavan yrityksen hallussa oleva big data-aineisto yhteisesti tutkijoiden käyttöön. Lagus kertoi tavoista, joilla projektissa oli pyritty luomaan edellytyksiä sille, että aineiston avaajat saavat myös akateemista tunnustusta työstään. Tässä keskeisenä välineenä oli viittauksittuudesta puhuminen hankkeen ja sen verkostojen piirissä sekä aineistoraportin julkaiseminen, jossa kerrotaan aineiston syntyprosessista ja luonteesta (Lagus et al. 2016). Tämä näyttäytyi melko onnistuneena kokeiluna.

Kyselytieto on kvantitatiivista tietoa, jota analysoidaan tilastollisissa analyyseissä. Se kerätään vastaajilta nimenomaista tieteellistä tutkimusta varten. Hanna Wass Helsingin yliopiston valtiotieteen tieteenalalta kertoi kokemuksistaan tämänkaltaisesta tiedosta BIBU-projektin puitteissa. BIBU on Suomen Akatemian strategisen tutkimuksen toimikunnan rahoittama ”Tackling biases and bubbles in participation” -hanke. Hankkeessa on tähän mennessä tehty viisi kyselyä, joista yksi yhdistettiin Suomen rekisteritietoihin. Wass kertoi, että itse kerätyille tiedoille avoin pääsy on helpommin saavutettavissa. Se voidaan kuitenkin tehdä vasta projektikauden päätyttyä, koska projektin jäsenten on ensin saavutettava julkaisutavoitteensa. Rekisteritietoja sisältävän kyselyn osan saattaminen avoimesti saataville tulee kuitenkin olemaan vaikeaa ellei mahdotonta. Tässä tulee huomioida tiedon omistajan, Tilastokeskuksen, ja GDPR:n ohjeet. Wass otti myös esiin tärkeän kohdan tietojen omistajuuden harkitsemisesta. Hänen käsityksensä mukaan tiedoista tulee julkista hyödykettä projektin pää-

tyttyä, koska projekti tiedon omistajana lakkaa olemasta. Hänelle tämä on luonnollinen argumentti tutkimustiedon saattamiseksi avoimesti saataville projektin päätyttyä.

Julkisesti saatavilla olevat tiedot ovat vapaasti saatavilla Internetissä, esimerkiksi hallitusten kotisivuilla. Sitä voi katsoa kuka tahansa ilman rajoituksia. Heikki Hiilamo Terveiden ja hyvinvoinnin laitoksesta ja Helsingin yliopiston sosiaalipolitiikan tieteenalasta kertoi kokemuksistaan tämänkaltaisista tiedoista COVID-19-yhteiskuntapoliittisia toimenpiteitä koskevassa hankkeessa. Hän koki, että vapaasti saatavilla olevan tiedon kanssa avoin pääsy itsessään oli vain pieni haaste. Hänen mukaansa pääkysymys oli projektin synnyttämän työtaakan hallinta. Myös projektin luonteesta johtuen tieto voitiin tuoda avoimesti saataville vain vaiheittain, jotta projektiryhmä voisi ensin täyttää julkaisutavoitteensa tietoa hyödyntäen.

Laadullista tietoa kerätään muun muassa haastattelujen ja havaintojen avulla. Se kerätään nimenomaista tieteellistä analyysiä varten. Helsingin yliopiston sosiaalipolitiikan tieteenalan tohtoriopiskelija Jawaria Khan kertoi kokemuksistaan tämänkaltaisista tiedoista. Samalla hän kuvasi jatko-opiskelijoiden näkemystä tutkimusaineiston avoimesta saatavuudesta. Khan selitti, että laadullinen data on luonteeltaan hyvin monimuotoista, eikä se ole aina itsestään selvää. Esimerkiksi havaintotieto tarvitsee ymmärrystä kontekstista, ja tutkimusprosessin aikana otetuissa valokuviissa voi olla järkeä vain, jos valokuva on jo kuvailtu ja tulkittu. Haasteena on myös se, että laadullisen tiedon anonymisointi voi olla vaikeaa, ja se voidaan joskus

kerätä vain silloin, kun luvataan tutkitaville, että tiedot pidetään luottamuksellisina. Siksi hän katsoo, että laadullisia tietokokonaisuuksia on harkittava tapauskohtaisesti, ja joissakin tapauksissa vain osittainen pääsy tai ei lainkaan avointa pääsyä tietoihin voi olla sopiva ratkaisu. Khan huomautti lisäksi, että tohtoriopiskelijoille avoimen pääsyn antaminen tutkimustietoihin voi aiheuttaa pelkoa. Yksi pelko on, että muut voivat analysoida tietoja ja julkaista niitä antamatta tunnustusta sen kerääjälle. Toinen pelko on, että muut voivat toistaa tohtoriopiskelijoiden tekemiä tutkimuksia ja saada erilaisia tuloksia, mikä saattaisi herättää epäilyksiä alkuperäisen työn laadusta (Khan et al. 2021). Nämä huomiot kokonaisuutena osoittavat, että avoimen pääsyn tarjoaminen tutkimustietoihin liittyy yhtä lailla teknisiin ja organisatorisiin näkökohtiin kuin kokemuksiin ja odotuksiin.

#### TUKIPALVELUT SOSIAALITETEIDEN TIEDON AVOIMEEN SAATAVUUTEEN

Kun yhteiskuntatieteilijät harkitsevat tarjoavansa tietonsa avoimen saatavuuden mukaisesti, he voivat saada apua. Esimerkiksi monet yliopistot ja data-arkistot tukevat tätä toimintaa. Tätä tukea tarjoavat henkilöt voivat hyödyntää laajaa kokemustaan tietojen avoimesta pääsystä antaakseen yhteiskuntatieteilijöille hyödyllisiä vinkkejä.

Monica Allardt Helsingin yliopiston tietopalveluista kertoi, kuinka yhteiskuntatieteilijät voivat parhaiten lähestyä tietonsa tarjoamista avoimen pääsyn kautta. Hän huomautti, että vaikka avoin pääsy on toivottavaa ja suositeltavaa, se ei ole ihanteellinen ratkaisu

kaikille tiedoille. Joissakin tapauksissa on parempi olla antamatta avointa pääsyä tutkimustietoihin. Näin on silloin, kun tiedoissa on arkaluonteisia tietoja, kun koko data tai osa tiedoista kuuluu jollekin muulle, kun vastaajille on luvattu, että heidän tietojaan ei jaeta tai, kun vastaajilta ei ole kysytty, voidaan heidän tietojaan jakaa. Tästä syystä yhteiskuntatieteilijöiden on arvioitava jokaisen aineiston osalta erikseen, onko se mahdollista saada käyttöön avoimen pääsyn kautta. Lisäksi Allardt korosti, että on tärkeää aloittaa tutkimustietojen avoimen saatavuuden suunnittelu-prosessi ajoissa. Jotta tutkimusprojektin aikana ei muodostu ongelmia, avoin pääsy tulisi suunnitella jo ennen projektin alkamista. Suunnittelu voidaan muotoilla *tiedonhallintasuunnitelmaksi*, joka on dokumentti, joka kuvaa miten tutkimusdataa käsitellään tutkimusprojektin aikana ja sen päätyttyä (Cauchick-Miguel et al. 2020). Projektin aikana tulee tehdä muistiinpanoja tiedonkeruu- ja käsittelyprosessista. Nämä muistiinpanot voivat olla hyödyllisiä luotaessa metadataa, jota tarvitaan myöhemmin tietojen saattamiseksi avoimesti saataville.

Helena Laaksonen Tietoarkistosta ja Tampereen yliopistosta kertoi Tietoarkiston roolista tutkimusaineiston avoimen pääsyn prosessissa. Tiedot tulee tallentaa luotettavaan tietoarkistoon, joka voidaan tunnistaa CoreTrustSeal merkistä (<https://www.coretrustseal.org/>). Suomessa tämä sinetti on myönnetty Tietoarkistolle (<https://www.fsd.tuni.fi/en/>) ja Kielipankille (<https://www.kielipankki.fi/>). Tietoarkisto antaa tietoa tietoaaineistojen arkistoinnin toiminnasta ja auttaa tutkijoita tässä tehtävässä. Arkistointia varten tutki-

joiden tulee koota tutkimustietonsa ja sitä vastaavat metatiedot. Metadata on tietoa tutkimuksen teon ja tiedonkeruun kontekstista, esimerkiksi milloin ja kuka tiedon on kerännyt ja miten se on laadittu (Davidson et al. 2019). Se arkistoidaan yhdessä tutkimusaineiston kanssa ja siirretään myös muille tiedon tarjoajille. Näin mahdollisimman moni saa tietää, mitä tutkimusaineistoa arkistosta löytyy.

#### CO-CREATION TAPAHTUMA

Tapahtuman CoCreation-osuudessa osallistujat jaettiin pienryhmiin keskustelemaan, ja heitä pyydettiin ryhmäkeskustelujen aikana kirjoittamaan yhteiseen, kevyesti strukturoituun dokumenttiin kukin omia huomioitaan ja kysymyksiään ranskalaisiksi viivoiksi. Tämä jaetun rakenteisen valkotalun työskentely mahdollisti osallistujille yhtäaikaisen kirjoittamisen asynkronisena ja nopeasti etenevänä prosessina, jossa kaikki koko ajan näkivät mitä on syntymässä, ja samalla saattoivat vapaasti keskustella pienryhmissä huomioistaan. Se myös mahdollisti niin anonyymit huomiot kuin omalla nimellään kommentoinnin. Aikaa yhdessä kirjoittamiselle ja keskustelulle oli käytettävissä noin 15 minuuttia, ja siihen osallistui noin puolet muutoin tilaisuudessa mukana olleista. Lopuksi pienryhmät kertoivat lyhyesti siitä mistä olivat keskustelleet.

Co-creation tapahtuman kommentit ja keskustelut antoivat käsityksen yhteiskuntatieteellisen tiedon avoimen saatavuuden nykyisistä käytännöistä. Eniten keskusteltu kysymys oli, mitä esteitä henkilöt olivat kohdanneet koskien tietojen avointa saatavuutta. Tämä ky-

symys sai 15 kommenttia. Kommentit käsittelevät tiedon laatuongelmia, dokumentaatiota, käyttöoikeuksia, tietoa tiedon avoimesta pääsystä, erimielisyyksiä projektitiimin ja instituutioiden sisällä, juridisia kysymyksiä ja muutoksia projektin sisällä. Toiseksi eniten kommentteja saanut kysymys oli, mitä tunteita ihmisillä oli tiedon avoimesta saatavuudesta (3 kommenttia). Kuva-  
tut tunteet koskivat pelkoa siitä, että ei saada tunnustusta tietojen keräämisestä, että joku julkaisee tulokset ennen itseä, ja siitä, että ei tiedetä, mitä datalle tapahtuu sen jälkeen, kun se ”vapautetaan luontoon”. Vastajat tunnistivat myös kolme avointa kysymystä, joita heillä oli yhteiskuntatieteellisen tiedon avaamisesta: he halusivat tietää, miten datan päivityksiä käsitellään, voitaisiinko tiedon avoimesta pääsystä järjestää seminaarisarjaa ja voisiko tiedon avoimesta saatavuudesta tulla pakollista. Helena Laaksonen kommentoi jälkimmäistä, että avoimen pääsyn harkitseminen on mahdollista kaikissa tapauksissa, mutta toteutus ei ole mahdollista kaikissa tapauksissa.

Co-creation tapahtuman kommentit ja keskustelut osoittavat selvästi, että tutkimusaineiston avoimen saatavuuden tarjoaminen on vielä nuorta toimintaa, joka sisältää paljon kysymyksiä ja epävarmuustekijöitä. Haasteet ja epäonnistumiset ovat edelleen yleisiä. Saatavilla on kuitenkin myös tukea, joka voi vastata esitettyihin kysymyksiin.

#### YHTEENVETO

Tutkimustiedon avoin saatavuus on yhä tärkeämpää myös yhteiskuntatieteissä. Koska yhteiskuntatieteissä kerättä-

vä tieto on hyvin monimuotoista, on avoin pääsy siihen parasta järjestää yksilöllisesti, tai aineistotyypeittäin haasteita jäsentäen ja niihin varautuen. Tämän avoimen pääsyn suunnittelu tulisi mieluiten tapahtua ennen projektin alkua ja se tavallisimmin tulee ajankohtaiseksi vasta projektin päätyttyä. Kaikissa esillä olleissa tapauksissa avoimen pääsyn rakentaminen tutkimusaineistoon edellyttää laajempaa aikahorisonttia kuin mitä itse tutkimushanke asettaa. Aineistojen avaaminen on näiden kokemusten pohjalta tarkasteltuna hyvin paljon työtä vaativa vaihe tutkimuksen prosessissa, ja sen tekijöille ei välttämättä ole nykyisellään riittäviä akateemisia kannustimia työstä. Vaikuttaa edelleen olevan avoin kysymys, millä eri tavoilla tutkimusaineiston avaamisesta voidaan tehdä sekä palkkatyönä että akateemisesti viranhauissa tunnustettua toimintaa. Tämä saattaa edellyttää sekä aineistojen avaamiseen keskittyviä akateemisia julkaisukanavia että akateemisten rahoitusmuotojen suuntaamista tähän siten, että ne ajoittuvat oikea-aikaisesti mahdollisesti jo päättyneiden tutkimushankkeiden kannalta.

#### KIITOKSET

Kiitämme Helsingin yliopiston humanistis-yhteiskuntatieteellistä instituuttia tuesta. Kiitämme Lauri Leukkusta teknisestä avusta ja tekstin oikolukemisesta.

#### KIRJALLISUUS

Cauchick-Miguel, Paulo & Moro, Suzana & Rivera, Roberto & Amorim, Marlene (2020) Data management plan in rese-

- arch: Characteristics and development. Teoksessa Rogero Mugnaini (toim.) *Data and Information in Online Environments*. Cham: Springer, 3–14. [https://doi.org/10.1007/978-3-030-50072-6\\_1](https://doi.org/10.1007/978-3-030-50072-6_1)
- Davidson, Emma & Edwards, Rosalind & Jamieson, Lynn & Weller, Susie (2019) Big data, qualitative style: A breadth-and-depth method for working with large amounts of secondary qualitative data. *Quality & Quantity* 53, 363–376. <https://doi.org/10.1007/s11135-018-0757-y>
- De Mauro, Andrea & Greco, Marco & Grimaldi, Michele (2016) A formal definition of Big Data based on its essential features. *Library Review* 65 (3), 122–135. <https://doi.org/10.1108/LR-06-2015-0061>
- Fecher, Benedikt & Friesike, Sascha (2014) Open science: one term, five schools of thought. Teoksessa Sönke Bartling & Sascha Friesike (toim.) *Opening science. The evolving guide on how the internet is changing research, collaboration and scholarly publishing*. Dordrecht: Springer, 17–48. [https://doi.org/10.1007/978-3-319-00026-8\\_2](https://doi.org/10.1007/978-3-319-00026-8_2)
- Khan, Jawaria & Rantanen, Visa & Kolmonen, Petteri (2021) Open access to qualitative research data: A guide for PhD students in the social sciences at the University of Helsinki. *Population ageing and its social consequences working paper no. 9*. Helsinki: University of Helsinki. <https://www2.helsinki.fi/en/researchgroups/population-ageing/working-papers>
- Lagus, Krista & Ruckenstein, Minna & Pantzar, Mika & Ylisiurua, Marjoriikka (2016) Suomi 24 – Muodonantaa aineistolle. *Kuluttajatutkimuskeskus, Valtiotieteellisen tiedekunnan julkaisuja* 2016:10. Helsinki: Helsingin yliopisto. <https://helda.helsinki.fi/handle/10138/163190>