

# Tampereen yliopiston tutkimusaineistoseelvitys 2017

Saila Huuskonen

*Tampereen yliopiston kirjasto*

saila.huuskonen@staff.uta.fi

<https://orcid.org/0000-0002-2316-285X>

Tomi Toikko

*Tampereen yliopiston kirjasto*

tomi.toikko@staff.uta.fi

Monien muiden eurooppalaisten (Tenopir ym., 2017) ja suomalaisten kirjastojen (Salmi & Kuusniemi, 2016) tavoin Tampereen yliopiston kirjasto on viime vuosina kehittänyt tutkimusdatapalveluita. Palveluiden kehittämisessä keskeistä on ymmärrys siitä, millaista tutkimusdataa tutkijat hyödyntävät.

Tampereen yliopistossa keväällä 2017 toteutetussa avoimen tieteen kyselyssä<sup>1</sup> selvisi, että vain pieni osa tutkijoista hyödyntää data-arkistoja aineistonsa tallentamiseen. Suurimmat esteet tutkimusaineistojen avaamiselle ovat tutkimuseettiset syyt, tietämättömyys aineistojen avaamisesta, avaamisen työmäärä ja tutkimusryhmässä tehdyn selvän sopimuksen puute (vrt. Tenopir ym., 2015). Noin puolet vastaajista oli kuitenkin hyödyntänyt muiden kokoamaan tutkimusaineistoa tutkimuksessa tai opetuksessa (vrt. Curty, Crowston, Specht, Grant, & Dalton, 2017; Kim & Yoon, 2017).

---

1 [https://www.uta.fi/sites/default/files/inline-files/Avoin\\_tiedekysely\\_%20tulokset.pdf](https://www.uta.fi/sites/default/files/inline-files/Avoin_tiedekysely_%20tulokset.pdf)

Jatko-osana avoimen tieteen kyselylle, selvitimme artikkeliaineistolla (1) millaisia tutkimusaineistoja yliopistossamme käytetään (2) millaisia määriä tutkimusaineistoja on (3) millaista työmäärää aineiston kerääminen voi vaatia (4) onko dataa avattu ja (5) millaisia eettisiä näkökulmia aineistoasioihin liittyy.

Aineistonamme on otos Tampereen yliopiston tutkijoiden alkuvuonna 2017 julkaistuista OKM:n luokituksen<sup>2</sup> mukaisista A1-artikkeleista (vertaisarvioidut tieteelliset artikkelit). Aineisto otettiin ulos Tampereen yliopiston tutkimustietojärjestelmästä<sup>3</sup> 10.5.2017, jolloin julkaisuja oli 307. Datasta tehtiin ensiksi 20% otanta kustakin OKM:n päätieteenalasta<sup>4</sup>. Toisessa vaiheessa otettiin jälleen 20% kustakin OKM:n päätieteenalasta. Lopulliseen analyysiin jäi 129 artikkelia eli 42% koko datasta.

Analysoiduista artikkeleista yli puolet oli lääke- ja terveystieteistä (n=72). Yhteiskuntatieteistä analyysiin sisältyi 35, luonnontieteistä 12, humanistisista tieteistä 8 ja tekniikan alalta 2 artikkelia. Artikkeleista hieman alle puolet (n=56) oli kansainvälisiä yhteisjulkaisuja. Julkaisufoorumitasoille artikkelit jakautuivat seuraavasti: taso 0: n=5, taso 1: n=82, taso 2: n=29 ja taso 3: n=13. Analyysiin otettiin mukaan vain suomenkieliset (n=4) ja englanninkieliset (n=125) artikkelit.

Otoksemme artikkeleissa oli käytetty yhteensä 155 eri aineistoa. Näistä 83 aineistoa oli itse kerätty. Ilman tutkimusta olemassa olevia aineistoja oli 46 ja toisen tutkijan keräämiä aineistoja 26 kappaletta. Kahdeksassa artikkelissa ei ollut mainintaa aineistoista.

Datasettien määrät vaihtelivat tutkimuksittain nollan ja kahdeksan välillä, mediaanituloksen ollessa kaksi datasettiä per artikkeli ja keskiarvon ollessa 2,16.

Artikkelitoksemme toi esille laajan kirjon erilaisia tutkimusaineistoja. Otoksessamme korostuivat erityisesti erilaiset lääketieteelliset näytteet, kyselyt ja erityyppiset haastattelut. Tämän lisäksi tutkimusaineistona käytettiin esimerkiksi tilastoja, blogeja ja potilastietoja.

Aineiston keruun ja hallinnan työmäärää on vaikea laskea ja yhteismitalistaa, koska otoksemme aineistot ovat hyvin heterogeenisia. Otoksemme artikkeleissa oli tehty näkyväksi erilaisia tutkimusdatan esityöstämisen vaiheita. Lääke- ja biotieteiden tutkimuksissa kuvattiin esimerkiksi soluviljelmien ja näytteiden valmistelua ja laadullisissa tutkimuksissa esimerkiksi haastattelujen litterointia. Otoksessamme näkyi myös, kuinka datan kerääminen voi ajoittua jopa yli vuosikymmenien tai vaatia vuosien kenttäjaksoja. Tutkimus myös sitoo resursseja esimerkiksi haastateltavilta, kyselyyn vastaajilta, koehenkilöiltä ja

---

2 <https://confluence.csc.fi/display/tutki/OKM%3An+julkaisutyypipiilokitus>

3 <https://solecris.uta.fi>

4 <https://confluence.csc.fi/pages/viewpage.action?pageId=3850296>

lääketieteelliseen tutkimukseen osallistuvilta. Osallistujien määrä voi vaihdella esimerkiksi muutamista haastateltavista kymmenien tuhansien henkilöiden lääketieteellisiin seurantatutkimuksiin.

Tutkimuseettiset asiat olivat vahvasti esillä otoksessamme. Yli puolet (n=69) artikkeleista sisälsi mainintoja erilaisista eettisistä luvista ja suostumuksista. Lupia oli haettu eettisiltä toimikunnilta ja valtakunnalliselta eläinkoelautakunnalta. Yksilötasolla suostumusta tutkimukseen osallistumisesta oli kysytty niin tutkittavilta (n=38) kuin näiden läheisiltä (n=13). Tutkimusluvan myöntäjinä olivat esimerkiksi rekisterinpitäjät ja organisaatiot. Eturistiriidoista tai niiden puuttumisesta mainittiin yli puolessa (n=66) artikkeleista. Lisäksi esille nousi esimerkiksi pseudonymisointi ja anonymisointi.

Artikkeliotoksessamme oli erittäin vähän mainintoja (n=10) aineiston tallennuksesta jatkokäyttöä varten. Aineiston mainittiin olevan saatavilla kirjoittajalta pyydettyäessä, data-arkistossa, tutkimusyhteisön verkkosivuilta tai lehden verkkosivuilta.

Tutkimusaineistoja kerätään ja käsitellään monin eri tavoin ja niiden määrä vaihtelee suuresti sekä tieteenalojen välillä että niiden sisällä. Tutkimusaineistojen avaaminen näyttää olevan vielä vähäistä. Tätä saattaa selittää esimerkiksi aineiston sensitiivisyys, omistajuus ja tutkijoiden halu ensikäyttää keräämäänsä aineistoa. Keskustelu aineistojen avaamisesta on ollut pinnalla vasta muutaman vuoden, joten avoimuuden ideologia ei välttämättä näy vielä tässä otoksessa.

Tutkimusdatan heterogeenisyys luo haasteita myös yliopistojen tuottamille datapalveluille, joiden on osattava auttaa tutkijoita monissa eri datan elinkaaren vaiheissa suunnittelusta mahdolliseen avaamiseen. Tätä varten tarvitaankin monien eri toimijoiden verkosto tukemaan datanhallintaa.

Kiitokset selvityksen tiedot keränneille Samppa Penttiselle ja Erno Juvoselle.

## Lähteet

- Curry, R. G., Crowston, K., Specht, A., Grant, B. W., & Dalton, E. D. (2017). Attitudes and norms affecting scientists' data reuse. *PLOS ONE*, 12(12). <https://doi.org/10.1371/journal.pone.0189288>
- Kim, Y., & Yoon, A. (2017). Scientists' data reuse behaviors: A multilevel analysis. *Journal of the Association for Information Science and Technology*, 68(12), 2709–2719. <https://doi.org/10.1002/asi.23892>
- Salmi, A., & Kuusniemi, M. E. (2016). Project MILDRED: Charting Ground for Research Data Management Services at University of Helsinki. *Informaatiotutkimus*, 35(3), 71–72. <https://journal.fi/inf/article/view/59446>
- Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., ... Dorsett, K. (2015). Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLOS ONE*, 10(8), e0134826. <https://doi.org/10.1371/journal.pone.0134826>

Tenopir, C., Talja, S., Horstmann, W., Late, E., Hughes, D., Pollock, D., ... Allard, S. (2017). Research Data Services in European Academic Research Libraries. *LIBER Quarterly*, 27(1), 23-44. <https://doi.org/10.18352/lq.10180>