

# Tutkimusaineiston avoin jakaminen – tutkimusorganisaatioiden jäsenten käsityksiä, kokemuksia ja mielipiteitä

Heidi Enwald

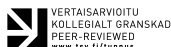
*Oulun yliopisto*

heidi.enwald@oulu.fi

<https://orcid.org/0000-0003-1953-2157>

Open research data is data that is free to access, reuse, and redistribute. This study focuses on the perceptions, opinions and experiences of staff and researchers of research institutes on topics related to open research data. Furthermore, the differences across gender, role in the research organization and research field were investigated. An international questionnaire survey, translated into Finnish and Swedish, was used as the data collection instrument. An online survey was distributed through an open science related network to Finnish research organizations. In the end, 469 responded to all 24 questions of the survey. Findings indicate that many are still unaware or uncertain about issues related to data sharing and long-term data storage. Women as well as staff and researchers of medical and health sciences were most concerned about the possible problems associated with data sharing. Those in the beginning of their scientific careers, hesitated about sharing their data.

Asiasanat: tutkimusaineisto; avoin tieto; tutkimuslaitokset; kyselytutkimus; tietoarkistot



Avoimeen tieteeseen liittyvät aiheet ovat viime aikoina olleet esillä niin tiedemaailmaan kuin yhteiskunnallisen vaikuttavuuteenkin liittyvissä keskusteluissa ja avauksissa. Tutkimusaineistojen ja -datan avoimen saatavuuden ja jaettavuuden lisääminen on osa avointa tiedettä. Avoin tiede voidaan nähdä tiedepoliittisena liikkeenä, joka on noussut kansainvälisesti merkittäväksi tavaksi edistää aineistojen käyttöä sekä tieteen näkyvyyttä ja vaikuttavuutta yhteiskunnassa. “Tieteellisen tiedon ja sen erilaisten tutkimustuotteiden, kuten julkaisujen ja

Artikkeli on lisensoitu Creative Commons Nimeä-EiKaupallinen-JaaSamoin 4.0 Kansainvälinen -lisenssillä

Pysyvä osoite: <https://doi.org/10.23978/inf.77411>

datan, vapaa saatavuus verkossa parantaa tieteen laatua, tukee demokraattista päätöksentekoa, lisää kansalaisten ymmärrystä ja tietoisuutta sekä palvelee elinkeinoelämää” (Suomen yliopistojen rehtorineuvosto UNIFI ry, 2018, s. 1).

Tämä artikkeli käsittelee kansainvälisen kyselyhankkeen Suomen aineiston tuloksia. Artikkelissa keskitytään erityisesti kyselyn avoimeen jakamiseen ja pitkäaikaissäilytykseen liittyviin kysymyksiin ja näiden kysymysten ja vastaajan sukupuolen, roolin tiedeyhteisössä sekä tieteenalan suhteeseen.

## Tutkimusaineiston jakaminen

Avoimen tieteen keskeisiin periaatteisiin kuuluu vastuullisuus: sen tavoitteena on olla niin avointa kuin mahdollista, mutta kuitenkin niin suljettua kuin välttämätöntä. Digitaalisessa muodossa olevien tutkimusaineistojen eli tutkimusdatan avoimeksi saattamisessa onkin otettava huomioon avaamiseen liittyvät rajoitteet, kuten aineiston omistus- ja käyttöoikeuksiin tai lainsäädäntöön liittyvät rajoitteet. Tutkimusdataa avatessa on myös varmistettava henkilötietojen suoja ja tämän takia aineisto pyritään yleensä anonymisoimaan. (Sarpawari, Kesselheim, Malin, Gagne, & Schneeweiss, 2014; Strasser, 2015.) Toisinaan ratkaisujen tekeminen yksityisyyden ja avoimen jakamisen välillä koetaan ongelmalliseksi (Strasser, 2015).

Tutkimusaineistojen ja -datan jakaminen muiden kanssa ei kuitenkaan ole uusi asia. Jakamista voi tapahtua saman tai eri instituutioiden tutkijakollegoiden kesken. Nykyään tutkimusdatan pitkäaikaistallentamista varten on tarjolla yhä enemmän avoimuuteen kannustavia aineiston tallennuspalveluita. Datat liittäminen tieteelliseen julkaisuun täydentävän materiaalin muodossa katsotaan myös aineiston jakamiseksi. (Borgman, 2012.) Osa kansainvälisistä tieteellisistä lehdistä onkin alkanut vaatia niin sanotun raakadatan ja tutkimusprosessin dokumentoinnin liittämistä lehtiin ehdolle lähetettäviin käsikirjoituksiin. Osa edellyttää tutkimusdatan tallentamista avoimiin tutkimusaineiston tallennuspalveluihin. (Tripathi, Shukla, & Sonkar, 2017.)

Aineistohallintasuunnitelma (data management plan, DMP) on dokumentti, jossa kuvataan hankkeen aikainen tutkimusaineiston hallinta ja hankkeen päättymisen jälkeiset toimet liittyen aineiston avaamiseen. Tänä päivänä useat yliopistot tarjoavat koulutusta ja neuvontaa liittyen aineistohallintasuunnitelman tekoon. Opastajana toimivat yleensä yliopistojen kirjastot. Myös monet rahoittajat ovat todenneet, että tutkimuksen suunnittelu aineistohallintasuunnitelman muodossa on hyödyllistä ja ne ovat alkaneet vaatia suunnitelmia rahoitushakemusten liitteeksi (Strasser, 2015, s. 2).

Tutkimusdatan säilytykseen liittyviä asioita ovat muun muassa varmuuskopioiden tallentaminen, versioiden hallinta, formaatin ja mahdollisen pitkäaikais säilytyspalvelun valinta. Digitaaliselle aineiston tulisi saada yksilöllinen ja pysyvä tunniste (esim. Digital Object Identifier, DOI tai Unified Resource Name, URN), jonka avulla aineistoon voidaan myös viitata (Mooney & Newton, 2012; Strasser, 2015). Web of Science-palvelussa ylläpidetään Clarivate Analyticsin aineistoviittauksista viitetietokantaa (Data Citation Index, DCI), joka mahdollistaa muun muassa tutkimusaineistojen ja julkaisujen yhteyksien tarkastelun (Clarivate Analytics, 2018).

Metadatan avulla tutkimusdatasta tulee löydettävää, ymmärrettävää ja käytökelpoista. Kontrolloidun sanaston ja standardien käyttö lisää löydettävyyttä ja hyödynnettävyyttä. (Strasser, 2015; Wilkinson ym., 2016.) Data-arkistoon tallentaminen takaa yleensä monien koneluettavuuden kannalta keskeisten asioiden, kuten formaatin, lisenssin ja tunnisteiden, huomioimisen.

## Aineiston jakamisen hyödyt ja esteet

Aineiston jakaminen voi lisätä näkyvyyttä, tutkimusyhteistyötä sekä poikia uusia projekteja. Eli samalla kun tutkimusdata on turvallisesti säilötty, avoimuus voi auttaa tutkijaa myös meritoitumisessa. Tutkimusdatan avoimuus lisää myös tutkimuksen läpinäkyvyyttä ja luotettavuutta. Se antaa muille tutkijoille mahdollisuuden toistaa tehdyt analyysit tai hyödyntää aineistoa sellaisiin tutkimuskysymyksiin vastaten, jotka eivät ole olleet aineiston keränneiden tutkijoiden intresseissä. Dataa voidaan hyödyntää myös esimerkiksi meta-analyyseissä ja opetusmateriaaleina. (Feldman, Patel, Ortmann, Robinson, & Popovic, 2012.) Tutkimusdatan jatkokäytön mahdollistamisesta puhuttaessa on korostettu myös yliopistojen mahdollisuuksia lisätä sen kautta vaikuttavuuttaan (McKiernan, 2017).

Tutkimusdatan avoimen saatavuuden hyötyjä on tutkittu erityisesti koskien aloja, joilla avoimuus on ollut jo pitempään yleistä. Esimerkiksi genomitutkimuksen tuottaman datan avoimeksi saattamisen on todettu lisäävän viittausten määrää (Piwowar & Vision, 2013). Astrofysiikan alalla on havaittu sama ilmiö (Dorch, Drachen, & Ellegaard, 2015; Drachen, Ellegaard, Larsen, & Dorch, 2016). Tutkimusdatan tallentamisen avoimiin tietoarkistoihin on havaittu myös olevan yhteydessä tutkimushankkeen lisääntyneeseen julkaisumäärään (Pienta, Alter, & Lyle, 2010). On kuitenkin esitetty, että laaja-alaisia tutkimuksia jakamisaktiivisuuteen ja hyötyihin liittyen kaivataan (Mongeon, Robinson-Garcia, Jeng, & Costas, 2017).

Tutkimuksessa on havaittu, että vaikka suurin osa tutkijoista kannattaa

ajatusta tutkimusdatan saattamisesta avoimeksi, harva kuitenkaan toimii näin (Scaramozzino, Ramírez, & McGaughey, 2012). Aineiston jakamiselle voidaan nähdä myös esteitä. Esteiden on tutkimuksissa esitetty liittyvän teknisiin tai organisatorisiin rakenteisiin, aineiston säilytyksen tukemiseen (e.g. Thessen & Patterson, 2011; Shearer & Furtado, 2017) tai eettisiin haasteisiin (e.g. ANDS, 2018). Myös datan avoimuuteen liittyvät käytännöt, ohjeistukset ja aktiivisuudesta palkitseminen voivat olla puutteellisia (Rowhani-Farid, Allen, & Barnett, 2017; Thessen & Patterson, 2011).

Tutkijoilla voi olla datan väärinkäyttöön tai väärinymmärrykseen liittyviä pelkoja. Lisäksi tiedostetaan, että datan ja metadatan saattaminen tallennuspalvelun edellyttämään muotoon voi vaatia paljon aikaa ja resursseja. (Koltay, 2017.) Toisaalta datan jakamisen ei koeta (vielä toistaiseksi) lisäävän tutkijan meritoitumista. Ratkaisuksi tähän on esitetty muun muassa niin sanottua jakamisindeksiä, S-indeksiä, tutkijan tuloksellisuuden ja julkaisutehokkuuden kuvaamiseen käytetyn H-indeksin tapaan (Olfson, Wall, & Blanco, 2017).

Kyselytutkimuksessaan Van den Eynden kollegoineen (2016) kartoittivat tutkimusdatan avaamisen syitä. Tutkijoiden kertomat pääasialliset syyt liittyivät rahoittajien vaatimuksiin sekä käsitykseen siitä, että aineiston jakaminen on hyvä käytäntö, joka voi lisätä yhteistyötä muiden tutkijoiden kanssa. Toisaalta Helsingin yliopistolla suoritetuissa tutkijahaastatteluissa keskeiseksi esteeksi ja kompastuskiveksi pitkäaikaissäilytykselle nousi metatietojen puutteellisuus. Myöskään metatietojen tuottamiseen tai jo olemassa olevien tietojen parantamiseen ei nähty olevan resursseja. (Ala-Kyyny, Korhonen, & Roinila, 2017.)

Myös demografisten tekijöiden yhteyttä halukkuuteen jakaa aineistoa on tutkittu. Linekin ja hänen kollegoidensa (2017) tutkimuksen perusteella naiset olivat miehiä haluttomampia jakamaan tutkimusdataansa. Heidän asenteensa olivat negatiivisempia ja heillä oli myös vähemmän kokemusta datan jakamisesta. Pelot vaivasivat naisia miehiä enemmän ja he kaipasivat myös lisää kannustimia.

Tietoarkisto toteutti toukokuussa 2015 Avoin tiede ja tutkimusaineistot-kyselyn kartoittaakseen humanististen sekä terveys- ja lääketieteen alojen tutkijoiden mielipiteitä sähköisten tutkimusaineistojen käsittelyyn ja jatkoehdyntämiseen liittyen. Selvityksen mukaan humanistit tunsivat suositukset tutkimusaineistojen avaamisesta tiedeyhteisön käyttöön terveystieteilijöitä paremmin. Vaikkakin vastaajat olivat todennäköisesti aiheesta kiinnostuneita, heistä vain puolet tiesi oliko heidän organisaatiossaan ohjeita tutkimusaineistojen säilyttämisestä tai tutkimusaineistojen avaamisesta muuhun kuin tutkijan/tutkimusryhmän omaan käyttöön. Tutkimusaineistojen jatkokäytön mahdollisen hyödyt nähtiin merkittävinä. Terveystieteilijät näkivät jatkokäyttöön liittyvinä pääasiallisina riskeinä tietosuoja- ja tutkimuseettisten

riskien lisääntymisen, tutkittavien informoinnin puutteen ja jatkokäyttäjien taitamattomuuden. Humanisteja huolestutti se, että tutkittavia ei ole informoitu aineistojen arkistoinnista, ja se että tutkijoiden työaika kuluu aineistojen muokkaamiseen. Tutkimusaineistojen avaamiseen parhaiten motivoivana tekijänä nähtiin oman tutkimuksen saama lisänäkyvyys. (Fält & Okuloff, 2016, ss. 6–15.)

## **Avoimen datan toimijat ja esitetyt linjaukset**

Avoimeen tieteen ja avoimen tutkimusdatan kenttään kytkeytyy monia kansainvälisiä ja kansallisia toimijoita. Lisäksi aiheeseen liittyen on esitetty useita linjauksia ja julistuksia. Esimerkkinä toimijoista voidaan esille nostaa Research Data Alliance (RDA), joka on kansainvälinen, yhteisövetoinen organisaatio. RDA:n päämääränä on edistää tieteellisen datan avoimuutta ja jaettavuutta maailmanlaajuisesti. (Research Data Alliance, 2018.) Open Research Funders Group (ORFG) sen sijaan tuo yhteen eri maiden tutkimusrahoittajia, joita yhdistää omistautuminen tutkimuksen avoimeen jakamiseen (Open Research Funders Group 2018).

Vuonna 2012 esitelty julkilausuma San Francisco Declaration on Research Assessment (DORA) (2018) painottaa sitä, että tutkimuksen arvioinnissa tulisi ottaa huomioon kaikkien tutkimukseen liittyvien tuotosten, kuten datan, arvo ja vaikutus. Joint Declaration of Data Citation Principles (2014) korostaa sitä, että tutkimuksen arvokkaana tuotteena myös dataan tulisi voida artikkelien tapaan viitata. Vuonna 2014 alkunsa sai myös vahvasti esille noussut FAIR-periaatelistaus, joka tarjoaa ohjenuoria digitaalisen aineiston löydettävyyden, saavutettavuuden, yhteentoimivuuden ja uudelleenkäytettävyyden takaamiseksi (FAIR: Findable, Accessible, Interoperable, Re-usable) (European Commission, 2016; Wilkinson ym., 2016).

Euroopan Komission ROARMAP ja PASTEUR4oA projektit ovat kartoittaneet eri maiden linjauksia liittyen avoimeen tieteseen (Swan, Gargouri, Hunt, & Harnad, 2015). League of European Research Universities (LERU) sen sijaan on julkaissut ohjeistuksen siitä, millainen rooli avoimella tietellä tulisi olla yliopistoissa (League of European Research Universities, 2018). Tutkimusdatan avoimuuden tulisi olla standardi, erityisesti julkisesti rahoitettujen tutkimusten kohdalla. Raportti nostaa esille, että suurimmat haasteet avoimen tieteen käytäntöihin siirtymisessä eivät liity pelkästään tekijänoikeuksiin ja maksuihin, vaan erityisen haastavaa on muutoksen aikaansaanti tutkimuskulttuurissa.

Suomessa opetus- ja kulttuuriministeriön Avoin tiede ja tutkimus (ATT) -hankkeessa (Avoin tiede ja tutkimus -hanke, 2018a) on vuosina 2014–2017 tehty

monitoimijaista ja -alaista yhteistyötä, jonka myötä Suomi on noussut kansainvälisesti tunnustetuksi avoimen tieteen edelläkävijäksi. Suomen yliopistojen rehtorineuvosto UNIFI:n johdolla on työstetty avoimen tieteen ja datan kansallinen toimenpideohjelma, joka julkistettiin toukokuussa 2018 (Suomen yliopistojen rehtorineuvosto UNIFI ry 2018). Opetus- ja kulttuuriministeriö tuki myös tätä työryhmätyöskentelyä. Toimenpideohjelman mukaisena päätavoitteena on pitää Suomi avoimen tieteen suunnannäyttäjänä ja saada avoin tiede Suomessa osaksi tieteen arkea. Toimenpideohjelman myötä avoimen tieteen kansallinen koordinaatio on jaettu kolmeen teemaan: FAIR-tutkimusdataan, avoimeen julkaisemiseen sekä avoimuuden kulttuuriin. Toimenpideohjelmassa korostetaan myös sitä, että ollakseen laadukkaita tutkimusdatan säilyttämiseen liittyvät palvelut vaativat paljon resursseja. Tieteellisten seurain valtuuskunta on ottanut vastuun koordinaatiovastuusta. Opetus- ja kulttuuriministeriön rahoittamana Suomessa on esitetty myös ”tiekartta” joka on ohjeistus tutkimusdataan liittyvistä viittaamiskäytännöistä (Finnish Committee for Research Data, 2018; Laine & Nykyri, 2018).

Suomessa ollaan myös havahduttu aineistonhallintasuunnitelmien ohjeistetun tekemisen tärkeyteen. Osana ATT-hanketta Helsingin yliopiston kirjaston koordinoima Tuuli-projekti keskittyi kansallisen DMPTuuli-aineistonhallintatyökalun kehittämiseen ja käyttöönottoon. Työkalu on kehitetty DCC:n eli Digital Curation Centren (UK) avoimen lähdekoodin DMPonline-työkalun pohjalta. (Tuuli-projekti, 2017.) Mukana projektissa oli noin neljäkymmentä edustajaa noin kahdestakymmenestä suomalaisesta yliopistosta ja tutkimusorganisaatiosta.

Julkaisujen avoimeen saatavuuteen kannustamisen lisäksi Suomen Akatemia edellyttää, että rahoitusta saaneet hankkeet avaavat tutkimusaineistonsa ja -menetelmänsä kansallisissa tai kansainvälisissä data-arkistoissa. Tampereen yliopiston yhteydessä toimii Tietoarkisto (2018), joka arkistoi ja välittää sähköisiä tutkimusaineistoja. ATT-hanke ylläpitää IDA-, Etsin- ja AVAA-palveluja. IDA-säilytyspalvelu mahdollistaa tutkimusdatan turvallisen säilytyksen ja Etsin on kehitetty metatietojen avointa jakamista varten. AVAA on avointen tutkimusaineistojen julkaisualusta. (Avoin tiede ja tutkimus -hanke, 2018b) Kansainvälisiä data-arkistoja ovat esimerkiksi Zenodo, EUDAT, Dryad, Figshare ja GitHub.

## **Aineiston keruu ja analyysi**

Tämän tutkimuksen päämääränä on lisätä tietoutta suomalaisten tutkimusorganisaatioiden jäsenten käsityksistä, kokemuksista ja mielipiteistä liittyen tutkimusaineiston jakamiseen ja pitkäaikaissäilytykseen. Huomio on erityisesti

tutkimusdatassa ja sen avoimessa jakamisessa. Aiheen selvittämistä varten toteutettiin kyselytutkimus. Tutkimuskysymykset ovat:

1. Millaisia käsityksiä, kokemuksia ja mielipiteitä tutkimusorganisaatioissa toimivilla on tutkimusaineistojen jakamiseen ja pitkäaikaissäilytykseen liittyen?
2. Onko miesten ja naisten käsityksissä, kokemuksissa ja mielipiteissä eroja?
3. Onko jatko-opiskelijoiden ja muissa työrooleissa olevien välillä havaittavissa eroja?
4. Onko kyselyn vastaajien edustamien tieteenalojen välillä eroja?

Artikkelin kyselytutkimus oli osa laajaa kansainvälistä Data Literacy and Research Data Management -yhteistyöprojektia. Projektissa sama kysely toteutettiin samanlaisena noin 24 eri maassa. Kyselyn suunnittelivat tutkijat Iso-Britanniasta, Turkista ja Ranskasta. Suomalaista kohdeyleisöä varten kysely käännettiin kielenkääntäjällä suomeksi ja ruotsiksi. Lopuksi käännökset vielä katsottiin läpi Oulun yliopiston tutkijoiden toimesta.

Kysely toteutettiin kaikissa maissa LimeSurvey-kyselypalvelua hyödyntäen. Vastaaja pystyi valitsemaan vastauskielekseen suomen, ruotsin tai englannin. Kysely sisälsi 24 kysymystä koskien vastaajan taustatietoja, tutkimuksessa tuotettuja ja käytettyjä aineistoja, metadatatietoja, aineistonhallintasuunnitelmaa sekä tutkimusaineiston jakamista ja pitkäaikaissäilytystä. Monissa kysymyksissä vastaajalla oli mahdollisuus valita useita vastausvaihtoehtoja. Osassa vastausvaihtoehtoon ”muu” pystyi antamaan tarkennuksen avovastauksen muodossa. Tässä artikkelissa keskitymme kyselyn sisältämiin tutkimusaineiston avoimeen jakamiseen ja pitkäaikaissäilytykseen liittyviin kysymyksiin ja näiden kysymysten suhteeseen vastaajan sukupuoleen, rooliin tiedeyhteisössä ja tieteenalaan.

Suomessa kysely kohdistettiin erilaisissa Suomen tutkimusorganisaatioissa tutkimuksen parissa toimiville henkilöille, eli niin henkilökunnalle kuin tohtorikoulutettavillekin. Tavoitteena oli siis saada vastauksia niin yliopistoista kuin myös muista tutkimusorganisaatioista. Tästä johtuen kyselyn levittämisessä päätettiin hyödyntää Tuuli-projektin (Tuuli 2016) verkostoa. Projektin edustajien katsottiin tarjoavan kattavan yhteyden erilaisiin tutkimusorganisaatioihin. Heitä lähestyttiin sähköpostitse pyynnöllä jakaa kyselyn saateviestiä ja linkkiä valitsemiensa reittien kautta organisaatioissaan. Otannan takia vastausaktiivisuutta kussakin organisaatiossa oli vaikea ennakoida. Aineistonkeruu toteutettiin kesä-heinäkuussa vuonna 2017. Vastaajista 60,5 prosenttia (441) valitsi vastauskieleksi suomen, 18,0 prosenttia (131) ruotsin ja 21,5 prosenttia (157) englannin. Vastaajista 612 oli vastannut vähintään yhteen kysymykseen

ja 469 kaikkiin 24:n kysymykseen. Analyysiä ei rajoitettu pelkästään kaikkiin kysymyksiin vastanneisiin.

Aineisto analysoitiin käyttäen IBM SPSS Statistics 24 tilastohjelmaa. Ristiintaulukoinnin avulla voi tarkastella kahden luokitteluasteikollisen muuttujan välistä riippuvuutta. Riippuvuutta voi testata Pearsonin khii toiseen tai Fisherin testillä. Tässä artikkelissa ryhmien välisen eron tilastollinen merkitsevyys testattiin Pearsonin khii toiseen testillä. Fisher's exact testiä käytettiin kun parametrisen testin käytön ehdot eivät täyttyneet. Tilastollisen merkitsevyyden rajana pidettiin yleisesti hyväksyttyä  $p < 0,05$ .

## Tulokset

Vastaajista 38,3 prosenttia (238) oli miehiä, 57,6 prosenttia (358) naisia ja 4,1 prosenttia (25) valitsi vaihtoehdon "muu" tai "en tahdo kertoa". Vastaajista 3,5 prosenttia (22) oli 18-25 vuotiaita, 40,3 prosenttia (250) 26-35 ja 23,5 prosenttia (146) 36-45 vuotiaita. 56-65 vuotiaita oli 11,8 prosenttia (73) ja "65+"-vuotiaisiin lukeutui 2,6 prosenttia (16) vastaajista. Kuusi vastaajaa (1,0 %) oli valinnut vaihtoehdon "en halua kertoa".

Kysyttäessä vastaajan tämänhetkistä pääasiallista roolia tutkimuksen parissa 56,9 prosenttia (348) valitsi vaihtoehdon "jatko-opiskelija", 36,5 prosenttia (222) määritteli olevansa "tutkija tai muu tutkimuslaitoksen (korkeakoulun) työntekijä" ja 6,9 prosenttia (42) näki kuuluvansa kolmanteen annettuun luokkaan eli "muu". Muutama oli vastannut "muu"-kohtaan jatko-opiskelija ja heidät luokiteltiin analyysissä mukaan "jatko-opiskelija"-luokkaan. Loput (jotka olivat määritelleet roolinsa avoimessa jatkokysymyksessä muun muassa kliinikoksi, sivutoiminen tutkijaksi, tilastotieteen asiantuntijaksi, opettajaksi, koordinaattoriksi, tutkimusavustajaksi, dosentiksi, professoriksi tai emeritukseksi) luokiteltiin analyysissä mukaan tutkija tai muu tutkimuslaitoksen työntekijä -luokkaan.

Kyselyssä tiedusteltiin myös sitä, kuinka pitkään vastaaja on tehnyt tutkimusta. Vastaajista 42,3 prosenttia (262) vastasi tehneensä tutkimusta alle 5 vuotta, 23,7 prosenttia (147) 5-10 vuotta, 9,9 prosenttia (61) 11-15 vuotta ja 22,8 prosenttia (141) yli 16 vuotta. Vastaajista kahdeksan (1,1 %) valitsi vaihtoehdon "en ole tehnyt tutkimustyötä".

Vastaajan tieteenalaa kysyttäessä vastaajalle annettiin 37 vastausvaihtoa. Näistä esimerkkeinä "insinööritieteet ja teknologia: sähkötekniikka, elektroniikka, tietotekniikka" ja "yhteiskuntatieteet: valtio-oppi". Vastausvaihtoehdoissa oli kuusi pääluokkaa eli esimerkiksi "insinööritieteet ja teknologia" sekä "yhteiskuntatieteet". Analyysissä päädyttiin luokittelemaan vastaukset niiden pääluokkien mukaisesti. Pääluokan "maataloustieteet"



(agricultural science) valitsi 2 vastaajaa joten analyysia varten se päätettiin yhdistää pääluokan “luonnontieteet” yhteyteen. Tilastollista analyysiä varten tieteenalaluokkia oli siis viisi: luonnontieteet ja maataloustieteet (20,1 %, 111), insinööritieteet ja teknologia (11,4 %, 63), lääke- ja terveystieteet (26,0 %, 144), yhteiskuntatieteet (29,5 %, 163) ja ihmistieteet (eli humanistiset alat, 13,0 %, 72).

Kyselyssä esitettiin avokysymys “Tutkimusinstituutio, jossa toimit”. Avovastaukseen ei ollut pakko vastata. Vastaukset luokiteltiin ja luokat vastausmäärineen on esitetty Taulukossa 1.

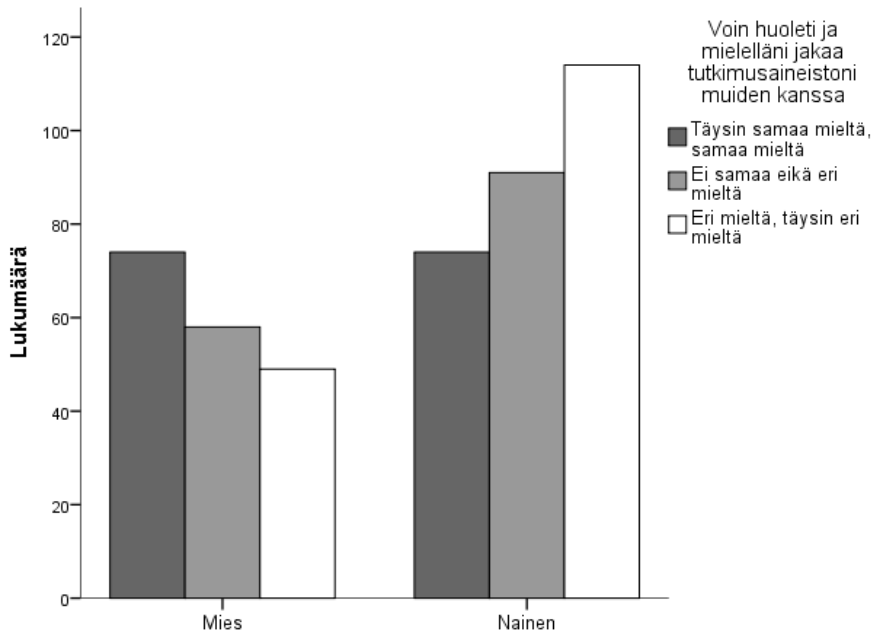
Organisaatio	%	lkm
Turun yliopisto	21,5	157
Oulun yliopisto	18,8	137
Åbo Akademi	15,6	114
Helsingin yliopisto	7,4	54
Tampereen yliopisto	2,9	21
Muu suomalainen yliopisto (Itä-Suomen yliopisto, Aalto-yliopisto, Vaasan yliopisto, Jyväskylän yliopisto, Lapin yliopisto)	2,2	16
Muut tutkimuslaitokset (Suomen ympäristökeskus SYKE, Teknologian tutkimuskeskus VTT, Terveysten ja hyvinvoinnin laitos THL, Syöpäjärjestöt, Työterveyslaitos ym.)	4,0	29
Ulkomainen yliopisto	0,7	5
Määrittelemätön (esim. “yliopisto” tai oppiaineen nimi ilman tietoa yliopistosta)	10,4	76
Tyhjä vastaus	15,9	116
Vastasi useamman organisaation	0,3	3

Taulukko 1: Vastaajien ilmoittamat tutkimusorganisaatiot avovastauksista luokiteltuina.

## Tutkimusaineiston jakaminen ja datan avoimuus

Tämän tutkimuksen fokuksessa ovat tutkimusaineiston jakamiseen liittyvät kysymykset. Nämä kyselyn kysymykset koskivat käsityksiä, kokemuksia ja mielipiteitä jakamiseen ja avoimuuteen liittyen. Kysyttäessä “suosittelee tutkimuslaitoksesi aktiivisesti aineistojen jakamista vapaan pääsyn (open access) periaatteella” vastaukset jakautuivat tasaisesti. Vastaajista 35,8 prosenttia (170) koki, että heidän tutkimusorganisaationsa kannustaa heitä jakamaan aineistoa avoimesti eli valitsi vastausvaihtoehdon “kyllä”. Vastaajista 32,6 prosenttia (155) valitsi vaihtoehdon “ei” ja lähes sama määrä eli 31,6 prosenttia (150) vaihtoehdon “en osaa sanoa”.

Vastauksissa ei ollut tilastollisesti merkitsevää eroa sukupuolten välillä. Sen sijaan rooli tiedeyhteisössä oli yhteydessä siihen, kokiko vastaaja että hänen tutkimuslaitoksensa suosittelee aktiivisesti aineiston avointa jakamista (Pearson  $p = 0,002$ ). Vastausvaihtoehdon “kyllä” oli tohtorikoulutettavista valinnut 29,2 prosenttia (78/267) ja muista (tutkija, henkilökuntaa, muu) 45,1 prosenttia

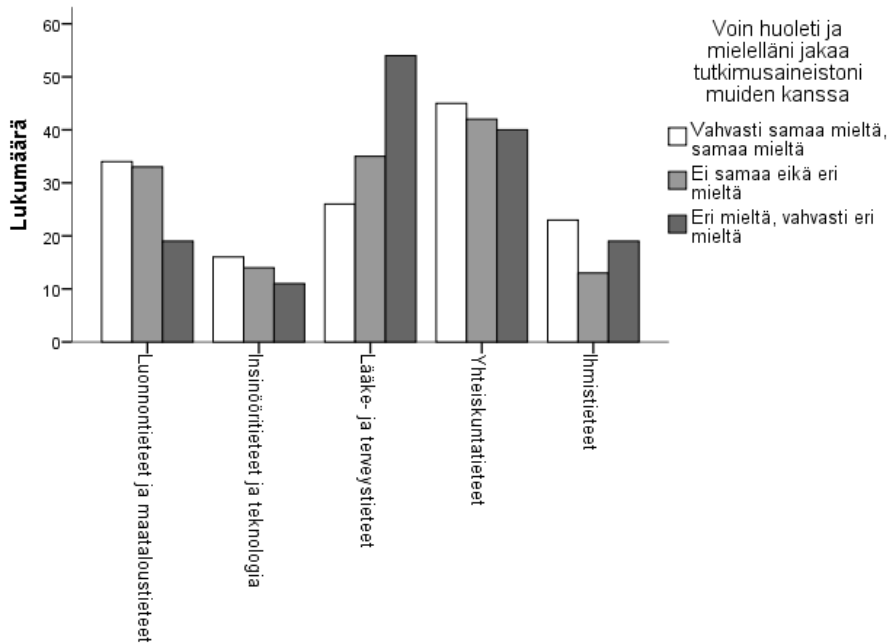


Kuvio 1: Sukupuolten välinen ero väitteen “voin huoletti ja mielelläni jakaa tutkimusaineistoni muiden kanssa” kohdalla. Kuvioissa yhdistetty luokat “vahvasti samaa mieltä” ja “samaa mieltä” sekä “eri mieltä” ja “vahvasti eri mieltä”.

(92/204). Jatko-opiskelijoiden keskuudessa “ei” oli vastauksista suosituin (36,7 %, 98/267). Tarkasteltaessa eri tieteenaloja erot eivät olleet tilastollisesti merkitseviä, vaikkakin lääke- ja terveystiedon alan edustajat erottuivat muista valiten muita vaihtoehtoja todennäköisemmin vastaukseen “ei”.

Väitteen “voin huoletti ja mielelläni jakaa tutkimusaineistoni muiden kanssa” kohdalla 8,9 prosenttia (42) vastasi olevansa “vahvasti samaa mieltä”, 24,2 prosenttia (114) “samaa mieltä”, 31,8 prosenttia (150) “ei samaa eikä eri mieltä”, 22,9 prosenttia (108) “eri mieltä” ja 12,3 prosenttia (58) “vahvasti eri mieltä”. Naiset vastasivat miehiä todennäköisemmin “ei samaa eikä eri mieltä” tai olivat eri mieltä väitteen kanssa (Pearson  $p = 0,006$ ) (Katso Kuvio 1.)

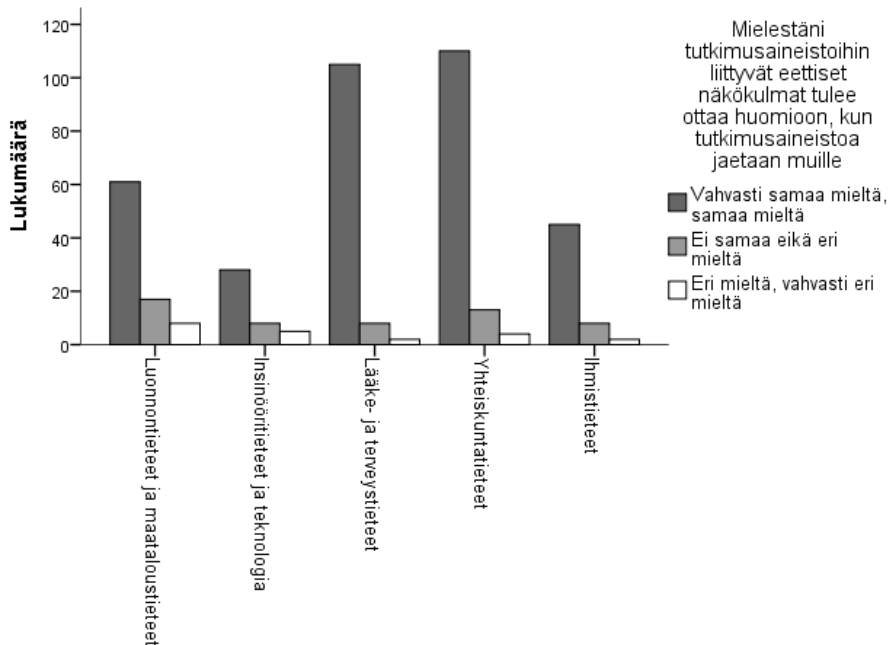
Jatko-opiskelijoiden ja muussa työroolissa toimivien välillä ei ollut merkittäviä eroja. Tieteenalojen välinen ero katsotaan tilastollisesti merkitseväksi (Fisher’s exact  $p = 0,048$ ) ja lääke- ja terveystieteen edustajat erottuivat muista ollen pääosin väitteen kanssa eri mieltä (47,0 %, 54/115) kuten Kuvioista 2 näkyy.



Kuvio 2: Tieteenalojen välinen ero väitteen “voin huolehti ja mielelläni jakaa tutkimusaineistoni muiden kanssa” vastausten kohdalla. Kuvioissa yhdistetty luokat “vahvasti samaa mieltä” ja “samaa mieltä” sekä “eri mieltä” ja “vahvasti eri mieltä”.

“En näe ongelmaa tutkimusaineistojeni jakamisessa” väitteen kanssa vastaajista 7,8 prosenttia (37) koki olevansa “vahvasti samaa mieltä”, 21,4 prosenttia (101) “samaa mieltä”, 26,5 prosenttia (125) “ei samaa eikä eri mieltä”, 32,6 prosenttia (154) “eri mieltä” ja 11,7 prosenttia (55) “vahvasti eri mieltä”. Naiset näkivät enemmän ongelmia tutkimusaineistojen jakamiseen liittyen, sillä heistä 50,9 prosenttia (142/279) kertoi olevansa väitteen kanssa eri mieltä, kun miehistä vastaava määrä oli 35,9 prosenttia (65/181) (Pearson  $p = 0,036$ ). Työrooliin tai tieteenaloihin liittyvät erot eivät olleet tilastollisesti merkitseviä.

Kysyttäessä mielipidettä väitteeseen “mielestäni tutkimusaineistoihin liittyvät eettiset näkökulmat tulee ottaa huomioon, kun tutkimusaineistoa jaetaan muille” 51,1 prosenttia (241) oli väitteen kanssa täysin samaa mieltä, 31,6 prosenttia (149) samaa mieltä, 11,9 prosenttia (56) ei samaa eikä eri mieltä, 3,2 prosenttia (15) eri mieltä ja 2,3 prosenttia (11) kertoi olevansa täysin eri mieltä. Naisista 85,7 prosenttia (239/279) oli väitteen kanssa samaa mieltä, kun osuus

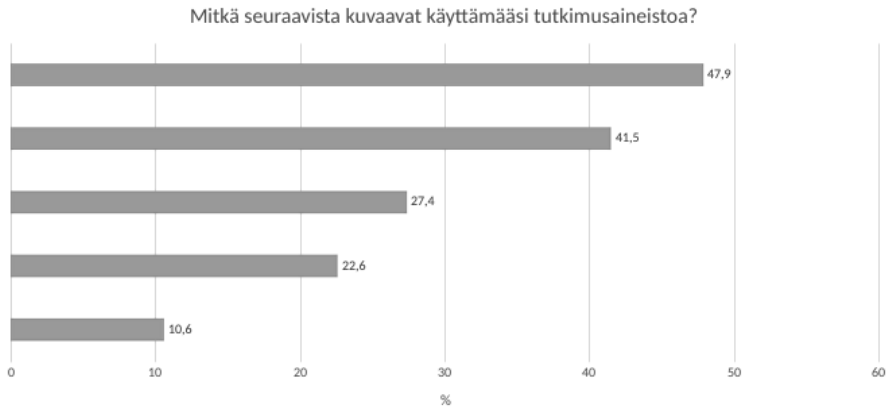


Kuvio 3: Tieteenalojen välinen ero väitteen “mielestäni tutkimusaineistoihin liittyvät eettiset näkökulmat tulee ottaa huomioon, kun tutkimusaineistoa jaetaan muille” vastausten kohdalla. Kuvioissa yhdistetty luokat “vahvasti samaa mieltä” ja ”samaa mieltä” sekä “eri mieltä” ja “vahvasti eri mieltä”.

miehistä oli 79,6 prosenttia (144/181). Ero oli tilastollisesti merkitsevä (Pearson  $p < 0,001$ ).

Jatko-opiskelijoista samaa mieltä väitteen kanssa oli 79,2 prosenttia (209/264) ja tutkimusorganisaatioissa muissa työrooleissa toimivista 86,7 prosenttia (177/204). Ero on tilastollisesti merkitsevä (Fisher’s exact  $p = 0,036$ ). Myös tutkimusalojen välillä oli eroja (Fisher’s exact  $p < 0,001$ ). Erityisesti lääke- ja terveystieteen (91,3 %, 105/115) sekä yhteiskuntatieteiden edustajat (86,6 %, 110/127) kokivat, että eettisten näkökulmien huomioiminen tutkimusdataa jaettaessa on keskeistä (Katso Kuvio 3.)

Kyselyssä esitettiin vastausvaihtoehtoja liittyen kysymykseen “mitkä seuraavista kuvaavat käyttämäsi tutkimusaineistoa?”. Vastaajia pyydettiin valitsemaan vaihtoehtoista kaikki heidän aineistoaan koskevat eli vastaaja pystyi valitsemaan useamman vaihtoehdon. Suosituimman vastausvaihtoehdon “ai-

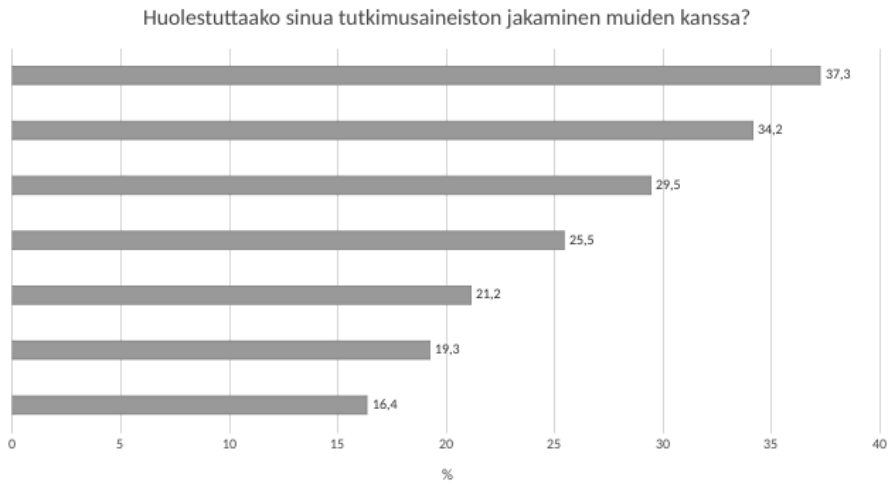


Kuvio 4: Vastausvaihtoehtojen valinnan jakautuminen kysyttäessä “Mitkä seuraavista kuvaavat käyttämäsi tutkimusaineistoa?” Vastaaaja pystyi valitsemaan useampia vaihtoehtoja.

neisto on käytettävissä pyynnöstä” valitsi 231 (47,9 %). Kuvio 4 havainnollistaa valittuja vastauksia.

Miesten ja naisten välillä ei ollut tilastollisesti merkitsevää eroa. Osan vaihtoehtoväittämien kohdalla ilmeni selkeä ero jatko-opiskelijoiden ja muiden tutkimusorganisaatioiden työntekijöiden välillä. Jälkimmäiset valitsivat selkeästi useammin vastaukset “aineistoni on vapaasti tutkimusryhmäni käytettävissä” (Pearson  $p < 0,001$ ) sillä tutkijoista ja muusta tutkimusorganisaatioiden työntekijöistä 49,0 prosenttia (102/208) katsoi vaihtoehdon kuvaavan heidän käyttämänsä tutkimusaineistoa, kun sen sijaan jatko-opiskelijoista tätä mieltä oli vain 34,1 prosenttia (92/270). Vastaavat prosenttiosuudet “aineisto on käytettävissä pyynnöstä” (Pearson  $p = 0,003$ ) väittämän kohdalla olivat 55,8 prosenttia (116/208) ja 41,9 prosenttia (113/270). Myös vaihtoehto “aineistoon on vain rajattu pääsy (esim. vain osa aineistosta on yleisesti saatavilla)” sai enemmän kannatusta muiden kuin jatko-opiskelijoiden toimesta: 36,1 prosenttia (75/208) verrattuna jatko-opiskelijoihin, joista sen valitsi 20,7 prosenttia (56/270) (Pearson  $p < 0,001$ ). Sen sijaan “aineistooni ei ole pääsyä muilla kuin minulla” oli selkeästi suosituimpi valinta jatko-opiskelijoiden (29,3 %, 79/270) keskuudessa verrattuna muihin (13,5 %, 28/208) (Pearson  $p < 0,001$ ).

Myös eri tieteenalojen välillä oli näkyvissä eroja osassa valinnoista. “Aineistoni on vapaasti kaikkien käytettävissä” vaihtoehdon valinneet kuuluivat pääasiassa luonnontieteilijöihin, yhteiskuntatieteilijöihin tai ihmistieteilijöihin (Fisher’s exact  $p < 0,001$ ). Humanististen alojen (ihmistieteet) edustajat olivat



Kuvio 5: Vastausvaihtoehtojen valinnan jakautuminen kysyttäessä “Huolestuttaako sinua tutkimusaineiston jakaminen muiden kanssa?” Vastaaaja pystyi valitsemaan useampia vaihtoehtoja.

valinneet muita vähemmässä määrin vaihtoehdon “Aineistoni on vapaasti vain tutkimusryhmäni käytettävissä” (Fisher’s exact  $p = 0,007$ ). Heistä tämän vaihtoehdon valinnoita oli 21,4 prosenttia (12/56). Toisaalta vaihtoehdon “aineistooni ei ole pääsyä muilla kuin minulla” oli valittu muita useammin yhteiskuntatieteilijöiden toimesta (31,3 %, 40/128) (Pearson  $p = 0,002$ ).

Kysely sisälsi kysymyksen liittyen siihen huolestuttaako vastaajaa tutkimusaineiston jakaminen muiden kanssa ja vaihtoehdoissa tarjottiin syitä mahdollisille huolille. Vastaaaja pystyi valitsemaan useamman esitetyistä vaihtoehdoista. Suosituimmaksi vaihtoehdoista osoittautui “Laillisuus ja eettisyys huolestuttaa”, jonka valitsi 180 (37,3 %) vastaajaa. Vastaaajien valinnat on esitetty Kuviossa 5.

Kysymykseen liittyi myös “muu” vaihtoehto, johon liittyen vastaaja pystyi kirjoittamaan vastauksensa. Avovastauksissa esille nostettiin se, että aineiston tuhoutuminen huolestuttaa, eettinen lupa ei kata avointa aineiston jakamista ja että data on arkaluontoista tai luottamuksellista. Osa vastaajista koki, että he tahtovat pitää aineiston vain omassa käytössään kunnes he ovat tehneet siitä kaikki tahtomansa analyysit ja julkaisseet siihen perustuvat artikkelit. Yksi vastaajista nosti myös esille, että työnantajan ohjeet rajaavat aineiston jakamista (salassapitovelvollisuus).

Miesten ja naisten välillä ei ollut vastausvaihtoehtojen valinnoissa

tilastollisesti merkitsevää eroa. Jatko-opiskelijoiden ja muiden väliset merkitsevät erot koskivat väittämiä “aineiston väärinkäyttö huolestuttaa” (Pearson  $p = 0,035$ ) ja “aineiston sisältämän tiedon väärä tulkinta huolestuttaa” (Pearson  $p = 0,012$ ). Molemmissa tapauksissa jatko-opiskelijoista muita harvempi oli valinnut nämä vaihtoehdot. Väärinkäyttö huolestutti 21,9 prosenttia (59/270) jatko-opiskelijoista, luvun ollessa muiden kohdalla 30,8 prosenttia (64/208). Datan sisältämän tiedon väärä tulkinta huolestutti 24,8 prosenttia (67/270) jatko-opiskelijoista ja muista 35,6 prosenttia (74/208).

Vastaajista pelko tieteellisen etumatkan menettämisestä huolestutti erityisesti luonnontieteilijöitä, joista 35,2 prosenttia (31/88) (Pearson  $p < 0,001$ ) valitsi tämän vaihtoehdon. Toisaalta laillisuuteen ja eettisyyteen liittyvät kysymykset huolestuttavat muita enemmän insinööritieteiden ja teknologioiden alojen edustajia, yhteiskuntatieteilijöitä ja ihmistieteilijöitä (Pearson  $p = 0,043$ ).

## Pitkäaikaissäilytys

Tutkimusaineistoja ja -dataa säilytetään usein myös erilaisten projektien loputtua jos toisin ei ole suunniteltu ja sovittu. Monesti joudutaan miettimään ja tekemään päätöksiä liittyen muun muassa pitkäaikaissäilytyksen sijaintiin ja kustannuksiin.

Väitteen “haluaisin säilyttää tutkimuksen aineistot myös projektin loputtua” kanssa vastaajista 45,1 prosenttia (213) oli täysin samaa mieltä, 31,6 prosenttia (149) samaa mieltä, 16,3 prosenttia (77) ei samaa eikä eri mieltä, 4,9 prosenttia (23) eri mieltä ja vain 2,1 prosenttia (10) vahvasti eri mieltä. Miehet (83,4 %, 151/181) ilmaisivat tahtovansa säilyttää aineistonsa todennäköisemmin kuin naiset (72,8 %, 203/279) (Fisher’s exact  $p = 0,020$ ). Jatko-opiskelijoista harvempi oli väittämän kanssa samaa mieltä (70,2 %, 185/264) kuin tutkijoista tai muissa tutkimusorganisaatioiden työrooleissa olevista (84,8 %, 173/204) (Fisher’s exact  $p < 0,001$ ). Tieteenalojen välillä ei ollut tilastollisesti merkitsevää eroja.

Kysyttäessä “kenen tulisi mielestäsi kustantaa kokoamasi aineiston säilyttäminen ja avoin pääsy siihen?” ja annettaessa mahdollisuus valita vaihtoehdoista kaikki sopivat, 71,8 prosenttia (346) valitsi “tutkimuslaitoksesi” vaihtoehdon ja 39,0 prosenttia (188) “kansallisen tahon”, 29,7 prosenttia (143) “rahoittajasi” ja 11,2 prosenttia (54) “itsesi tai ryhmäsi”. Vaihtoehdon “muu” valinneet saivat esittää avovastauksen kysymykseen. Vastauksissa esille nostettiin, että aineistoa ei tarvitse säilyttää, EU tai muu kansainvälinen taho voisi olla vastuussa säilytyksestä tai käyttäjiltä voitaisiin periä pieni maksu.

Sukupuolten välillä havaittiin tilastollisesti merkitsevää ero vaihtoehdon “itsesi tai ryhmäsi” kohdalla, johon 15,1 prosenttia (28/186) miehistä ja 8,8 pro-

senttia (25/284) naisista vastasi kyllä (Pearson  $p = 0,036$ ). Jatko-opiskelijat (76,3 %, 206/270) nostivat muita (66,3 %, 138/208) useammin esille vaihtoehdon “tutkimuslaitoksesi” (Pearson  $p = 0,016$ ). Toisaalta useampi muista kuin jatko-opiskelijoista katsoi, että maksajana voisi toimia “kansallinen taho” (45,7 %, 95/208 verrattuna 34,1 %, 92/270) (Pearson  $p = 0,010$ ). Tieteenalojen välillä ei ollut tilastollisesti merkitseviä eroja.

Aineiston pitkäaikaisen säilytyksen sijaintia kysyttäessä vastaajista 75,3 prosenttia (363) valitsi vaihtoehdon “tutkimuslaitokselle”, 45,6 prosenttia (220) tutkimusorganisaation “ulkopuoliseen eli”ulkoiseen tallennustilaan (maksuton)“, 13,5 prosenttia (65) ”ulkoiseen tallennustilaan (maksullinen)” ja 6,8 prosenttia (33) vaihtoehdon “rahoittajalle”. Vastaaja pystyi valitsemaan useampia vaihtoehtoja ja kommentoimaan myös avovastauksella. Avovastauksissa esille nostettiin muun muassa “kansallinen taho (kirjasto tms.)”, “arkisto” ja “yhteiskunnallinen tietoarkisto”.

Vaihtoehtojen valinnoissa ei ollut merkitsevää eroa naisten ja miesten välillä. Sen sijaan rooli tutkimusyhteisössä vaikutti vaihtoehtojen “ulkoiseen tallennustilaan (maksuton)” (Pearson  $p = 0,019$ ) ja “ulkoiseen tallennustilaan (maksullinen)” (Pearson  $p = 0,017$ ) valintoihin. Ensimmäinen oli suosittu muiden kuin jatko-opiskelijoiden keskuudessa (51,9 %, 108/208 verrattuna 41,1 %, 111/270) ja jälkimmäinen oli muita useammin jatko-opiskelijoiden valinta (16,7 %, 45/270 verrattuna 9,1 %, 19/208).

Tieteenalakohtaiset erot näkyivät vaihtoehtojen “tutkimuslaitoksesi/yliopistosasi” (Pearson  $p = 0,022$ ), “rahoittajasi” (Fisher’s exact  $p = 0,023$ ) ja “ulkoiseen tallennustilaan (maksuton)” (Pearson  $p = 0,035$ ) valinnoissa. Kaikki nämä vaihtoehdot olivat erityisen suosittuja insinööritieteiden ja teknologian alojen edustajien keskuudessa. Esimerkiksi “rahoittajasi” vaihtoehdon valitsi 18,8 prosenttia (9/48) tekniikan alan edustajista verrattuna vaikkapa lääketieteilijöiden vastaavaan lukuun, joka oli 5,2 prosenttia (6/116).

## Tutkimuksen vahvuudet ja rajoitteet

Kyselytutkimukseen pyrittiin saamaan edustajia erilaisista tutkimusorganisaatioista ympäri Suomea. Siksi ei pidättäydytty ainoastaan yhden tai useamman yliopiston saavuttamisessa, vaan päädyttiin hyödyntämään Tuuli-verkostoa kyselyn levittämässä. Tätä kautta tavoitettiin myös muissa tutkimusorganisaatioissa työskenteleviä ja kautta linjan monipuolisesti eri alojen tutkijoita. Yhteen yliopistoon keskittyessä kysely olisi saattanut päätyä tavoittamaan vain rajattujen tieteenalojen edustajia. Eri organisaatioiden tavoitettujen vastaajien määrät eivät kuitenkaan ole suhteessa näiden tahojen tutkija- ja henkilökuntamääriin



ja tästä johtuen luovuttiin ajatuksesta tehdä taustaorganisaatioiden välistä vertailua. Turun yliopisto, Oulun yliopisto ja Åbo Akademi olivat lopulta vahviten edustettuja organisaatioita.

Kyselyn kysymyksiä ei laadittu itse ja joitain asioita olisi voinut kysyä erilaista kysymyksenasettelua käyttäen tai täsmentäviä lisäkysymyksiä hyödyntäen. Alun perin kysely on suunniteltu englanniksi jakaikki kysymykset eivät mahdollisesti sellaisenaan toimineet parhaiten suomennettuina. Kyselyssä “research data” käännettiin “tutkimusaineistoksi”. Vaikka kyselyn pääpaino oli digitaalisessa aineistossa, sanan tutkimusdata katsottiin voivan olla monelle vastaajalle yhä vieras. On siis mahdollista, että vastatessaan osa saattoi siis ajatella myös muunlaisen kuin digitaalisen aineiston jakamista ja säilyttämistä. Näin toki on englanniksikin, koska “research data” viittaa myös muunlaiseen kuin digitaaliseen aineistoon.

Tutkimusorganisaatioiden työntekijöiden ja tutkijoiden tavoitettavuuden vaihtelevuus vaikuttaa tutkimuksen yleistettävyyteen ja kattavuuteen. Tilastollinen analyysi oli myös suhteellisen yksinkertaista.

“Ei samaa eikä eri mieltä” sai osassa kysymyksistä suhteellisen paljon kannatusta. Tämä saattaa kertoa siitä, etteivät avoimuuteen ja tutkimusaineiston jakamiseen liittyvät kysymykset ole vielä kaikille tuttuja. Kyselyn aihe saattoi myös tuntua vieraalta osalle kyselyä täyttämään aloittaneista. Tämä saattoi johtaa keskeytykseen.

## Pohdintaa

Tässä artikkelissa esiteltävä tutkimus lisää ymmärrystämme suomalaisten tutkimusorganisaatioiden jäsenten käsityksistä, kokemuksista ja mielipiteistä tutkimusaineiston ja -datan jakamiseen ja pitkäaikaissäilytykseen liittyen. Tarkastelussa esille nostetaan myös sukupuolten, työroolien ja tieteenalojen väliset mahdolliset eroavaisuudet ja pohditaan niihin liittyviä syitä.

Kysymyksissä “suosittelee tutkimuslaitoksesi aktiivisesti aineistojen jakamista vapaan pääsyn (open access) periaatteella” ja “voin huoletti ja mielelläni jakaa tutkimusaineistoni muiden kanssa” vastaajien kannat jakautuivat sangen tasaisesti samaa tai eri mieltä olevien, sekä niiden välillä, jotka eivät osanneet ottaa kantaa. Osa näki aineistojen jakamiseen liittyvän ongelmia ja huolenaiheina annetuista vaihtoehdoista esille nousivat erityisesti laillisuuteen ja eettisyyteen, aineiston väärinkäyttöön ja sen sisältämän tiedon väärinymmärrykseen liittyvät pelot. Suurin osa vastaajista kuvasi käyttämänsä tutkimusaineiston olevan vapaasti heidän tutkimusryhmänsä käytössä tai muiden tutkijoiden käytettävissä erillispyynnöstä. Suurin osa vastaajista toivoi aineistonsa säilyvän myös tutki-

musprojektin loputtua ja pitkäaikaissäilytyksen kustannusten katsottiin kuuluvan tutkimusorganisaatiolleen tai “kansalliselle taholle”. Säilytyksen sijainniksi suurin osa toivoi joko tutkimusorganisaatiotaan tai ulkopuolista maksutonta aineistonsäilytyspalvelua. Myös Fältin ja Okuloffin (2016) tutkimuksessa havaittiin, että vaikkakin vastaajat olivat todennäköisesti aiheesta kiinnostuneita, heistä vain puolet tiesi oliko heidän organisaatiossaan ohjeita tutkimusaineistojen säilyttämisestä tai tutkimusaineistojen avaamisesta muuhun kuin tutkijan/tutkimusryhmän omaan käyttöön.

Tutkimuksen perusteella naiset näkivät miehiä enemmän ongelmia tutkimusaineiston jakamiseen liittyen ja heissä jossain määrin useampi koki, että eettisten näkökulmien huomioiminen tutkimusdataa jaettaessa on keskeistä. Tämä tulos tukee Linekin ja hänen kollegoidensa havaintoa sukupuolten välisistä eroista (2017). Yksi syy eroihin voi löytyä eri alojen tutkijoiden sukupuolijakauksista ja eri aloihin liittyvistä ominaispiirteistä. Nais- ja miesvastaajat saattavat käyttää työssään erityyppisiä tutkimusaineistoja. Täsmällisemmin kysytyjen huolien ja pelkojen suhteen sukupuolten välillä ei kuitenkaan ollut tilastollisesti merkitseviä eroja.

Rooli tiedeyhteisössä oli yhteydessä muun muassa siihen, kokiko vastaaja että hänen tutkimuslaitoksensa suosittelee aktiivisesti aineiston avointa jakamista. Jatko-opiskelijat kokivat muita harvemmin näin olevan. Lisäksi he vastasivat muissa työrooleissa toimivia harvemmin että määritelmä “aineistoni on vapaasti tutkimusryhmäni käytettävissä”, “aineisto on käytettävissä pyynnöstä” tai “aineistoon on vain rajattu pääsy (esim. osa aineistosta on yleisesti käytettävissä)” kuvasi heidän käyttämäänsä tutkimusaineistoa. Sen sijaan jatko-opiskelijoiden kohdalla “aineistoon ei ole pääsyä muilla kuin minulla” oli suosittu vaihtoehto (29,3 %). Toisaalta, aineiston väärinkäyttö tai sen sisältämän tiedon väärä tulkinta eivät huolestuttaneet jatko-opiskelijoita yhtä paljon kuin muita vastaajia. Tutkijan uran alkuvaiheessa oleville ei ehkä vielä ole korostettu aineiston avoimen jakamisen mahdollisuuksia ja toisaalta siihen liittyvät huoletkaan eivät ole vielä nostaneet päätään siinä määrin kuin kauemmin asiaa pohtineilla. Tietylnaista arkuutta aiheeseen liittyen voi myös olla havaittavissa. Tutkimustulokset heijastelevat myös sitä tosiasiaa, että yhä edelleen moni jatko-opiskelija kerää aineistonsa yksin ja mahdollisesti ilman yhteyttä mihinkään tutkimusryhmään.

Tieteenalojen välillä havaittiin kiinnostavia eroja. Lääke- ja terveystieteen edustajat erottuivat muista ollen pääosin eri mieltä siitä, että he voivat huolehti ja mielellään jakaa tutkimusaineistoa muiden kanssa. Lisäksi he jakoivat yhteiskuntatieteiden edustajien kanssa vahvimmin näkemyksen, että eettisten näkökulmien huomioiminen tutkimusdataa jaettaessa on keskeistä. Luonnon-tieteilijät, yhteiskuntatieteilijät ja ihmistieteilijät vastasivat muita useammin, et-

tä heidän aineistonsa on vapaasti kaikkien käytettävissä. Toisaalta vaihtoehdon “aineistooni ei ole pääsyä muilla kuin minulla” oli valittu muita useammin yhteiskuntatieteilijöiden toimesta. Pelko tieteellisen etumatkan menettämisestä huolestutti erityisesti luonnontieteilijöitä ja laillisuuteen ja eettisyyteen liittyvät kysymykset huolestuttivat muita enemmän insinööritieteiden ja teknologioiden alojen edustajia, yhteiskuntatieteilijöitä ja ihmistieteilijöitä. Pitkäaikaissäilytyksen sijaintiehdotuksia mietittäessä erityisesti insinööritieteiden ja teknologian alojen edustajat suosivat vaihtoehtoja “tutkimuslaitoksesi/yliopistosasi”, “rahoittajasi” ja “ulkoiseen tallennustilaan (maksuton)”. Oli kuitenkin kiinnostavaa, että monessa pitkäaikaissäilytykseen liittyvässä kysymyksessä ei ollut tieteenalakohtaisia tilastollisesti merkitseviä eroja. Aiemmissä tutkimuksissa eri alojen ominaispiirteet nousivat esille myös Fältin ja Okuloffin (2016) selvityksessä, joskin siinä tarkastelu kohdistui vain rajattuihin aloihin.

## Tutkimuksen merkitys ja jatkotutkimusideat

Koulutuksessa ja tutkimusdatan avoimeen jakamiseen kannustamisessa voitaisiin entistä paremmin huomioida myös kohdeyleisön asema tutkimusyhteisössä ja eri tieteellisille aloille ominaiset piirteet, joita tämänkin tutkimuksen tulokset heijastelevat. Myös pitkäaikaissäilytykseen liittyvien toiveiden ja käsitysten voisi olettaa kiinnostavan niin tutkimusorganisaatioita kuin rahoittajakin.

Avoimen tieteen ja tutkimusdatan avoimen jakamisen aiheiden ollessa tapetilla tutkimusorganisaatioissa, kattavaa käsitysten ja kokemusten kartoittamista olisi syytä tehdä lisää. League of European Research Universities -verkoston raportti (2018) nosti suurena ongelmana esille myös tutkimuskulttuurin muutoksen. Muutos vaatii avoimuuden huomioimista aineistonkeruun ja -hallinnan joka vaiheessa. Pienempiäkin kehitysaskelia, muun muassa totuttujen tutkimusaineistojen keruun saateviestien sanamuodoissa, tulisi seurata tarkkaavaisesti. Myös eri tieteenalojen tutkijoiden näkemysten vertailu on ilmiselvästi yhä vähäistä. Lisäksi olisi mielenkiintoista selvittää tutkimusdataansa avanneiden ja avattuja dataa käyttäneiden kokemuksia.

Kysely, johon tässä artikkelissa esitelty tutkimus perustui, sisältää lisäksi kysymyksiä muun muassa metadatatista ja aineistonhallintasuunnitelman teosta. Näitä aiheita tullaan käsittelemään tarkemmin tulevassa artikkelissa. Kuten aiemmin mainittiinkin, tämän tutkimuksen kysely on toistettu ainakin 24 eri maassa ja kyselyjen tuloksista tullaan myös julkaisemaan vertaileva tutkimus. Oletus on että Suomi erottuu useissa käsitellyissä aiheissa kärkimaana, vaikka toisaalta moni asia, kuten avoimuuteen liittyvät huolet, todennäköisesti yhdistävät eri maiden tutkimusorganisaatioiden työntekijöitä ja tutkijoita.

## Johtopäätökset

Artikkelissa esitellyt tulokset antavat yleiskuvaa suomalaisten tutkimusorganisaatioiden jäsenten käsityksistä ja mielipiteistä tutkimusdatan jakamiseen ja säilyttämiseen liittyen. Lisäksi tarkastellaan näiden kysymysten suhdetta vastaajan sukupuoleen, rooliin tiedeyhteisössä ja tieteenalaan.

Suhtautuminen avoimeen tutkimusaineistojen ja -datan jakamiseen jakaa yhä mielipiteitä. Tulokset kertovat myös siitä, ettei aihe ole kaikille tutkimusorganisaatioiden työntekijöille ja tutkijoille vielä tuttu.

Vastaajien välillä havaittiin tilastollisesti merkitseviä eroja. Naiset esimerkiksi näkivät miehiä enemmän ongelmia tutkimusaineiston jakamiseen liittyen ja heissä jossain määrin useampi koki, että eettisten näkökulmien huomioiminen tutkimusdataa jaettaessa on keskeistä. Sukupuolella voi olla yhteys myös tutkittavien tieteenalaan ja tämä voi osaltaan selittää osaa havainnoista. Rooliin tiedeyhteisössä kytkeytyy havainto siitä, että tutkijan uran alkuvaiheessa oleville ei ehkä vielä ole korostettu aineiston avoimen jakamisen mahdollisuuksia. Myös arkuutta aiheeseen liittyen voi myös olla havaittavissa. Tieteenalakohtaisessa vertailussa eniten ongelmia avoimeen jakamiseen liittyen näkivät lääke- ja terveystieteiden edustajat. Erot tieteenalojen välillä kertonevat erilaisten tutkimusmenetelmien ja -aineistojen lisäksi myös eri alojen kulttuureista ja perinteisistä tavoista suhtautua aineistoihin.

Tulosten perusteella toiveita pitkäaikaissäilytykseen liittyen on, mutta miten saada ihmiset myös avaamaan datansa tietokoneella, muistitikuilla tai ulkoisilla kova-aseilla säilyttämisen sijaan. Avoimen tieteen ja tutkimusdatan avoimen jakamisen aiheiden ollessa tapetilla tutkimusorganisaatioissa, kattavaa käsitysten ja kokemusten kartoittamista olisi syytä tehdä lisää.

## Lähteet

- 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>
- ANDS. (2018). Guides and resources. Working with data. Ethics and data sharing. *Australian national data service*. <https://www.ands.org.au/working-with-data/sensitive-data/ethics-and-data-sharing>
- Avoin tiede ja tutkimus -hanke. (2018a). Avoin tiede ja tutkimus. <https://avointiede.fi/>
- Avoin tiede ja tutkimus -hanke. (2018b). Avoin tiede ja tutkimus. Avoimen tieteen kansalliset palvelut. <https://avointiede.fi/fi/data/palveluita>
- Borgman, C. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6), 1059–1078. <https://doi.org/10.1002/asi.22634>
- Clarivate Analytics. (2018). Data Citation Index. [http://wokinfo.com/products\\_tools/multidisciplinary/dci/](http://wokinfo.com/products_tools/multidisciplinary/dci/)

- Dorch, B., Drachen, T., & Ellegaard, O. (2015). The data sharing advantage in astrophysics. *Proceedings of the International Astronomical Union*, 11(A29A), 172–175. <https://doi.org/10.1017/S1743921316002696>
- Drachen, T., Ellegaard, O., Larsen, A., & Dorch, S. (2016). Sharing data increases citations. *LIBER Quarterly*, 26(2), 67–82. <https://doi.org/10.18352/lq.10149>
- European Commission. (2016). H2020 Programme – guidelines on FAIR data management in Horizon 2020. Version 3.0. [https://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)
- Fält, K., & Okuloff, A. (2016). *Avoin tiede ja tutkimusaineistot –kysely 2015. Raportti terveystieteiden ja humanististen alojen tutkijoille suunnatun kyselyn tuloksista* (No. 12/2016). Tampere: Yhteiskunnallinen tietoaarkisto. [http://www.fsd.uta.fi/fi/julkaisut/julkaisusarja/FSDjs12\\_tutkimusaineistot-kysely.pdf](http://www.fsd.uta.fi/fi/julkaisut/julkaisusarja/FSDjs12_tutkimusaineistot-kysely.pdf)
- Feldman, L., Patel, D., Ortmann, L., Robinson, K., & Popovic, T. (2012). Educating for the future: another important benefit of data sharing. *The Lancet*, 379(9829), 1877–1878. [https://doi.org/10.1016/S0140-6736\(12\)60809-5](https://doi.org/10.1016/S0140-6736(12)60809-5)
- Finnish Committee for Research Data. (2018). *Tracing data: data citation roadmap for Finland*. Finnish Committee for Research Data. <http://urn.fi/URN:NBN:fi-fe201804106446>
- Joint Declaration of Data Citation Principles. Martone M. (ed.). (2014). *FORCE11*. <https://doi.org/10.25490/a97f-egy>
- Koltay, T. (2017). Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science*, 49(1), 3–14. <https://doi.org/10.1177/0961000615616450>
- Laine, H., & Nykyri, S. (2018). Dataviittaamisen tiekartta tutkijalle. *Informaatiotutkimus*, 37(2). <https://doi.org/10.23978/inf.72999>
- League of European Research Universities. (2018). *Open Science and its role in universities: a roadmap for cultural change*. <https://www.leru.org/files/LERU-AP24-Open-Science-full-paper.pdf>
- Linek, S., Fecher, B., Friesike, S., & Hebing, M. (2017). Data sharing as social dilemma: Influence of the researcher's personality. *PLOS ONE*, 12(8), e0183216. <https://doi.org/10.1371/journal.pone.0183216>
- McKiernan, E. (2017). Imagining the "open" university: Sharing scholarship to improve research and education. *PLOS Biology*, 15(10), e1002614. <https://doi.org/10.1371/journal.pbio.1002614>
- Mongeon, P., Robinson-Garcia, N., Jeng, W., & Costas, R. (2017). Incorporating data sharing to the reward system of science: Linking DataCite records to authors in the Web of Science. *Aslib Journal of Information Management*, 69(5), 545–556. <https://doi.org/10.1108/AJIM-01-2017-0024>
- Mooney, H., & Newton, M. (2012). The Anatomy of a Data Citation: Discovery, Reuse, and Credit. *Journal of Librarianship and Scholarly Communication*, 1(1), eP1035. <https://doi.org/10.7710/2162-3309.1035>
- Olfson, M., Wall, M., & Blanco, C. (2017). Incentivizing data sharing and collaboration in medical research – the S-Index. *JAMA Psychiatry*, 74(1), 5–6. <https://doi.org/10.1001/jamapsychiatry.2016.2610>
- Pienta, A., Alter, G., & Lyle, J. (2010). The enduring value of social science research: the use and reuse of primary research data. Teoksessa *The Organisation, Economics and Policy of Scientific Research" workshop*. Torino, Italy. <http://hdl.handle.net/2027.42/78307>
- Piwovar, H., & Vision, T. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1, e175. <https://doi.org/10.7717/peerj.175>
- Research Data Alliance. (2018). Research Data Alliance. <https://www.rd-alliance.org/>

- Rowhani-Farid, A., Allen, M., & Barnett, A. (2017). What incentives increase data sharing in health and medical research? A systematic review. *Research Integrity and Peer Review*, 2(1), 4. <https://doi.org/10.1186/s41073-017-0028-9>
- San Francisco Declaration on Research Assessment. (2018). San Francisco Declaration on Research Assessment. <https://sfedora.org/read/>
- Sarpatawari, A., Kesselheim, A., Malin, B., Gagne, J., & Schneeweiss, S. (2014). Ensuring Patient Privacy in Data Sharing for Postapproval Research. *New England Journal of Medicine*, (371), 1644–1649. <https://doi.org/10.1056/NEJMs1405487>
- Scaramozzino, J., Ramírez, M., & McGaughey, K. (2012). A Study of Faculty Data Curation Behaviors and Attitudes at a Teaching-Centered University. *College and Research Libraries*, 73(4), 349–365. <https://doi.org/https://doi.org/10.5860/cr1-255>
- Shearer, K., & Furtado, F. (2017). *COAR Survey of research data management: Results*. Confederation of Open Access Repositories (COAR). <https://www.coar-repositories.org/files/COAR-RDM-Survey-Jan-2017.pdf>
- Strasser, C. (2015). *Research Data Management*. NISO. <http://wiki.lib.sun.ac.za/images/2/24/PrimerRDM-2015-0727.pdf>
- Suomen yliopistojen rehtorineuvosto UNIFI ry. (2018). Avoin tiede ja data – Toimenpideohjelman suomalaiselle tiedeyhteisölle. *Suomen yliopistojen rehtorineuvosto UNIFI ry*. <http://www.unifi.fi/toiminta/hankeet/avoin-tiede/>
- Swan, A., Gargouri, Y., Hunt, M., & Harnad, S. (2015). *Working Together to Promote Open Access Policy Alignment in Europe – Work Package 3 report: Open Access Policies. March 2015. PASTEUR4oA Project, European Commission*. PASTEUR4oA Project, European Commission. <https://arxiv.org/ftp/arxiv/papers/1504/1504.02261.pdf>
- Thessen, A., & Patterson, D. (2011). Data issues in the life sciences. *ZooKeys*, 150, 15–51. <https://doi.org/10.3897/zookeys.150.1766>
- Tietoarkisto. (2018). Tietoarkisto. *Tietoarkisto*. <https://www.fsd.uta.fi/fi/>
- Tripathi, M., Shukla, A., & Sonkar, S. (2017). Research data management practices in university libraries: a study. *Journal of Library and Information Technology*, 37(6), 417–424. <https://doi.org/10.14429/djlit.37.6.11336>
- Tuuli-projekti. (2017). Tuuli – aineistonhallintasuunnitelmatyökalu. <https://wiki.helsinki.fi/display/Tuuli/Tuuli-projekti>
- Van den Eynden, V., Knight, G., Vlad, A., Radler, B., Tenopir, C., Leon, D., ... Corti, L. (2016). Towards open research: practices, experiences, barriers and opportunities. <https://doi.org/10.6084/m9.figshare.4055448>
- 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*, 3. <https://doi.org/10.1038/sdata.2016.18>