

Tiina Kautonen

**AVOIN TIEDE JA TUTKIMUSAINEISTOT:
TUTKIJOIDEN NÄKEMYKSIÄ JA
KÄYTÄNTÖJÄ**

Informaatioteknologian ja
viestinnän tiedekunta
Pro gradu -tutkielma
Toukokuu 2021

TIIVISTELMÄ

Tiina Kautonen: Avoin tiede ja tutkimusaineistot – Tutkijoiden näkemyksiä ja käytäntöjä
Pro Gradu
Tampereen yliopisto
Informaatiotutkimuksen maisteriohjelma
Toukokuu 2021

Tämä pro gradu -tutkielma käsittelee avointa tiedettä ja tutkimusaineistoja. Avoin tiede on tiedemaailmassa vaikuttava ajattelutapa, jonka mukaan tutkimustulokset tulisi jakaa laajalle yleisölle ilman maksua. Yksi avoimen tieteen osa-alue on tutkimusaineistojen avoimuus.

Tutkielman ensimmäinen tutkimuskysymys koskee sitä, mitä käytäntöjä informaatiotutkimuksen ja sosiologian tutkijoilla on tutkimusaineistojen hallinnassa. Toinen tutkimuskysymys on, miten informaatiotutkimuksen ja sosiologian tutkijat suhtautuvat aineistojen avoimuuteen, ja mitä hyötyjä ja ongelmia aineistojen avoimuudessa nähdään. Kolmas tutkimuskysymys on, miten informaatiotutkimuksen ja sosiologian tutkijat näkevät tilanteen aineistojen avoimuuden suhteen omilla tieteenaloillaan.

Tutkimuksen empiirinen osa on toteutettu haastattelemalla kuutta Tampereen yliopiston tutkijaa. Tutkimusmenetelmänä on teemahaastattelu ja aineiston analyysimenetelmänä on teoriaohjaava sisällönanalyysi.

Tulosten mukaan aineistohallintasuunnitelma tehtiin yleensä, jos tutkimusrahoittaja sitä vaati. Tutkimuksen jälkeen aineistot vaihtelevasti joko tallennettiin avoimesti saataville, tuhottiin tai säilytettiin itsellä. Tutkijat suhtautuvat aineistojen jakamiseen positiivisesti, mutta käytännössä siihen vaikuttavat monet eri tekijät. Aineiston avoimuuden keskeisiä hyötyjä ovat aineistojen uudelleenkäyttö uusissa yhteyksissä, tutkijoiden ajan ja vaivan säästyminen ja tutkimuksen läpinäkyvyyden ja todennettavuuden lisääntyminen.

Keskeisin haaste aineistojen avoimuudessa liittyy tutkittavien tietosuojaan. Aineiston anonymisointi vaatii aikaa ja osaamista ja voi hankaloittaa aineiston jatkokäyttöä. Muita aineistojen jakamiseen vaikuttavia tekijöitä olivat muun muassa aineiston omistusoikeudet, tutkijoiden ajan puute ja aineiston relevanssi muille tutkijoille.

Sosiologiassa aineistojen avoimuus on yleistä ja Suomessa aineistojen tallennuspaikka on Yhteiskuntatieteellinen tietoarkisto. Informaatiotutkimuksessa aineistoja jaetaan useissa eri paikoissa, ja sensitiivisiä aineistoja jaetaan vähemmän kuin neutraaleja aiheita käsitteleviä aineistoja. Aineistojen avoimuus on lisääntynyt viime vuosina molemmilla aloilla.

Avainsanat: Avoin tiede, tutkija, tutkimusaineisto, tieteenala, aineistohallinta

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -ohjelmalla.

Sisällysluettelo

1	JOHDANTO	1
2	TUTKIMUSAINEISTOT JA TIETEENALAT	3
	2.1 Tutkimusaineistojen määrittelyä	3
	2.2 Tutkimusaineistot eri tieteenaloilla	5
3	AVOIN TIEDE JA TUTKIMUSAINEISTOT	9
	3.1 Avoimen tieteen taustaa.....	9
	3.2 Avoimen tieteen hankkeita ja toimijoita	12
4	TUTKIMUSAINEISTOJEN AVOIMUUDEN KÄYTÄNTÖJÄ.....	15
	4.1 Tutkimusaineistojen hallinta ja elinkaari	15
	4.2 Aineistojen löydettävyys ja FAIR-data periaatteet.....	18
	4.3 Data-arkistojen, tieteellisten lehtien ja tieteellisten kirjastojen rooli	20
5	NÄKÖKULMIA AINEISTOJEN AVOIMUUTEEN	23
	5.1 Hyödyt ja esteet	23
	5.2 Tutkijoiden näkemyksiä.....	27
6	TUTKIMUSASETELMA.....	33
	6.1 Tutkimuskysymykset.....	33
	6.2 Tutkimusmenetelmä.....	34
	6.3 Aineiston analyysi	35
7	TUTKIMUSTULOKSET	38
	7.1 Taustatiedot	38
	7.2 Aineistonhallinta	39
	7.3 Avoimuuden hyödyt.....	42
	7.4 Avoimuuteen vaikuttavat tekijät, ongelmat ja esteet.....	44
	7.5 Näkemyksiä avoimuuden tilanteesta.....	50
8	JOHTOPÄÄTÖKSET JA POHDINTA.....	55
	LÄHTEET:	61

LIITE 1: TEEMAHAASTattelun kysymykset

1 JOHDANTO

Avoimuus ja tiedon jakaminen ovat tieteellisen tutkimuksen keskeisiä periaatteita. Viime vuosikymmeninä tieteessä on siirretty kohti avoimempaa kulttuuria mm. internetin kehityksen myötä. Avoin tiede on tiedemaailmassa vaikuttava liike ja ajattelutapa, jonka tarkoitus on julkaista tieteellisen tutkimuksen tulokset laajalle yleisölle ilman maksua. Tutkimusjulkaisujen lisäksi avoimuuden tavoitteet ulottuvat koskemaan koko tieteenteon prosessia, tutkimusaineistoja ja oppimisympäristöjä. Avointa tiedettä edistämään on perustettu kansallisia ja kansainvälisiä hankkeita, tutkimusrahoittajat ovat alkaneet edellyttää avoimuutta ja tutkimusorganisaatiot sekä yliopistojen kirjastot ovat alkaneet kehittää avoimen tieteen linjauksia ja palveluita.

Kiinnostuin avoimesta tieteestä ilmiönä opintojen aikana. Tekemäni työharjoittelu tieteellisessä kirjastossa ja tutustuminen aiheesta tehtyyn aikaisempaan tutkimukseen vahvisti käsitystäni siitä, miten ajankohtainen aihe on kyseessä.

Tämä pro gradu -tutkielma keskittyy tarkemmin yhteen avoimen tieteen osa-alueeseen, eli tutkimusaineistojen avoimuuteen. Tieteellisessä tutkimusprosessissa syntyä ja käytetään tieteenalasta riippuen hyvin erilaisia aineistoja vahvistamaan tutkimustuloksia. Tutkijat voivat jakaa tutkimusaineistojaan suoraan toisilleen, mutta niiden jakamiseen on myös esimerkiksi kaikille tutkijoille avoimia data-arkistoja. Aineistojen jakaminen mahdollistaa toisten tutkijoiden keräämien aineistojen hyödyntämisen tutkimustyössä. Aineistojen avoimuuteen kytkeytyvät kiinteästi tutkimusaineistojen hallintaan liittyvät kysymykset. Aineistonhallinnalla tarkoitetaan aineistojen käsittelyä koko sen elinkaaren ajalta keräämisestä jatkokäyttöön. Jo aineiston keruuvaiheessa tehdyt päätökset voivat vaikuttaa siihen, voiko aineistoa jakaa myöhemmin.

Käytännössä avointa tiedettä ja sen osana tutkimusaineistojen avoimuutta toteuttavat tutkijat jokapäiväisessä tutkimustyössään. Miten yleistä aineistojen jakaminen todella

on? Miten tutkijat toteuttavat aineistonhallintaa ja mikä vaikuttaa käytännössä siihen, päättävätkö he tallentaa aineistonsa muiden käytettäväksi? Mitä hyötyjä ja haasteita aineistojen avoimuuteen liittyy? Näihin kysymyksiin lähdin hakemaan tutkielmallani vastausta. Tutkielman empiirinen osa on tehty haastattelemalla Tampereen yliopiston tutkijoita. Menetelmäksi valikoitui teemahaastattelu, jonka avulla on mahdollista saada tutkijoilta syvällisempiä vastauksia kuin kyselytutkimuksella. Tutkielman laajuus huomioon ottaen päädyin tarkastelemaan tilannetta kahdella itseäni kiinnostavalla tieteenalalla: informaatiotutkimuksessa ja sosiologiassa.

Luvut 2–5 luovat tutkimukselle viitekehyksen. Luvussa 2 määritellään sitä, mitä tutkimusaineistot voivat olla eri tieteenaloilla. Luvussa 3 taustoitetaan avointa tiedettä ja luku 4 keskittyy aineistojen avoimuuden käytäntöihin ja aineistonhallintaan. Luvussa 5 käsitellään kirjallisuudessa tunnistettuja aineiston avoimuuden hyötyjä ja ongelmia sekä vastaavia aikaisempia tutkimuksia tutkijoiden näkemyksistä. Luvussa 6 esitellään tutkimusasetelma: tutkimuksen tarkoitus ja tutkimuskysymykset sekä tutkimusmenetelmä. Luku 7 esittää tutkimuksen tulokset. Viimeisessä luvussa vastataan vielä tutkimuskysymyksiin ja verrataan tuloksia aiempaan vastaavaan tutkimukseen.

2 TUTKIMUSAINIESTOT JA TIETEENALAT

Tässä työssä tarkastellaan *sosiologian* ja *informaatiotutkimuksen* alojen *tutkimusaineistoja*. Luvussa 2.1 määritellään termiä tutkimusaineisto ja esitetään yleisellä tasolla sitä minkälaisia tutkimusaineistot voivat olla. Luvussa 2.2 eritellään sitä, miten eri *tieteenalojen* tutkimusaineistot eroavat toisistaan. Tässä luvussa myös esitellään tarkastelun kohteena olevat tieteenalat sosiologia ja informaatiotutkimus.

2.1 Tutkimusaineistojen määrittelyä

Tutkimusaineistoille (englanniksi *research data*) ei ole yksiselitteistä määritelmää. Kun puhutaan tutkimusaineistoista, on tyypillistä, että jokainen kirjoittaja määrittelee termin omalla tavallaan. Dietrich ja Wiebe (2013, 17) määrittelevät tutkimusaineiston tarkoittavan tieteellisessä tutkimusprosessissa syntynyttä aineistoa, joka voi olla monessa eri muodossa. Toisaalta tutkimuksissa käytetään myös valmiita aineistoja, jotka eivät ole tutkimusta varten luotuja. OECD:n (2007, 13–14) määritelmän mukaan tutkimusaineisto voi olla numeerista aineistoa, tekstiä, kuvia tai ääntä, joita käytetään tieteellisen tutkimuksen keskeisenä lähteenä. Määritelmän mukaan aineiston tulee olla tieteellisen yhteisön hyväksymää ja tarpeellista tutkimustulosten aikaansaamiseksi. Suomen kielessä käytetään osin päällekkäin termejä tutkimusaineisto ja tutkimusdata. Tutkimusdatalla tarkoitetaan yleensä digitaalisessa muodossa olevaa tutkimuksessa kerättyä tai syntynyttä aineistoa (Late ja Talja 2017, 45). Tässä työssä käytetään termiä tutkimusaineisto, ja sillä viitataan kaikkeen tieteellisessä tutkimusprosessissa syntyneeseen tai kerättyyn aineistoon. Käsite tutkimusaineisto ymmärretään laajasti ja tutkielman empiirisessä osassa tutkimusaineistona käsiteltiin kaikkea, minkä tutkija itse näki aineistona.

Cox ja Verbaan (2018, 22) listaavat tutkimusaineistojen tyyppejä seuraavasti:

- Tieteellisten kokeiden tulokset
- Kenttätutkimusten mittaustulokset
- Tietokoneohjelmat
- Haastattelujen äänitteet ja puhtaaksikirjoitukset
- Kyselytutkimusten vastaukset
- Kuvat
- Liikkuvat kuvat
- Historialliset dokumentit
- Fyysiset objektit
- Sosiaalisen median aineistot kuten twiitit
- Webliikenteen lokitiedot

Listauksesta havaitaan, että laajasti käsitettynä tutkimusaineistoja voi olla hyvin erilaisia. Suuri osa tutkimusaineistosta on nykyisin digitaalisessa muodossa. Kuitenkaan kaikki tutkimusaineisto ei ole digitaalisessa muodossa (Cox ja Verbaan 2018, 22).

Tutkimusaineisto voi olla kvantitatiivista eli määrällistä tai kvalitatiivista eli laadullista. Kvantitatiivinen tutkimus keskittyy keräämään ja analysoimaan aineistoa, joka on strukturoitua ja joka voidaan esittää numeerisessa muodossa. Kvantitatiivista aineistoa voidaan mitata ja analysoida tilastollisesti. (Goertzen 2017, 12.) Kvalitatiivista aineistoa taas on kaikki ihmistieteissä syntynyt tai kerätty aineisto, joka ei ole määrällisessä muodossa. Yleensä kvalitatiivinen tutkimus kuvaa kokemuksia elämästä ja merkityksiä, joita ihmiset antava näille kokemuksille. Tyypillisiä laadullisia aineistoja ovat yksilö- ja ryhmähaastattelujen aineistot, kenttätutkimuksen päiväkirjat, havainnot, strukturoidut ja strukturoimattomat päiväkirjat, henkilökohtaiset dokumentit, valokuvat ja videot.

Jos ihmisistä kerätty aineisto on muutettu numeeriseen muotoon, on kyseessä kvantitatiivinen aineisto. (Corti 2007, 38–39.)

2.2 Tutkimusaineistot eri tieteenaloilla

Eri tieteenaloilla syntyvät tutkimusaineistot eroavat toisistaan. Käsitteelle tieteenala ei kuitenkaan ole olemassa yksiselitteistä määritelmää. Krishnanin (2009, 9) mukaan tieteenaloihin sisältyy organisoitua oppimista ja uuden tiedon systemaattista tuottamista. Hän määrittelee tieteenalalle kuusi keskeistä ominaisuutta, jotka eivät välttämättä kaikki toteudu kaikilla tieteenaloilla. Ensinnäkin tieteenalalla on oma tutkimuskohteensa ja toiseksi tutkimuskohteeseen liittyvää kertynyttä tietoa. Kolmanneksi tieteenaloilla on myös teorioita ja käsitteitä, jotka auttavat organisoimaan tietoa. Neljänneksi tieteenalalla on tutkimuskohteeseen sopivaa terminologiaa ja käsitteitä. Viidentenä kohtana tieteenalat ovat myös kehittäneet tutkimuskohteeseen sopivia metodeja. Kuudenneksi tieteenaloilla on opetukseen ja tutkimukseen liittyvä asema yliopistolla.

Eri tieteenaloilla käytetään erilaisia tutkimusaineistoja, ja myös tieteenalojen sisällä aineistot vaihtelevat tutkimuskohteen mukaan. Luonnontieteellisillä aloilla on tyypillistä, että aineistot on kerätty tutkimusta varten. Tutkijat käyttävät sekä itse keräämäänsä että toisten tutkijoiden keräämiä aineistoja. Esimerkiksi biologiassa tyypillisiä aineistoja ovat näytteet, lääketieteessä testitulokset ja kemiassa proteiiniirakenteet. Luonnontieteellisiä aineistoja ovat myös esimerkiksi laboratorio- tai kenttämuistiinpanot ja laskentataulukot. (Borgman 2008, 31.)

Myös yhteiskuntatieteissä esiintyy monenlaisia aineistoja alasta riippuen. Aineistoja voivat olla esimerkiksi mielipidetutkimukset, kyselytutkimukset, haastattelut, kokeelliset tutkimukset, väestönlaskennat, äänestystulokset tai taloudelliset indikaattorit. Tyypillisimmin yhteiskuntatieteilijät keräävät itse aineistonsa, mutta

myös muiden keräämiä aineistoja käytetään. Luonnontieteellisistä aineistoista yhteiskuntatieteelliset aineistot eroavat siinä, että niissä käytetään myös aineistoja, jotka eivät ole kerätty tutkimusta varten. (Borgman 2008, 31.) Yhteiskuntatieteissä käytetään sekä kvantitatiivisia että kvalitatiivisia aineistoja.

Humanististen tieteiden aineistot eroavat kaikkein eniten toisistaan, sillä lähes mikä tahansa ihmisen toiminnasta syntynyt tallenne voi toimia aineistona. Aineistoa voivat olla esimerkiksi sanomalehdet, kirjat, valokuvat, kirjeet, syntymä - ja kuolintodistukset, artikkelit tai vaikka kirkkojen ja tuomioistuinten dokumentit. (Borgman 2008, 32.) Esimerkiksi historiantutkimuksessa käytetään paljon ensisijaisia lähteitä, joita ovat arkistoaineistot ja alkuperäiset dokumentit. Nämä lähteet voidaan nähdä silloin tutkimusaineistoina. (Cox ja Verbaan 2018, 20.) Humanistisille aloille tyypillistä on, että aineiston ja lähteen välinen raja voi olla epäselvä. Humanistit myös keräävät harvemmin itse lähteensä verrattuna yhteiskuntatieteisiin ja luonnontieteisiin ja käyttävät lähteenä enemmän ulkopuolisia tahoja, esimerkiksi kirjastoja ja arkistoja. (Borgman 2008, 32.) Erittäin teoreettisilla aloilla, kuten filosofiassa, taas ei välttämättä ole lainkaan varsinaisia tutkimusaineistoja (Cox ja Verbaan 2018, 20).

Tässä työssä tarkastellaan kahdella tieteenalalla, sosiologiassa ja informaatiotutkimuksessa, käytettyjä tutkimusaineistoja. Seuraavaksi esitellään lyhyesti nämä tieteenalat.

Sosiologia on yleinen yhteiskuntatiede, joka tarkastelee yksilön ja yhteiskunnan suhdetta. Sosiologian tavoitteena on ymmärtää yhteiskuntaa, sen yksilölle tarjoamia toimintaedellytyksiä ja näiden muutosta. Sosiologiassa muodostetaan yleistyksiä yhteiskunnasta ja ihmisten toiminnasta siinä. Painopiste on yksilöiden sijaan sosiaalisissa yhteisöissä. Ihmisten toiminta ja vuorovaikutus tietynä aikana muodostaa sosiaalisen järjestelmän, joka on sosiologian kiinnostuksen kohteena. (Erola ja Räsänen 2014a, 11.) Sosiologin tutkimat yhteiskunnalliset ilmiöt ovat erittäin moninaisia ja

tutkimuskenttä on laaja. Erityisesti sosiologia on kiinnostunut moderneissa yhteiskunnissa elävistä ihmisyhteisöistä. Näitä tutkii myös moni muu tieteenala, jotka ovat lähellä sosiologiaa. Erityisesti sosiaali-antropologia, taloustiede, valtio-oppi, kasvatustiede, psykologia ja historia ovat nähty sosiologian lähitieteinä. Sosiologian ominaispiirre verrattuna näihin aloihin on se, että ilmiöitä tutkitaan sosiaalisina. (Erola ja Räsänen 2014a, 16–19.) Suomessa sosiologiaa opetetaan Helsingin yliopistossa, Jyväskylän yliopistossa, Itä-Suomen yliopistossa, Lapin yliopistossa, Tampereen yliopistossa, Turun yliopistossa ja Åbo Akademiassa.

Sosiologia on empiirinen eli kokemuseräinen tiede. Tällä tarkoitetaan sitä, että sosiologian tutkimuksessa keskeistä ovat havainnot ympäröivästä maailmasta. (Erola ja Räsänen 2014a, 14.) Sosiologiselle tutkimukselle tyypillistä on, että muodostettuja teorioita ja olettamuksia koetellaan analysoimalla ihmisiä koskevia havaintoja tutkimusaineistojen avulla. Sosiologiassa tehdään sekä kvantitatiivista että kvalitatiivista tutkimusta. Kvantitatiivista aineistoa kerätään tyypillisesti kyselytutkimusten avulla. Myös muuhun tarkoitukseen kerättyä määrällistä aineistoa hyödynnetään; esimerkiksi julkishallinnon käyttöön tarkoitetut rekisteritiedot ovat tyypillisiä sosiologian kvantitatiivisia tutkimusaineistoja. Tyypillisiä sosiologian kvalitatiivisia aineistoja ovat haastatteluaineistot, erilaiset teksti- ja kuva-aineistot sekä havainnointi-aineistot. (Erola ja Räsänen 2014b, 67–70.)

Informaatiotutkimus on tieteenala, joka tarkastelee tiedon hankintatapoja ja tiedon käyttöä sekä tiedon organisointia, välittämistä ja säilyttämistä. Alan tavoitteena on kehittää käsitteitä, menetelmiä ja järjestelmiä, joiden avulla tieto saadaan sitä tarvitseville helposti omaksuttavassa muodossa. Kirjastot ja arkistot ovat informaatiotutkimukselle läheisiä instituutioita, vaikka niitä ei tieteenalan määritelmässä mainita. Informaatiotutkimus oli aikaisemmin enemmän instituutiokeskeistä ja viime vuosikymmeninä on siirrytty enemmän ihmisen informaatiokäyttäytymistä painottavaan lähestymistapaan. Näiden näkökulmien välinen jännite näkyy alalla edelleen. (Mäkinen 2010, 1–4.) Informaatiotutkimuksen

osa-alueita ovat mm. tiedonhankinnan tutkimus, tiedonhaun tutkimus, tietohallinnon tutkimus sekä arkisto- ja asiakirjahallinnan tutkimus.

Informaatiotutkimus on soveltava ala, jolle tyypillistä on monet erilaiset lähestymistavat ja tutkimusperinteet. Informaatiotutkimuksessa yhdistyvät monet eri yhteiskunta- ja hallintotieteelliset, humanistiset ja teknologiset tutkimusmenetelmät ja traditiot. 1970-luvulta lähtien yhteiskuntatieteet on nähty alalle läheisinä; erityisesti tiedonhankintatutkimus on lähellä yhteiskuntatieteitä. Myös tietojenkäsittelytieteet ovat lähellä informaatiotutkimusta, erityisesti tiedonhaun tutkimuksessa käytetään tietojenkäsittelytieteiden menetelmiä. Myös journalistiikassa, psykologiassa ja kasvatustieteissä voidaan nähdä yhtäläisyyksiä informaatiotutkimuksen kanssa. (Mäkinen 2010, 2–4.) Suomessa informaatiotutkimus kuuluu Tampereen yliopistossa informaatioteknologian ja viestinnän tiedekuntaan, Oulun yliopistossa humanistiseen tiedekuntaan ja Åbo Akademiassa yhteiskuntatieteiden ja kauppatieteiden tiedekuntaan.

3 AVOIN TIEDE JA TUTKIMUSAINEISTOT

Tässä työssä tarkastellaan informaatiotutkimuksen ja sosiologian tutkijoiden näkemyksiä *tutkimusaineistojen avoimuudesta*. Avoimilla tutkimusaineistoilla tarkoitetaan tutkimusaineistoja, jotka on tehty julkisesti saatavaksi ja uudelleen käytettäväksi mahdollisimman vähäisin rajoituksin (Higman, Bangert ja Jones 2019, 3). *Tutkimusaineistojen jakaminen* voidaan määritellä aineistojen järjestämiseksi ja säilyttämiseksi avoimesti saataville niin, että tarkoituksena on mahdollistaa aineistojen jatkokäyttö (Chawinga ja Zinn 2019, 110). Suomenkielisessä kirjallisuudessa ja tässä työssä käytetään myös ilmaisua *aineistojen avaaminen*, joka on aineiston jakamisen synonyymi. *Jatkokäytöllä* tarkoitetaan toisen tutkijan tai tutkijaryhmän keräämän aineiston hyödyntämistä tutkimuksessa. Tutkija voi myös itse käyttää keräämäänsä aineistoa uudessa tutkimuksessa. (Ketola 2014, 7.) Tutkimusaineistojen avoimuus on osa *avoimen tieteen* kehitystä, jota tarkastellaan luvussa 3.1. Luvussa 3.2 esitellään avoimen tieteen keskeisiä toimijoita ja ajankohtaisia hankkeita.

3.1 Avoimen tieteen taustaa

Avoimuus ja informaation jakaminen ovat perusta tieteen edistymiselle (The national academies of Sciences, Engineering and Medicine 2018, 1). Niin sanottu ensimmäinen tieteellinen vallankumous sijoittuu 1500–1700 -luvulle. Silloin kehittyivät ensimmäiset tieteenalat ja tieteellinen julkaisutoiminta. (Bartling ja Friesike 2014, 7.) Tieteellinen julkaisutoiminta sai alkunsa siis aikana, jolloin painettu teksti oli ainut keino viestiä. Tieteellisten aikakauslehtien synty mahdollisti tutkijoille informaation jakamisen ja kommunikoinnin ajasta ja paikasta riippumatta (The national academies of Sciences, Engineering and Medicine, 2018, 1). Tutkijoiden keskuudessa muotoutui ajatus siitä, että tutkimusten avoimuus muille tutkijoille on yleinen etu. (Bartling ja Friesike 2014, 7–8.)

Viimeisen kahdenkymmenen vuoden ajan tieteessä on tapahtunut merkittävä muutos kohti avoimempaa kulttuuria (Rentier 2016, 155). Muutos kohti avoimempaa tiedettä on niin suuri, että voidaan puhua meneillään olevasta toisesta tieteellisestä vallankumouksesta (Bartling ja Friesike 2014, 7–12). Muutokseen on kaksi keskeistä syytä. Ensimmäinen syy on tieteelliseen julkaisuun pääsyn hintojen nousu monilla tieteenaloilla. (Rentier 2016, 155.) Tiedekustantajien jatkuvasti nousseet tilaus- ja lisenssimaksut alkoivat viedä yhä suuremman osan tieteellisten kirjastojen budjeteista (Holopainen ja Koskinen 2017, 43). Kohoavat maksut ovat luoneet kysyntää vaihtoehtoisille julkaisuväylille. Toinen syy on internetin kehitys ja sen mahdollistamat uudet julkaisutavat. (Rentier 2016, 155.) Nykyisen informaatioteknologian jatkuva kehitys vie tutkimusta kohti uutta avoimen tieteen ympäristöä (the national academies of Sciences, Engineering and Medicine, 2018, 11).

Avoimella tieteellä tarkoitetaan tieteellistä kulttuuria, joka tähtää tutkimustulosten nopeaan ja tehokkaaseen jakamiseen laajalle yleisölle ilman maksua (The national academies of Sciences, Engineering and Medicine 2018, 1; Bartling ja Friesike 2014, 10–11). Keskustelu avoimesta tieteestä liittyi pitkään lähinnä tieteellisiin julkaisuihin (Lilja 2017). Avoimen julkaisun yhteydessä puhutaan vihreästä ja kultaisesta tiestä. Vihreällä tiellä tarkoitetaan sitä, että artikkeli tallennetaan kirjoittajan tai hänen edustajansa julkaisuarkistoon. Kustantaja voi määrittää artikkelille viiveajan (embargo), jonka jälkeen artikkeli on luettavissa. Kultainen tie taas tarkoittaa sitä, että artikkeli julkaistaan suoraan vapaasti luettavaksi open access -lehdessä. Tämä vaihtoehto on usein tekijälle maksullinen. (Gued ja Ramjoué 2014, 12.) Kultaisen ja vihreän tien lisäksi puhutaan niin sanotusta hybridimallista, jossa artikkeli avataan maksua vastaan perinteisessä tilausmaksua perivässä lehdessä (Holopainen ja Koskinen 2017, 43).

Vähitellen vaatimus avoimuudesta on alkanut koskea julkaisujen lisäksi koko tieteen prosessia. Avoimen tieteen osa-alueista julkaisujen lisäksi

tutkimusaineistojen avoimuus on noussut vahvasti esille. (Lilja 2017, 1–2.) Julkaisujen ja tutkimusaineiston lisäksi avoimen tieteen osa-alueita ovat oppimisen ja oppimisympäristöjen avoimuus, saavutettavuus ja esteettömyys sekä tiedeyhteisöjen avoin toimintakulttuuri (Avoimen tieteen kansallinen koordinaatio 2019a). Avoin tiede on noussut merkittäväksi tavaksi edistää tutkimusta ja lisätä tieteen vaikuttavuutta yhteiskunnassa. Avoin tiede on myös suuri kulttuurinmuutos tiedeyhteisölle: se muokkaa tutkimuksen tekoa, tutkimuksen rahoitusta, tulosten jakamista ja johtaa uudenlaiseen avoimeen toimintakulttuuriin. (Avoimen tieteen kansallinen koordinaatio 2020c.)

Avoimeen tieteeseen liittyvä termi *open access* viittaa tieteellisen tiedon vapaaseen ja maksuttomaan saatavuuteen (Sitek ja Bertelmann 2014, 140). Open access -julkaisut ovat internetissä kaikkien saatavissa ja mahdollisuuksien mukaan uudelleen käytettävissä ilmaiseksi. Tutkimusaineistojen yhteydessä open access viittaa tutkimusaineistoon, joka on saavutettavissa, louhittavissa, hyödynnettävissä ja uudelleen käytettävissä ilman maksua. (Gued ja Ramjoué 2014, 11–13.)

Avoimen tieteen käytännöillä on todettu tutkimuksissa oleva paljon hyötyjä. Avoin tiede tekee tutkimuksesta tehokkaampaa, sillä avoin saatavuus vähentää tieteen teon kustannuksia. Lisäksi avoin tiede tekee saatavilla olevasta tietomäärästä suuremman, mikä auttaa tutkimuksia onnistumaan paremmin. Avoimen tieteen käytännöt myös edistävät yhteistyötä tieteenalan sisällä mutta myös tieteenalojen välillä, mikä edistää uusien ideoiden syntymistä. Avoin tiede tekee myös tieteestä demokraattisempaa. Kustannusten väheneminen on tärkeää erityisesti niille, joilla ei ole varaa maksaa pääsystä informaatioon. Avoimen tieteen käytäntöjen ansiosta tieteellinen tieto tulee saataville kaikille ihmisille, ei vain tutkijoille. Avoimen tieteen on myös nähty lisäävän tutkijoiden ja yhteisön välistä yhteistyötä, mikä edistää tieteen soveltamista käytännön ongelmiin. (Arza ja Fressoli 2017, 465–439.) Avoin tiede tekee myös tutkimuksesta läpinäkyvämpää (Tenopir ym. 2020, 2). Avoimeen tieteeseen liittyy myös ongelmia.

Esimerkiksi mahdollisuus sensitiivisen tiedon leviämiseen ja tunnustuksen saaminen tieteellisestä työstä on nähty avoimen tieteen yhteydessä haasteellisena. (Jomier 2017, 365.)

3.2 Avoimen tieteen hankkeita ja toimijoita

Sekä tutkimusaineistojen että julkaisujen avoimuuteen liittyviä hankkeita on käynnistetty sekä Suomessa että kansainvälisesti. Avoimuuden määrässä on kuitenkin vielä suuria eroja maiden välillä. (Holopainen ja Koskinen 2017, 43.)

Euroopan komissio on edistänyt avointa tiedettä vuodesta 2006 lähtien, keskittyen ensin avoimeen julkaisuun ja myöhemmin myös avoimiin tutkimusaineistoihin. Euroopan komission tutkimuksen ja innovoinnin Horizon 2020 -hankkeessa määritellään unionin avoimen tieteen politiikka. Ohjelman mukaan tutkimustulosten ja tutkimusaineistojen avoin saatavuus lisää tieteen laadukkuutta ja vaikuttavuutta, nopeuttaa innovaatioita ja lisää tieteellisen prosessin läpinäkyvyyttä. Hankkeen keskeinen periaate on, että julkisrahoitteisten eurooppalaisten tutkimustulosten tulisi olla kaikkien kiinnostuneiden saatavilla. Hankkeeseen kuuluu myös avoimen tutkimusaineiston pilottiohjelma, jonka tavoitteena on lisätä tutkimusaineistojen avaamista rahoittamalla projekteja. Euroopan unionin avoimen tieteen politiikka on edesauttanut myös muita rahoittajia ja organisaatioita kehittämään omaa avoimen tieteen politiikkaansa. (Gued ja Ramjoué 2014, 11–13.)

Avointa tiedettä edistää myös syyskuussa 2018 perustettu hanke Plan S. Hankkeen takana on cOAlition S, jonka perusti ryhmä tutkimusrahoittajia Euroopan komission ja Euroopan tutkimusneuvoston tuella. Plan S:n tavoite on saada kaikki sekä julkisin että yksityisin varoin rahoitetut tutkimukset välittömästi avoimesti saataviksi ilman embargo-aikaa vuoteen 2021 mennessä. (European Science Foundation 2020.)

Suomessa opetus- ja kulttuuriministeriö järjesti Avoin tiede ja tutkimus -hankkeen vuosina 2014 - 2017 tiedon saatavuuden ja avoimuuden edistämiseksi. ATT-hanke perustui ministeriöiden, korkeakoulujen, tutkimuslaitosten ja tutkimusrahoittajien yhteistyöhön. Hankkeen aikana luotiin mm. avoimen tieteen tiekartta ja avoimen tieteen käsikirja ja pidettiin koulutuksia ja osallistuttiin kansainväliseen yhteistyöhön. (Forsström ja Kuttilainen 2019.) Tällä hetkellä avoimen tieteen koordinaatiosta Suomessa vastaa Tieteellisten seurain valtuuskunta Opetus- ja kulttuuriministeriön rahoittamana. Koordinaatio perustuu tutkimusyhteisön yhteistyölle: mukana ovat yliopistot, ammattikorkeakoulut, tutkimuslaitokset, rahoittajat, kirjastot ja arkistot. Koordinaation tavoitteena on lisätä yhteistyötä ja tietoisuutta avoimen tieteen mahdollisuuksista, haasteista ja ratkaisuista ja lisätä keskustelua avoimen tieteen tavoitteista ja keinoista. (Avoimen tieteen kansallinen koordinaatio 2019b.)

Monet näkyvät tutkimusten rahoittajat ovat ottaneet avoimen tieteen osaksi politiikkaansa. Rahoituksen edellytyksenä voi olla julkaisujen ja tutkimusaineistojen avoimuus. (Eisfeld-Reschke ym. 2014.) Erityisesti 2010-luvulla monissa maissa tärkeimmät kansalliset tutkimusrahoittajat ovat alkaneet vaatia julkaisujen avaamista ja suosittelemaan myös tutkimusaineistojen avoimuutta (Holopainen ja Koskinen 2017, 43). Esimerkiksi Suomessa Suomen Akatemia on sitoutunut avoimen tieteen periaatteiden ja käytäntöjen edistämiseen. Akatemian rahoituksen edellytyksenä on sitoutuminen tutkimustuloksista kertovien tiedejulkaisujen avoimeen saatavuuteen ja hankkeen tuottamien tutkimusaineistojen avaamiseen. Avoimuuden aste voi kuitenkin vaihdella perustelluista syistä. (Suomen Akatemia 2019.)

Yliopistot ja tutkimuslaitokset ovat myös kehittäneet omia avoimen tieteen linjauksiaan. Instituutioiden tasolla tehtävä avoimuuden politiikka auttaa avoimuuden kulttuurin toteutumisessa (Patel 2016). Tampereen yliopiston avoimen tieteen linjauksessa määritellään avoimuuden yleiset periaatteet sekä julkaisujen ja

tutkimusaineistojen avoimuuden periaatteet (Tampereen yliopisto n.d.). Linjauksen mukaisesti Tampereen yliopisto arvostaa tieteen avoimuutta ja tukee sen toteutumista. Tutkimusaineistojen osalta linjauksessa todetaan, että siltä osin kuin on mahdollista, tutkimusaineistot tallennetaan ja avataan kansallisissa tai kansainvälisissä arkistoissa tai tallennuspalveluissa. Linjauksen mukaan tutkimusaineistojen avoimuuden taso voi kuitenkin vaihdella perustelluista syistä ja esimerkiksi kaupallinen hyödyntäminen ja oikeuksien turvaaminen otetaan huomioon aineistoja avattaessa. (Tampereen yliopisto 2018a, 1–3.) Avoimen tieteen linjauksia toteutetaan ammattikorkeakoulun ja yliopiston toimenpideohjelmien avulla (Tampereen yliopisto n.d.). Toimenpideohjelman mukaisesti yliopisto esimerkiksi tarjoaa tukea ja neuvontaa aineiston elinkaaren kaikkiin vaiheisiin (Tampereen yliopisto 2018b, 2).

4 TUTKIMUSAINEISTOJEN KÄYTÄNTÖJÄ

AVOIMUUDEN

Tässä luvussa tarkastellaan tutkimusaineistojen avoimuuden käytäntöjä. *Tutkimusaineistojen hallinta* ja *aineiston elinkaari* ovat keskeisessä osassa tutkimusaineistojen avoimessa jakamisessa, ja tätä tarkastellaan tarkemmin luvussa 4.1. Luku 4.2 keskittyy aineistojen löydettävyyteen ja *FAIR-data* periaatteisiin. Tutkimusaineistojen avoimuuteen liittyy erilaisia toimijoita. Luvussa 4.3. esitellään *data-arkistojen*, *tieteellisten lehtien* ja *tieteellisten kirjastojen* roolia aineistojen avoimuudessa.

4.1 Tutkimusaineistojen hallinta ja elinkaari

Tutkimusaineistojen hallinta tarkoittaa aineistojen luomista, löytämistä, järjestämistä, säilyttämistä ja jakamista tutkimusprosessin aikana (Cox ja Verbaan 2018, 4). Tutkimusaineistojen avoimuus edellyttää suunnitelmallisuutta ja hyvää aineistojen hallintaa (Cox ja Verbaan 2018, 53; Tenopir ym. 2020, 2). Jo tutkimuksen alussa tehdyt ratkaisut aineistojen hallinnassa vaikuttavat siihen, onko aineistoja mahdollista avata myöhemmin (Higman ym. 2019, 2). Aineistohallinta on noussut vahvasti keskusteluun erityisesti 2010-luvulla. Tehokkaasta aineistohallinnasta on hyötyä tutkijalle, vaikka aineistoa ei olisikaan tarkoitus jakaa. (Cox ja Verbaan 2018, 41, 53.)

Tutkimusaineistojen hallintaa voi hahmottaa aineistojen elinkaarta tarkastelemalla. Tutkimusaineistojen elinkaaresta on olemassa useita malleja. Cox ja Verbaan (2018, 26) esittävät UK Data Archivea lainaten tutkimusaineistojen elinkaaren kuusivaiheisena. Ensimmäisessä vaiheessa aineisto luodaan. Tämä vaihe sisältää aineiston keräämisen suunnittelua ja toteutusta tai valmiin aineiston paikallistamista ja aineiston dokumentointia. Toisessa vaiheessa aineisto siivotaan ja valmistellaan

analyysiin sopivaksi. Kolmannessa vaiheessa aineisto analysoidaan ja analyysi julkaistaan. Neljännessä vaiheessa aineisto saatetaan oikeaan formaattiin ja dokumentoidaan säilyttämistä varten. Viidennessä vaiheessa aineisto tallennetaan niin että se on saatavilla ja löydettävissä. Aineiston elinkaaren kuudes vaihe on aineiston jatkokäyttö muissa yhteyksissä.

Borg & Kuula (2007, 14) puolestaan esittävät julkisrahoitteisen tutkimusaineiston elinkaaren DDI Alliancen (2004) aineiston elinkaarimallia mukaillen kymmenvaiheisena:

1. Tutkimussuunnitelma
2. Eettinen arviointi ja luvat
3. Rahoitushakemus
4. Rahoittaja / rahoituspäätös
5. Aineiston keruun valmistelu/keruu
6. Ensikäyttö
7. Arkistointi ja julkaiseminen
8. Jatkokäytön tuki
9. Jatkokäyttö
10. Julkaisut ja aineistot (tietokannat)

Tieteellinen tutkimussuunnitelma on tutkimusta ohjaava dokumentti, jossa esitetään tutkimuksen tavoitteet, aineistot ja menetelmät uuden tiedon tuottamisen näkökulmasta. Tutkimusaineistojen elinkaari voidaan usein hahmottaa jo tutkimussuunnitelman laatimisen aikana. Tutkimussuunnitelmaa seuraa eettinen ennakoarviointi, joka Suomessa on lakisääteinen velvoite esimerkiksi ihmisiin kohdistuvassa lääketieteellisessä tutkimuksessa tai eläinkokeissa. Seuraavana aineistojen elinkaarimallissa on tutkimusrahoituksen hakeminen ja rahoituspäätös. Tutkimusrahoittajat edellyttävät hyvän tutkimusetiikan noudattamista, ja aineistojen kohdalla se tarkoittaa eettistä tapaa kerätä ja käyttää tutkimusaineistoja. (Borg &

Kuula 2007, 15–17.) Nykyään yhä useammin tutkimusten rahoittajat vaativat myös mahdollisuutta tutkimusaineiston jatkokäyttöön (Pampell ja Dallmeir-Tiessen 2014, 213).

Rahoituspäätöstä seuraa aineiston keruu. Tutkimusaineiston laatu määrittää sen, mitä aineiston keruun valmisteluun ja toteutukseen liittyy. Jo keruun alkuvaiheessa tehdyt toimenpiteet ovat ratkaisevia aineistojen jatkokäytön kannalta. Tehdyt sopimukset, aineiston keruun onnistuminen ja toimenpiteiden kirjaaminen vaikuttavat siihen, onko aineistoa mahdollista käyttää uudelleen myöhemmin. Aineiston keruun jälkeen vuorossa on aineiston ensikäyttö, jolla tarkoitetaan käyttöä tutkimussuunnitelmassa esitetystä tarkoituksessa. (Borg & Kuula 2007, 19–21.)

Usein aineiston elinkaari päättyy jo ensikäyttöön. Jos aineistoa halutaan julkaista avoimesti, ensikäyttöä seuraa aineiston arkistointi ja julkaiseminen. Julkaisun voi tehdä esimerkiksi jossain tutkimusaineistoja arkistoivassa organisaatiossa. Julkaisuprosessiin liittyy aineistosta riippuen kuvailutietojen lisäämistä, anonymisointia ja kvantitatiivisen aineiston kohdalla muuttujien tarkistusta. Kun aineisto on julkaistu, sitä on mahdollista käyttää uudestaan esimerkiksi tutkimukseen tai opetukseen. Jatkokäytön toteutuminen vaatii myös tukea. Jatkokäytön tukeen liittyy kysymykset siitä, mihin aineisto on arkistoitu, miten sitä voi käyttää, millaisia ovat jatkokäytön neuvontapalvelut ja onko jatkokäyttö maksullista vai ei. Ensi- ja jatkokäytön tuottamat julkaisut tallennetaan usein erilaisiin julkaisu- ja hanketietokantoihin, mutta ne sisältävät harvoin tietoa taustalla olevasta aineistosta. Julkaisujen taustalla olevien tutkimusaineistojen helppo löytyminen olisi kuitenkin uuden tiedon tuottamisen näkökulmasta oleellista. (Borg ja Kuula 2007, 21–25.)

Yhteiskuntatieteellinen tietoarkisto (2017) esittää aineiston elinkaarimallissaan tutkimussuunnitelman yhteydessä *aineistonhallintasuunnitelman* tekemisen. Aineistojen jatkokäyttö on helpompaa, jos aineiston hallinta on suunnitelmallista.

Aineistohallintasuunnitelma on dokumentti, joka kattaa koko aineiston elinkaaren. Siinä kuvataan, miten aineisto hankitaan ja miten sitä säilytetään tutkimuksen aikana ja tutkimuksen jälkeen. Monet tutkimusrahoittajat ovat huomanneet suunnitelmallisuuden merkityksen aineiston hallinnassa ja vaativat aineistohallintasuunnitelmaa rahoitushakemuksen liitteeksi. (Enwald, Huuskonen ja Kuusniemi 2016, 29; Strasser 2015, 2.) Tampereen yliopiston (2018, 3) avoimen tieteen linjauksessa todetaan, että tutkija laatii suunnitteluvaiheessa aineistohallintasuunnitelman riippumatta siitä, onko aineisto mahdollista avata.

Tyypillisessä aineistohallintasuunnitelmassa on viisi keskeistä kohtaa. Ensimmäinen aineistohallintasuunnitelmassa täytyy määrittellä, millaista aineistoa tutkimuksen aikana syntyy. Toiseksi kuvataan mitä standardeja esimerkiksi *metatietojen* luomisessa käytetään. Kolmanneksi kuvataan aineiston keräämiseen liittyvät säännöt ja velvoitteet, esimerkiksi miten aineistoon pääsee käsiksi jälkepäin, onko aineiston avaamisen suhteen vaatimuksia ja miten sensitiivistä aineistoa käsitellään. Aineistohallintasuunnitelmassa tulee myös olla selvitys siitä, miten aineiston pitkäaikaissäilytys toteutetaan. Lopuksi suunnitelmassa tulee kuvata, mitä resursseja (aikaa, välineitä, rahaa) suunnitelman toteuttaminen vaatii. (Strasser 2015, 2–4.) Suomessa on otettu käyttöön kaikille tutkimusorganisaatioille yhteinen työkalu DMP Tuuli aineistohallintasuunnitelmien tekemiseen (Enwald ym. 2016, 30).

4.2 Aineistojen löydettävyyden ja FAIR-data periaatteet

Tutkimusaineistojen löydettävyyden ja jatkokäytön kannalta on oleellista, että aineistot on kuvailtu huolellisesti. Aineistojen kuvailutietoja kutsutaan metatiedoiksi. Epävirallisia metatietoja ovat esimerkiksi tutkimuksen aikana tehdyt kuvailutiedot ja muistiinpanot, jotka eivät noudata mitään metatietoformaattia. Yleensä aineistoja julkaistaessa metatiedot tehdään jonkin standardin mukaisesti. Standardin mukaan tehty metatieto esittää aineiston tiedot tietyssä muodossa ja

erityistä sanastoa käyttäen. Erilaisia metatietostandardeja on useita, ja standardin valintaan vaikuttaa esimerkiksi tieteenala ja käytetty data-arkisto. Standardoidun metatiedon käyttö on hyödyllistä aineiston löydettävyyden, koneluettavuuden ja aineistojen yhteen toimivuuden kannalta. (Strasser 2015, 7.) Aineistojen avaamisen kannalta tärkeää olisi ohjeistus tutkijoille metatietojen tuottamiseen (Lilja 2017, 4).

Tutkimusaineistojen yhteydessä puhutaan usein avoimuuden lisäksi FAIR-periaatteista. Termi sai alkunsa vuonna 2014, ja periaatteet julkaistiin vuonna 2016 (FORCE11 2017). Periaatteet ovat saavuttaneet laajan hyväksynnän (Higman ym. 2019, 2; Cox ja Verbaan 2018, 50; Tenopir ym. 2020, 2). FAIR-periaatteiden on tarkoitus tehdä aineistot löydettäviksi, saavutettaviksi, yhteentoimiviksi ja uudelleenkäytettäviksi (Fairdata.fi 2020). Termin alle kuuluu 15 alakohtaa, jotka ohjaavat julkaisijan ja muiden sidosryhmien toimintaa. FAIR-periaatteet eivät tarkoita samaa kuin aineistojen avoimuus; tutkimusaineisto voi esimerkiksi noudattaa FAIR-periaatteita niin, että vain aineiston metatiedot ovat saatavilla (Higman ym. 2019, 2–3).

FAIR-periaatteissa löydettävyyden osalta määritellään, että aineistolle annetaan ainutkertainen standardoitu pysyvä tunniste. Tutkimusaineistojen tyypillisimpiä tunnisteita ovat DOI, URN ja Handle -tunnistejärjestelmät. Aineistolla tulee olla myös ylläpitäjä, joka huolehtii siitä, että aineisto pysyy verkossa ja tunniste osoittaa siihen. Aineistolla tulisi olla kattavat metatiedot, joista tulee ilmi aineiston pysyvä tunniste. Lisäksi aineistot kuvaillaan ja rekisteröidään johonkin hakupalveluun. (Fairdata.fi 2020.)

Saavutettavuuden osalta FAIR-periaatteissa määritellään, että joko aineisto tai sen metatiedot ovat haettavissa avointa, maksutonta ja yleisesti käyttöön otettavaa yhteyskäytäntöä noudattaen. Metatiedot tulisi olla saavutettavissa, vaikka aineistoa ei enää olisi. Jos aineistot eivät ole avoimesti saatavilla, tulisi yhteyskäytännön tukea tunnistamista ja valtuuttamista, eli mahdollistaa esimerkiksi käyttöluvan hakemisen.

(Fairdata.fi 2020.) Yhteentoimivuus FAIR-periaatteissa tarkoittaa sitä, että sekä aineisto että metatiedot on tuotettu standardoitua sekä ihmis- että koneluettavaa kieltä käyttäen. Uudelleenkäytettävyydellä taas viitataan siihen, että aineisto on huolellisesti dokumentoitu. Aineiston tulee olla tiedeyhteisön hyväksymää ja kuvailusta pitää tulla ilmi mm. miten aineisto on kerätty ja käsitelty ja miten siihen viitataan. (Fairdata.fi 2020; Cox ja Verbaan 2018, 50.)

4.3 Data-arkistojen, tieteellisten lehtien ja tieteellisten kirjastojen rooli

Tutkimusaineistojen avaamiseen on kolme keskeistä tapaa. Ensimmäinen on julkaista aineisto itsenäisenä yksikkönä aineistoja säilyttävässä data-arkistossa. Toinen mahdollisuus on julkaista aineistot tekstiaineistona datalehdessä. Kolmas vaihtoehto on julkaista aineisto tutkimusjulkaisun yhteydessä. (Pampell ja Dallmeier-Tiessen 2014, 219.)

Tutkimusaineistojen avoimuutta on pyritty tukemaan luomalla aineistoja säilyttäviä, välittäviä ja dokumentoivia organisaatioita. Nämä organisaatiot toimivat kolmantena osapuolena aineistojen luojien ja jatkokäyttäjien välillä. (Borg ja Kuula 2007, 10.) Data-arkistoja on sekä kansallisia että kansainvälisiä, ja osa niistä on tieteenalakohtaisia ja osa kaikki tieteenalat kattavia. Suomessa keskeinen organisaatio on vuodesta 1999 asti toiminut Yhteiskuntatieteellinen tietoarkisto (FSD) (Borg ja Kuula 2007, 10). Yhteiskuntatieteellinen tietoarkisto (n.d) arkistoi ja välittää sähköisiä tutkimusaineistoja tutkimukseen, opetukseen ja opiskeluun. Se on keskittynyt erityisesti yhteiskuntatieteellisten, humanististen alojen, kasvatustieteiden ja terveystieteiden tutkimusaineistoihin. Muita suomalaisia tallennuspalveluja ovat esimerkiksi avointen tutkimusaineistojen julkaisupalvelu AVAA ja kieliaineistoja tallentava FINCLARIN. Kansainvälisiä data-arkistoja on paljon; keskeisiä ovat esimerkiksi Zenodo, EUDAT, Dryad, Figshare ja GitHub (Enwald 2018, 31).

Tieteelliset lehdet ovat keskeinen toimija tutkimusten julkaisuissa. Viime vuosina monet tieteelliset lehdet eri tieteenaloilla ovat ottaneet tutkimusaineistojen avoimuuden osaksi politiikkaansa (Chawinga ja Zinn 2019, 117). Tieteellisiin lehtiin on alkanut kohdistua odotuksia liittyen tutkimusaineistojen huomioimiseen artikkelien valinnassa ja viittauskäytännöistä ohjeistamisessa. Tieteellisten lehtien käytännöt ja vaatimukset tutkimusaineistojen suhteen ovat kuitenkin hyvin vaihtelevia. (Lilja 2018, 2.) Tieteellisten lehtien suhtautumisen aineistoihin voi jakaa neljään tasoon. Nollatason lehti on aineistojen avaamisen ja niihin viittaamisen suhteen passiivinen. Ykköstason lehti ohjeistaa viittaamiseen ja vaatii kirjoittajaa ilmaisemaan, onko aineisto käytössä. Kakkostasolla aineistojen tallentaminen ja avaaminen silloin, kun se on mahdollista, on ehto julkaisulle. Kolmannella tasolla lehti vaatii läpinäkyvyyttä tutkimussuunnitelmasta lähtien ja ei julkaise artikkeleita, jos aineistoa ei ole avattu. (Center of Open Science, ks. Lilja 2017, 4.) Käytännössä kolmas taso ei kuitenkaan vielä juuri toteudu (Lilja 2017, 5). On myös olemassa erityisiä tutkimusaineistoihin keskittyviä datalehtiä, joissa on mm. tutkimusaineistojen metatietoja ja linkityksiä aineistoihin. Näiden lehtien määrä on kasvussa maailmalla. (Candela, Castelli, Manghi ja Tani 2015.)

Myös tieteellisillä kirjastoilla on perinteisesti ollut merkittävä rooli tieteelliseen tietoon pääsyssä (Tenopir ym. 2017, 25). Viime vuosina tieteellisten kirjastojen rooli on laajentunut tieteellisen informaation hallintaan, tutkimusaineistojen hallintaan, rinnakkaistallentamiseen ja metatietoon liittyvään osaamiseen (Forsström, Kuttilainen ja Väisänen 2015, 1). Erityisesti tutkimusaineistojen hallinnan merkityksen kasvu on noussut maailmanlaajuisesti puheenaiheeksi tieteellisissä kirjastoissa. Tutkimusaineistopalvelujen tarve akateemisissa yhteisöissä on saanut kirjastot vastaamaan tarpeeseen ja tarjoamaan palveluja aineistojen hallintaan. Tieteellisten kirjastojen työntekijät ovat yhä enemmän mukana koko tutkimusprosessissa, niin aineiston keräämisessä, julkaisussa kuin säilyttämisessäkin. (Tenopir ym. 2015, 2, 25.)

Kuitenkin kirjastojen aineistojen hallintaan liittyvät palvelut ovat vielä varsin uusia (Late ja Talja 2016, 45).

Eurooppalaisten yliopistokirjastojen neuvosto (LIBER1) selvitti vuonna 2016 eurooppalaisten tieteellisten kirjastojen tarjoamia tutkimusaineistopalveluja. Kartoituksessa kysyttiin kirjastojen johtajilta, mitä tutkimusaineistoihin liittyviä palveluja kirjasto tällä hetkellä tarjoaa ja mitä palveluja on suunnitelmassa tarjota tulevaisuudessa. Suurin osa kirjastoista tarjoaa konsultatiivisia palveluja tai aikoo tarjota niitä lähitulevaisuudessa. Yli puolet vastanneista kirjastoista käy keskustelua tutkimusaineistoon liittyvistä asioista organisaatiossaan ja yli 40 % antaa neuvontaa aineistonhallintasuunnitelmiin ja metatietoihin liittyen. Teknisten tutkimusaineistopalvelujen tarjoaminen ei ole yhtä yleistä. Kuitenkin lähes 40 % kirjastoista antaa teknistä tukea data-arkistojen käyttöön ja 40 % ylläpitää data-arkistoa. Suurin osa (67 %) vastaajista näki, että kirjastojen on tarjottava tutkimusaineistopalveluja ja oltava edelläkävijä niiden tarjoamisessa. (Late ja Talja 2016; Tenopir ym. 2017.) Monien muiden eurooppalaisten ja suomalaisten kirjastojen tavoin myös Tampereen yliopiston kirjasto on kehittänyt tutkimusaineistopalvelujaan viime vuosina (Huusko ja Toikkonen 2018, 47).

5 NÄKÖKULMIA AINEISTOJEN AVOIMUUTEEN

Tutkimusaineistojen avoimuus on ajankohtainen aihe ja viime vuosina on tehty melko paljon tutkimusta aiheesta. Tutkimuksissa on kartoitettu aineiston avoimuuden hyötyjä ja esteitä, ja näitä käsitellään luvussa 5.1. Tutkijoiden näkemyksistä ja käytännöistä tutkimusaineistojen avoimuuden suhteen on tehty useita kansainvälisiä ja kansallisia tutkimuksia. Luvussa 5.2 esitellään muutamia kansainvälisiä ja suomalaisia tutkimustuloksia aiheesta.

5.1 Hyödyt ja esteet

Tutkimusaineistojen avoimuus on aihe, josta keskustellaan laajasti eri tieteenaloilla. Kulttuuri aineistojen jakamisen suhteen vaihtelee tieteenalojen välillä. (Pampell ja Dallmeir-Tiessen 2014, 213.) On olemassa tieteenaloja, joilla tutkijat tekevät kiinteästi yhteistyötä ja aineistojen jakaminen on vakiintunut osa tutkimuksen tekemistä. Toisilla aloilla taas tutkijat pitävät aineiston tutkimuksen aikana itsellään ja mahdollinen aineiston jakaminen tapahtuu julkaisun jälkeen. On myös aloja, joilla tutkimusaineistoja ei juuri jaeta. (Berghman ym. 2017, 5.) Aineistojen jakaminen on yleistä joillain luonnontieteellisillä aloilla, esimerkiksi biolääketieteessä, maantieteessä ja meteorologiassa. Harvinaisempaa se taas on esimerkiksi yhteiskuntatieteissä, lääketieteessä ja psykologiassa. (Pampell ja Dallmeier-Tiessen 2014, 213, 126.) Jokaisella tieteenalalla on omat haasteensa ja ominaispiirteensä, jotka vaikuttavat aineistojen jakamiseen (Zuiderwijk ja Spiers 2019, 228).

Tutkimusaineistojen avaamisella on yleisesti ottaen useita hyötyjä. Ensinnäkin se mahdollistaa aineistojen uudelleenkäytön muissa yhteyksissä (Pampell ja Dallmeir-Tiessen 2014, 214). Samasta aineistosta voi esimerkiksi eri analyysimenetelmillä tulla

uusia löytöjä (Holopainen & Koskinen 2017, 45). Mahdollisuus aineistojen jatkokäyttöön vähentää päällekkäistä tiedonkeruuta (Borg ja Kuula 2007, 43). Aineistojen jakaminen säästää tutkijoiden aikaa, kun he voivat aineiston keräämisen sijaan keskittyä analysoimaan jo olemassa olevaa aineistoa (Patel 2016, 227). Ajan säästön lisäksi aineistojen jakamisesta tulee myös taloudellista hyötyä (Chawinga ja Zinn 2019, 113). Aineistojen jakaminen voi myös lisätä tiedeyhteisön välistä vuorovaikutusta ja yhteistyötä (Borg ja Kuula 2007, 43). Se voi esimerkiksi edistää uusien yhteistyökumppaneiden löytymistä, mistä voi hyötyä koko tutkimusyhteisö (Holopainen ja Koskinen 2017, 45). Niinpä aineistojen avoimuuden voi nähdä edistävän tiedettä (Chawinga ja Zinn 2019, 113).

Aineistojen avaaminen tekee tutkimuksesta läpinäkyvämpää (Patel 2016, 227). Se lisää tutkimuksen todennettavuutta ja mahdollistaa julkaisun johtopäätösten tarkastamisen. Avoimet aineistot antavat mahdollisuuden varmistaa, että tutkimus on toteutettu hyvän tieteellisen käytännön periaatteilla. (Holopainen ja Koskinen 2017, 45; Pampell ja Dallmeir-Tiessen 2014, 214.) Koska puutteellisesti kerättyä aineistoa ei haluta julkaista, voi avoimuus parantaa tutkimuksen laatua (Lilja 2017, 2).

Aineistojen avaaminen voi myös tuoda tutkijalle lisää näkyvyyttä ja siten hyödyttää tutkijaa itseään. Aineistojen kerääjät saavat tunnustusta aineistoihin viittaamisen kautta (Patel 2016, 227). Useissa tutkimuksissa on myös osoitettu, että julkaisut joiden yhteydessä tutkimusaineistot ovat avoimesti saatavilla, saavat enemmän viittauksia verrattuna vastaaviin julkaisuihin ilman saatavilla olevaa aineistoa (Piwovar ja Vision 2013).

Vaikka aineistojen jakamista pyritään edistämään monilta tahoilta ja siitä on todettu olevan hyötyjä, melko harvat tutkijat vielä käytännössä jakavat aineistojaan (Fecher, Friesike ja Hebing 2015, 1). Tutkimuksissa on löydetty monia tekijöitä, jotka estävät

aineistojen jakamista. Näiden syiden merkitys vaihtelee tieteenalojen välillä (Zuiderwijk ja Spiers 2019, 228).

Tutkimusaineistojen jakamiseen liittyy eettisiä ja lainsäädännöllisiä seikkoja, eikä kaikkea aineistoa ole mahdollista avata (Chawinga ja Zinn 2019, 118). Erityisesti ihmistieteissä keskeistä on huomioida henkilötietojen suoja (Ketola 2014, 3). Tutkimusaineistojen säilytystä ja jatkokäyttöä säätelee henkilötietolaki (Borg ja Kuula 2007, 27). Monissa tutkimuksissa korostetaan, että henkilökohtaisten tietojen käsittelyyn on oltava tutkittavan suostumus (Fecher ym. 2015, 15). Tutkittavalta tulisi myös pyytää suostumus aineistojen jakamiseen. Eettisesti ongelmallisia ovat tilanteet, joissa tutkittavat ovat lapsia tai nuoria, sillä heidän ei voida katsoa kykenevän tekemään päätöstä asiasta. (Chawinga ja Zinn 2019, 118). Tutkijat eivät myös välttämättä kaikissa tilanteissa halua pyytää suostumusta aineistojen jakamiseen, sillä se voi vaikuttaa negatiivisesti halukkuuteen osallistua tutkimukseen (Higman ym. 2019, 6).

Erityisesti arkaluonteista tietoa sisältävien aineistojen arkistoinnissa on huolehdittava siitä, että tutkittavien tietosuoja ei vaarannu. Aineistot tuleekin anonymisoida ennen tallentamista. (Fält ja Okuloff 2016, 12.) Anonymisoinnilla tarkoitetaan aineiston käsittelyä niin, että yksittäinen henkilö ei ole tunnistettavissa (Yhteiskuntatieteellinen tietoarkisto 2019). Kuitenkin joskus anonymisointi voi hankaloittaa jatkokäyttöä tai jopa muuttaa aineiston käyttökelvottomaksi (Fält ja Okuloff 2016, 12). Avoimuuden ja yksityisyyden välinen rajanveto ei ole aina yksinkertaista (Strasser 2015, 11).

Lainsäädännöllisiä tekijöitä aineistojen avoimuudessa ovat myös kysymykset siitä, kuka aineistot omistaa (Fecher ym. 2015, 15). Yleisimmin aineistot ovat tutkimuslaitoksen omaisuutta, vaikka tutkijat saattavatkin nähdä aineistot henkilökohtaisena omaisuutena (Strasser 2015, 10; Borg & Kuula 2007, 34). Tutkimusaineistot saattavat saada tekijänoikeuden suojan, jos niin sanottu teoskynnys ylittyy eli ne katsotaan

luovan työn omaperäiseksi tuotteeksi. Vaikka aineisto ei olisikaan tekijänoikeudella suojattua, pitäisi aineiston jakamisesta sopia kaikkien tutkimukseen osallistuneiden tutkijoiden kanssa jo tutkimuksen alussa. Jos aineistoon sisältyy muiden tuottamia osia, myös niiden käytöstä tulee sopia. (Yhteiskuntatieteellinen tietoaarkisto 2016.) Joskus tekijänoikeudet voivat olla aineiston avaamisen este. Näin on esimerkiksi silloin, jos aineistona käytetään julkisia tekijänoikeussuojaa saavia aineistoja. Nämä ongelmat tulevat esiin erityisesti taiteentutkimuksessa ja kulttuurintutkimuksessa. Erityisesti teknisillä aloilla tutkimusaineistojen avaamisessa on otettava huomioon patenttioikeus ja muut keksintöjä suojaavat oikeudet. (Borg ja Kuula 2007, 26, 34.) Patentteihin voi esimerkiksi liittyä suoja-aikoja, jotka estävät aineistojen avaamisen (Alakyyry, Korhonen ja Roinila 2017, 27).

Tutkijoiden kesken esiintyy myös haluttomuutta aineistojen jakamiseen. Yksilötasolla keskeinen syy on ajan puute. Tutkijat ovat kiireisiä ihmisiä, ja aineistojen tallentaminen on heille ylimääräistä työtä. (Chawinga ja Zinn 2019, 110, 114.) Esimerkiksi metatietojen tuottaminen vie aikaa, eikä kaikilla aloilla myöskään ole selvää, mitä metatietostandardia käytetään (Pampell ja Dallmeier-Tiessen 2014, 218). Tutkijat ovat yleensä nähneet aineistojen keräämiseen paljon vaivaa, joten heillä ilmenee myös halua säilyttää kontrolli aineistoon ja tietää, mihin aineistoa käytetään. Lisäksi tutkijoita huolestaa se, että heidän aineistoaan saatetaan tulkita ja käyttää väärin. (Chawinga ja Zinn 2019, 114.)

Tutkijat työskentelevät yleensä yliopistossa tai tutkimusorganisaatiossa. Näiden instituutioiden tasolla aineiston jakamisessa keskeistä on se, onko aineistojen jakamiseen riittävää osaamista. Jos tutkijoilla ei ole riittävää tietoa ja osaamista aiheesta, ei aineistojen jakamista tehdä. Aineistojen jakamiseen puuttuu myös kannustimia eikä siitä saa samanlaista tunnustusta kuin tutkimusjulkaisujen tekemisestä. Myös instituution tutkimuspolitiikka ja kulttuuri vaikuttaa halukkuuteen jakaa aineistoja. Tutkimusaineistojen jakaminen edellyttää myös hyviä teknisiä

ratkaisuja aineiston käsittelyyn ja säilyttämiseen. Monissa tutkimuksissa on havaittu tässä puutteita. (Chawinga ja Zinn 2019, 114, 117–118.)

5.2 Tutkijoiden näkemyksiä

Tenopirin ym. (2020, 3) tutkimuksessa selvitettiin asenteita ja käytäntöjä liittyen tutkimusaineistojen kuvailuun, säilyttämiseen ja jakamiseen. Kyselyyn vastasi vuosina 2017 - 2018 yhteensä 2184 tutkijaa 87 eri maasta (Tenopir ym. 2020, 7). Tutkimuksen mukaan yleisimmät paikat säilyttää aineistoa olivat henkilökohtainen tietokone tai instituution palvelimet. Hieman alle puolet vastaajista säilytti aineistoaan instituution, tieteenalakohtaisissa tai yleisissä data-arkistoissa. Asenteet aineistojen jakamiseen olivat positiivisia: 87 % tutkijoista olisi sekä valmis jakamaan omia aineistojaan että käyttämään muiden keräämiä aineistoja. (Tenopir ym. 2020, 11, 14.) Asenteiden ja todellisten käytäntöjen välillä on siis huomattava ero. Kuitenkin sekä asenteet että käytännöt olivat avoimuuden suhteen positiivisempia verrattuna vastaaviin vuosina 2011 ja 2015 julkaistuihin tutkimuksiin.

Tutkijoilta selvitettiin myös syitä siihen, mistä syistä aineistot on jätetty avaamatta. Merkittävin syy oli se, että tutkijat halusivat itse julkaista ensin; seuraavaksi merkittävimmät syyt olivat ajan puute ja puuttuvat oikeudet aineistojen avaamiseen. (Tenopir ym. 2020, 18, 23.) Vajaa puolet vastaajista ilmoitti rahoittajan vaativan aineistonhallintasuunnitelman tekemistä. Tutkimus toi esiin nykyistä suuremman tarpeen apuun aineistonhallintaan instituutioissa; vain noin kolmannes vastasi, että instituutio tarjosi apua aineistonhallintaan (Tenopir ym. 2020, 9, 13).

Berghman ym. (2017) selvittivät tutkijoiden käytäntöjä ja näkemyksiä aineistojen avoimuudesta kyselytutkimuksella, johon vastasi 1200 tutkijaa vuonna 2016. Myös heidän tutkimuksessaan tuli esiin ero asenteiden ja käytäntöjen välillä. Vastanneista kolmannes ei ollut tallentanut aineistojaan avoimesti mihinkään, vain 13 % oli tallentanut data-arkistoihin ja loput muualle kuten julkaisun yhteyteen tai datalehtiin.

Asenteet aineistojen avoimuuden suhteen olivat positiivisia: 65 % oli halukas jakamaan aineistoa ja 73 % koki avoimet aineistot hyödyllisenä itselleen ja 69 % hyödyllisenä tieteenalalle. Yksi mahdollinen syy sille, miksi aineistoa ei jaettu enempää, oli se että vain 25 % oli saanut harjoitusta aineistojen jakamiseen. (Berghman ym. 2017, 20–21.)

Kim, Schuler ja Pechenina (2017) tarkastelevat tutkimuksessaan aineistojen jakamiseen vaikuttavia tekijöitä. Tutkimusaineistona toimi vuonna 2015 kerätty kyselytutkimusaineisto (Kim ym. 2017, 5). Ihmistieteellisten aineistojen käyttäjillä aineistojen jakaminen ja jatkokäyttö oli muita vähäisempää todennäköisesti yksityisyydensuojaan liittyvien kysymysten takia. Tutkimuksen mukaan vanhemmat ja tutkimusta pidempään tehneet henkilöt ovat nuorempia tutkijoita valmiimpia jakamaan aineistojaan. Tutkimusrahoittajan vaatimuksilla oli vaikutusta - aineistoja jaettiin enemmän, jos tutkimusrahoittaja vaati aineistonhallintasuunnitelman tekemistä. Myös organisaation rahallinen tuki aineistonhallintaan ja selvät prosessit, miten aineistoa hallitaan ja säilytetään, vaikuttivat positiivisesti aineistojen jakamiseen. (Kim ym. 2017, 15–16.)

Zenk-Möltgenin ym. (2018) tutkimuksen mukaan tutkimusaineistojen jakaminen ei vielä ole sosiologiassa itsestään selvä käytäntö. Artikkeliotos vuosien 2012–2014 sosiologian lehdistä osoitti, että vaikka hieman yli puolissa tapauksissa tutkimusaineisto mainittiin saatavilla olevaksi, alle kolmasosassa lehtiartikkeleista aineisto oli todella saatavilla. Saatavuus oli kuitenkin kasvanut kyseisellä tarkastelujaksolla. Sosiologian tutkijoille osoitetun kyselytutkimuksen mukaan neljäsosassa julkaistuista tutkimusartikkeleista myös tutkimusaineistoja oli jaettu. Tutkijoiden oma asenne, aikaisemmat käytännöt ja sosiaalinen paine vaikuttivat positiivisesti aineistojen jakamisen käytäntöihin. Tutkimuksen mukaan suositusten ja käytännön välisen eron poistamiseen tarvittaisiin sekä tieteellisten lehtien datapolitiikkaa, toimivaa infrastruktuuria että instituutioiden ohjeistusta. (Zen-Möltgen ym. 2018.)

Myös Suomessa on tehty useita tutkimuksia ja selvityksiä tutkimusaineistojen avaamiseen liittyen. Borg ja Kuula selvittivät jo vuonna 2006 tutkijoiden suhtautumista

aineistojen avoimuuteen kyselyllä, joka kohdennettiin kaikille Suomen yliopistojen humanististen, yhteiskuntatieteellisten ja käyttäytymistieteellisten alojen professoreille. Kyselyyn vastasi 150 henkilöä ja vastausprosentti oli 28. Vielä vuonna 2006 yliopistot eivät aktiivisesti suositelleet sähköisten tutkimusaineistojen säilyttämistä, sillä 90 % vastaajista totesi, että laitoksella ei ole mitään ohjeita asiasta. Tutkimuksen jälkeen aineistot jäivät yleisimmin tutkijoiden haltuun. Aineistojen arkistointi erilliseen aineistojäilyttävään organisaatioon oli yleinen käytäntö vain 10 % vastaajista mielestä. Kyselyssä avoimuudelle tunnistettiin sekä hyötyjä että haittoja, mutta hyödyt arvioitiin haittoja suuremmiksi. Suurin hyöty nähtiin monipuolisten tutkimusasetelmien toteutumisessa ja tiedeyhteisön vuorovaikutuksen lisääntymisessä. (Borg ja Kuula 2007, 28–39.)

Kysyttäessä avaamisen esteistä ja haitoista kaikki tarjotut vaihtoehdot saivat kannatusta. Merkittävien este oli se, että aineistot olivat puutteellisesti järjestettyjä ja dokumentoituja ja tähän liittyen haittoina nähtiin tutkijoiden työajan kulumisen. 31 % vastaajista näki tekijänoikeuskysymykset vähintään melko tärkeinä syinä ja 47 % prosenttia sen, että aineiston omistajuudesta ja käyttöoikeuksista ei ollut sovittu aineistoa kerätessä. Jatkokäytön esteistä 43 % katsoi vähintään melko tärkeäksi sen, että tutkimusaineistot eivät vastaa uusiin tutkimusongelmiin ja 46 % pelkäsi, ettei jatkotutkimuksessa aineistoa käytetä oikein. Joka neljäs näki kilpailuedun menettämisen vähintään melko suurena haittana. Lähes puolet näki esteenä tietosuoja- ja tutkimuseettiset ongelmat. Kysymykseen siitä, kuinka sähköisten aineistojen avoimuuteen suhtaudutaan, kuitenkin 76 % vastasi ehdottoman tai melko myönteisesti. (Borg ja Kuula 2007, 28–39.) Tutkimuksesta voi siis päätellä, että vuonna 2006 aineistojen avaaminen oli vielä harvinaista, mutta tutkijoiden suhtautuminen esteistä huolimatta myönteistä. Toisaalta tutkimuksen vastausprosentti oli melko pieni, ja myönteisesti asiaan suhtautuvat voivat korostua vastaajissa.

Tampereen yliopistossa toteutettiin vuonna 2017 avoimeen tieteeseen liittyvä kysely, jossa selvitettiin yliopiston tutkijoiden avoimeen julkaisuun ja tutkimusaineistoihin liittyviä käytäntöjä ja palveluntarpeita. Vastauksia saatiin 146 ja vastausprosentti oli

12. Tuloksista kävi ilmi, että suurin osa (71 %) tutkijoista ei ole käyttänyt data-arkistoja. Jos aineistoja oli avattu, sitä perusteltiin rahoittajien tai lehtien vaatimuksilla (16 %) ja sillä, että sen katsottiin kuuluvan tieteen tekemiseen (33 %). Tärkeimmät avaamista estävät tekijät olivat tutkimuseettiset syyt (56 %) ja tiedon puute (38 %). Myös avaamisen työläys ja tutkimusryhmässä tehdyn sopimuksen puute sekä lainsäädännölliset ja kaupalliset syyt mainittiin estävinä tekijöinä. (Tampereen yliopisto 2017.) Jatkona avoimen tieteen kyselylle Huuskonen ja Toikko (2018, 49) selvittivät otannalla vuonna 2017 Tampereen yliopistossa julkaisuista artikkeleista käytäntöjä tutkimusaineistojen suhteen. Artikkeleissa oli hyvin vähän mainintoja aineiston tallennuksesta jatkokäyttöä varten, minkä perusteella aineistojen avaaminen vaikuttaa olevan vielä vähäistä.

Tutkimusaineistojen avaamisen esteitä on selvitetty myös Helsingin yliopistossa. Selvityksen mukaan tutkimusaineistojen avaaminen vaatii useimmiten ulkoista painetta ja tutkimusaineistojen elinkaariajattelu ei ole vielä vahva osa tutkimuskulttuuria. Aineistonhallinta on parempaa, jos aineiston säilytysratkaisu on olemassa. Säilyttämisen arvoisiin aineistoihin liittyvissä näkemyksissä löytyi eroja tieteenalojen välillä: yhteiskuntatieteellisillä ja humanistisilla aloilla korostuivat aineiston yhteiskunnallinen ja historiallinen merkitys, muilla aloilla näiden lisäksi kaupalliset tekijät ja lääketieteessä terveyteen liittyvät tekijät. Aineistojen jatkokäytölle yksi keskeinen haaste on sensitiivinen aineisto, jota muodostuu lääke- ja terveystieteissä sekä humanistisissa ja yhteiskuntatieteissä mm. haastatteluaineistojen muodossa. Sensitiivinen aineisto vaatii anonymisointia, mikä voi joissain tapauksissa olla haastavaa. Joillain tieteenaloilla aineistojen avaamisen esteenä on myös patentteihin liittyviä suoja-aikoja. Kaikilla tieteenaloilla esiin nousi kysymys aineiston metatiedoista, jonka tekemisen tutkijat kokivat ylimääräisenä työnä ja johon tarvittaisiin lisää koulutusta ja kannustimia. (Ala-Kyyny ym. 2017, 25–29.)

Enwald tutki suomalaisten tutkimusorganisaatioiden jäsenten käsityksiä, kokemuksia ja mielipiteitä tutkimusaineistojen jakamisesta ja pitkäaikaissäilytyksestä vuonna 2018. Tutkimus toteutettiin kyselytutkimuksena, jonka kohteena olivat suomalaisten

yliopistojen ja tutkimusorganisaatioiden jäsenet. Kysymykseen siitä, suositteleeko tutkimusorganisaatio aineistojen avaamista, vastattiin suunnilleen saman verran kyllä, ei tai en osaa sanoa. Myös vastaukset väitteeseen “voin huoletti ja mielelläni jakaa tutkimusaineistoni muiden kanssa” jakoutuivat tasaisesti. Tutkimusaineistojen saatavuuteen liittyen suurin osa vastasi tutkimusaineiston olevan vapaasti heidän tutkimusryhmänsä käytössä ja muiden tutkijoiden käytettävissä pyynnöstä. Suurin osa vastaajista toivoi aineiston säilyvän tallessa myös tutkimuksen jälkeen. Kyselyssä selvitettiin myös vastaajien huolenaiheita aineiston avaamiseen liittyen. Suurimmat huolenaiheet (37 % vastaajista) liittyivät laillisuuteen ja eettisyyteen. Myös aineistojen väärinkäyttöön ja väärinymmärrykseen liittyvät pelot nousivat esille. Avovastauksissa mainittiin aineiston arkaluontoisuus, mahdollinen tuhoutuminen ja se, että eettinen lupa ei kata aineiston avaamista. (Enwald 2018, 34–42.)

Yhteenvetona voidaan todeta, että aikaisempien tutkimuksien perusteella tutkijat suhtautuvat pääosin varsin myönteisesti aineistojen avoimuuteen. Kuitenkaan aineistojen avaaminen ei vielä käytännön tasolla ole yhtä yleistä, mitä asenteet antavat olettaa. Kulttuuri aineistojen avaamisen suhteen myös vaihtelee tieteenaloittain. Tutkimusaineistojen avoimuuden haasteita on tunnistettu useita. Keskeisiä ovat eettisyyteen, laillisuuteen ja tekijänoikeuksiin liittyvät syyt. Kuitenkin myös henkilökohtaiset syyt, kuten ajan puute, nousevat esiin tutkimuksissa. Tutkimuksista käy myös ilmi tarve aineistohallinnan ja avoimuuden koulutukseen organisaatioissa.

Tutkimusaineistojen avoimuus on erittäin ajankohtainen aihe tutkimusorganisaatioissa. Aineistojen jakamiseen liittyviä kysymyksiä ei ole vielä täysin tutkittu ja ymmärretty (Chawinga ja Zinn 2019, 110). Eri tieteenalojen tutkijoiden näkemysten vertailu on toistaiseksi ollut vähäistä (Enwald 2018, 44). Tehdyt tutkimukset ovat myös yleensä yleisillä tasoilla tehtyjä, ja syvällisesti tietyn tieteenalan tilanteeseen pureutuvia tutkimuksia ei ole paljoa (Zuidervijk ja Spiers 2018, 228). Esimerkiksi Enwaldin (2018, 44) mukaan tutkimusaineistoja avanneiden ja avattujen aineistojen käyttäneiden kokemusten ja näkemysten selvittämistä olisi syytä tehdä

lisää. Fecher ym. (2015, 20) näkevät että erityisesti tieteenalakohtaisia avoimuuden esteitä ja kannustimia pitäisi vielä tutkia.

6 TUTKIMUSASETELMA

6.1 Tutkimuskysymykset

Tämän tutkimuksen taustalla on yksi avoimen tieteen osa-alue eli tutkimusaineistojen avoimuus. Tähän liittyvät kiinteästi myös aineistojen hallintaan liittyvät kysymykset. Tieteen avoimuus, avoimet aineistot ja aineistonhallinta ovat tämän hetken tutkimusmaailmassa kehittyvä ja ajankohtainen ilmiö. Tässä tutkimuksessa ilmiötä tarkastellaan tutkijoiden näkökulmasta. Tarkoitus on tuoda esiin tutkijoiden näkemyksiä ja käytäntöjä liittyen aineistojen hallintaan ja avoimuuteen. Tilanne aineistojen avoimuuden suhteen vaihtelee suuresti tieteenaloittain. Tarkastelun kohteena ovat informaatiotutkimuksen ja sosiologian tutkijat. Tutkimuksessa kartoitetaan aineiston avaamisen hyötyjä sekä ongelmia ja esteitä kyseisillä tieteenaloilla. Tutkimuksessa selvitetään myös tutkijoiden näkemyksiä aineistojen avoimuuden yleisyydestä, käytännöistä ja siihen liittyvistä ilmiöistä heidän alallaan. Pohdintaosassa tutkimuksen tuloksia verrataan aikaisempiin vastaaviin tutkimuksiin.

Tutkimuskysymykset ovat seuraavat:

- Mitä käytäntöjä informaatiotutkimuksen ja sosiologian tutkijoilla on tutkimusaineistojen hallinnassa?
- Miten informaatiotutkimuksen ja sosiologian tutkijat suhtautuvat tutkimusaineistojen avoimuuteen?
 - o Mitä hyötyjä aineistojen avoimuudessa nähdään?
 - o Mitä ongelmia aineistojen avoimuudessa on ja mikä estää aineistojen avoimuutta?
- Millaisena sosiologian ja informaatiotutkimuksen tutkijat näkevät tilanteen aineistojen avoimuuden suhteen alallaan?

6.2 Tutkimusmenetelmä

Tutkimuksen empiirinen osa toteutettiin haastattelemalla Tampereen yliopiston tutkijoita. Hirsjärven ja Hurmeen (2008, 11) mukaan haastattelututkimus soveltuu käytettäväksi, kun halutaan syvällistä tietoa tutkittavasta aiheesta. Haastattelututkimukseen päädyttiin, koska sillä uskottiin saatavan kyselytutkimusta syvällisempää tietoa. Lisäksi oletettiin haastattelujen tavoittavan ihmiset hyvin. Kyselytutkimuksen vaarana olisi se, että vastausprosentti jää pieneksi. Aikaisemmin Suomessa tehdyissä kyselytutkimuksissa ja selvityksissä tästä aiheesta vastausprosentti on ollut melko pieni.

Tämän tutkimuksen menetelmänä on teemahaastattelu. Hirsjärven ja Hurmeen (2008, 48) mukaan teemahaastattelussa ihmisten tulkinnat asioista ja heidän asialle antamansa merkitykset ovat keskeisiä. Teemahaastatteluun päädyttiin, sillä tarkoituksena oli selvittää tutkijoiden näkemyksiä ja kokemuksia. Teemahaastattelussa pyritään löytämään merkityksellisiä vastauksia tutkimusongelmiin ja valitut teemat rakentuvat aiemmin tiedetyn eli tutkimuksen viitekehyksen mukaan (Tuomi ja Sarajärvi 2013, 75). Teemahaastattelu on puolistrukturoitu haastattelumenetelmä, ja se sijoittuu strukturoidun lomakehaastattelun ja avoimen haastattelun väliin. Teemahaastattelussa kaikkien haastateltavien kanssa käydään läpi tietyt aihepiirit eli teemat. Kysymysten tarkka muoto ja järjestys kuitenkin vaihtelevat. Teemahaastattelu mahdollistaa tarkentavien kysymysten tekemisen haastattelun aikana. (Hirsjärvi ja Hurme 2008, 47–48.)

Graduun käytettävä aika huomioon ottaen päädyttiin haastattelemaan kuutta tutkijaa kahdelta eri tieteenalalta. Informaatiotutkimuksen ja sosiologian valikoitumiseen tarkasteltaviksi tieteenaloiksi vaikutti henkilökohtainen mielenkiinto. Laadullisessa tutkimuksessa haastateltavien valinnassa on tärkeää, että henkilöt, joilta tietoa

kerätään, tietävät tutkittavasta ilmiöstä mahdollisimman paljon ja heillä on kokemusta asiasta. (Tuomi ja Sarajärvi 2013, 85). Niinpä haastateltavien valinnan ei pidä olla satunnaista, vaan harkittua ja tarkoitukseen sopivaa (Tuomi ja Sarajärvi 2013, 86). Hirsjärvi ja Hurme (2008, 58–59) puhuvat haastateltavien valinnan yhteydessä harkinnanvaraisesta näytteestä. Tässä työssä tarkasteltava aihepiiri eli tutkimusaineistojen hallinta ja aineistojen avoimuus ovat jossain määrin tuttuja kaikille tutkijoille, joten haastateltaviksi sopivat periaatteessa kaikki valittujen tieteenalojen tutkijat. Informaatiotutkimuksen tutkijoiden valinnan taustalla oli kuitenkin myös pohdintaa siitä, kenellä tutkijoilla olisi paljon kokemusta aineistojen käytöstä ja sanottavaa asiasta. Sosiologian tutkijoiden kohdalla valintaan yhtenä vaikuttavana tekijänä oli se, ketkä tutkijat olivat tallentaneet aineistojaan yhteiskuntatieteelliseen tietoarkistoon. Toisaalta kolme sosiologian tutkijaa kieltäytyi haastattelusta, joten sosiologien kohdalla haastateltavien valinta oli melko sattumanvaraista.

Ennen varsinaista haastattelua tehtiin koehaastattelu, jossa testattiin haastattelurungon toimivuutta. Sen jälkeen haastateltavia lähestyttiin sähköpostilla, jossa kerrottiin tutkimuksen tarkoituksesta, haastattelun teemoista ja haastatteluun liittyvistä käytännön seikoista kuten tutkimusluvasta ja aineiston käsittelystä. Haastateltaviksi tavoitettiin kolme informaatiotutkimuksen ja kolme sosiologian tutkijaa. Haastattelukysymykset lähetettiin haastateltaville tutustuttaviksi etukäteen. Haastattelut toteutettiin aikavälillä 5.10–2.11.2020. Viisi haastattelua toteutettiin etäyhteydellä zoomin välityksellä ja yksi haastattelu kasvotusten tapaamalla. Haastattelut kestivät 29–54 minuuttia. Haastattelut nauhoitettiin.

6.3 Aineiston analyysi

Haastattelunauhoitukset puhtaaksi kirjoitettiin eli litteroitiin. Litteroinnin tarkkuuden tarve vaihtelee tutkimusotteen ja tutkimustehtävän mukaan (Hirsjärvi ja Hurme 2008, 139). Tässä tutkimuksessa litteroitiin koko haastattelut ja litterointi tehtiin

sanantarkasti. Kuitenkaan esimerkiksi keskusteluanalyysissä tyypillisesti huomioituja äänenpainoja ja taukoja ei huomioitu.

Haastatteluaineistojen analysointiin on paljon erilaisia analyysitekniikoita ja työskentelytapoja. Standardoituja menetelmiä on vähän. Tyypillistä on, että aineiston analyysia tehdään useissa tutkimusprosessin vaiheissa. Haastatteluaineiston käsittelystä voidaan erottaa aineiston luokittelu, analysointi ja tulkinta. Tutkijasta riippuu, miten erillisinä toimintoina nämä nähdään. (Hirsjärvi ja Hurme 2011, 136).

Tässä tutkimuksessa aineiston analyysi perustuu sisällönanalyysiin, joka on laadullisessa tutkimuksessa käytettävä perusanalyysimenetelmä (Tuomi ja Sarajärvi 2013, 91). Sisällönanalyysin tarkoitus on luoda sanallinen kuvaus tutkittavasta ilmiöstä ja etsiä tekstistä merkityksiä. Sisällönanalyysin tarkoitus on tehdä tyypillisesti runsaasta ja hajanaisesta haastatteluaineistosta selkeää ja yhtenäistä informaatiota, joka mahdollistaa johtopäätösten tekemisen. (Tuomi ja Sarajärvi 2013, 104, 108.)

Laadullisen aineiston käsittelyssä oleellista on lukea aineistoa riittävän paljon (Hirsjärvi ja Hurme 2011, 143) ja tehdä valinta siitä, mikä aineistossa kiinnostaa (Tuomi ja Sarajärvi 2013, 92). Tätä tutkimusta tehdessä litteroidut haastatteluaineistot luettiin useita kertoja ja analyysiin päätyvät asiat valikoituivat tutkimusongelmien perusteella.

Seuraava vaihe on aineiston luokittelu, jonka sijaan voidaan käyttää myös tässä työssä hyödynnettyä teemoittelua. Siinä aineistoa pilkotaan ja ryhmitellään aihepiirien mukaan, minkä jälkeen aineistosta etsitään teemoja. Keskeistä on etsiä kutakin teemaa kuvaavia näkemyksiä. Teemahaastattelussa haastattelurunko luo jo luonnostaan jäsenyyttä teemojen etsintään. (Tuomi ja Sarajärvi 2013, 93.) Tässä työssä teemat valikoituivat pääasiassa tutkimuskysymysten ja haastattelurungon perusteella, mutta teemoja nousi myös aineiston pohjalta. Samaa teemaa kuvaavat näkemykset kerättiin aineistosta. Yhtenä teemana esimerkiksi olivat aineistojen avoimuuden hyödyt, jolloin kaikki tutkijoiden haastattelun aikana esiin tuomat avoimuuden hyödyt listattiin ja ryhmiteltiin.

Sisällönanalyysi voidaan jakaa aineistolähtöiseen, teorialähtöiseen ja teoriaohjaavaan sisällönanalyysiin. Aineistolähtöisessä analyysissä analyysiyksiköt eivät ole etukäteen sovittuja vaan nousevat aineistosta ja päättely on luonteeltaan induktiivista eli yksityisestä yleiseen etenevää. Teorialähtöisessä analyysissä aineiston luokittelu perustuu aiempaan viitekehykseen ja päättely on luonteeltaan deduktiivista eli yleisestä yksittäiseen etenevää. (Tuomi ja Sarajarvi 2013, 95–97, 108–113) Tässä työssä analyysi on lähinnä teoriaohjaavaa sisällönanalyysia. Teoriaohjaava sisällönanalyysi etenee aineiston ehdoilla, mutta analyysissä käytetyt teoreettiset käsitteet tuodaan aiemmin tiedetyn perusteella. Päättely on luonteeltaan abduktiivista, eli aineistolähtöisyys ja valmiit mallit vaihtelevat. (Tuomi ja Sarajarvi 2013, 97, 117.) Tässä tutkimuksessa analyysi on edennyt aineistosta nousevien asioiden pohjalta, mutta analyysin teossa on hyödynnetty viitekehyksen ja haastattelurungon mukaisia ryhmittelyjä ja käsitteitä.

7 TUTKIMUSTULOKSET

7.1 Taustatiedot

Tutkimuksessa haastateltiin kolmea informaatiotutkimuksen tutkijaa ja kolmea sosiologian tutkijaa Tampereen yliopistosta. Suorissa lainauksissa informaatiotutkimuksen tutkijoihin viitataan koodeilla H1, H2 ja H3 ja sosiologian tutkijoihin koodeilla H4, H5 ja H6. Tutkijoiden työuran pituus vaihteli kahdesta vuodesta useisiin vuosikymmeniin.

Tutkijoilta kysyttiin, minkälaisia aineistoja heidän tutkimuksissaan syntyy ja millaisia aineistoja he käyttävät. Tutkijoista kaksi käytti ainoastaan kvalitatiivisia aineistoja ja neljä sekä kvalitatiivisia että kvantitatiivisia aineistoja. Tässä tutkimuksessa tutkimusaineistona on käsitelty kaikkea, mitä tutkijat itse näkivät aineistoina. Taulukossa 1 on kuvattu tutkijoiden käyttämät aineistotyyppit ja kuinka moni haastatelluista tutkijoista on käyttänyt kyseisen tyyppistä aineistoa tutkimuksissaan.

Taulukko 1. Tutkijoiden käyttämät tutkimusaineistot.

Aineistotyyppi	Tutkijoiden lukumäärä
Haastatteluaineistot	6
Kyselyaineistot	4
Videokuva	3
Kenttä-/havainnointimuistiinpanot	3
Valmiit tilastoaineistot	2

Lokiaineistot	2
Kansainväliset vertailuaineistot	1
Arkistoaineistot (esim.kirjeet, sanomalehtiartikkelit)	1
Tutkimusjulkaisut	1
Tietokoneohjelmakoodit	1

7.2 Aineistonhallinta

Tutkijoilta kysyttiin, tekevätkö he aineistonhallintasuunnitelmaa ja kokevatko he sen hyödyllisenä. Yksi tutkija oli vasta ensimmäisessä tutkimushankkeessaan, jossa suunnitelma oli tehty hankkeen puolesta. Hän näki yleisesti aineistonhallintasuunnitelman kaltaisen suunnitelmallisuuden hyvänä tutkimuksen tekemisessä. Vain yksi informaatiotutkimuksen tutkija sanoi tekevänsä yleensä aineistonhallintasuunnitelman. Hän näki suunnitelman hyödyllisenä ensimmäistä isoa tutkimushanketta tehdessä tai jos aineistoja olisi tarkoitus jakaa kansainvälisesti. Hänen mukaansa tarkat aineistonhallintasuunnitelmat ovat kuitenkin yleensä rahoittajia varten tehtyjä ja vain suuntaa antavia – käytännössä tutkimus ei mene täsmälleen suunnitelman mukaan. Suunnitelmaa voisi päivittää tutkimuksen edetessä, mutta käytännössä sitä ei tule tehtyä. Kolme tutkijoista oli tehnyt aineistonhallintasuunnitelman vain niissä tutkimuksissa, missä tutkimusrahoittaja sitä vaati. Yksi heistä totesi aineistonhallintasuunnitelman teon olevan tutkimusmaailmassa varsin uusi asia. Hän näki sen kuitenkin hyödyllisenä, sillä suunnitelmaa tehdessä tuli pohdittua esimerkiksi aineiston jakamiseen liittyviä eettisiä ja juridisia kysymyksiä. Kaksi tutkijoista korosti sitä, että vaikka tutkijat yleensä eivät

aineistonhallintasuunnitelmaa tee, he kuitenkin muuten suunnittelevat aineistonhallintaa ennen tutkimusta.

Tutkijat säilyttivät yleisimmin aineistojaan omalla tietokoneellaan ja yliopiston tarjoamissa verkkokansioissa. Yksi tutkija mainitsi verkkokansion menevän helposti täyteen, jolloin säilytyspaikkana olivat myös ulkoinen kovalevy ja yliopiston pilvipalvelut. Hänen lisäksi yksi tutkija mainitsi yhdeksi säilytyspaikaksi yliopiston pilvipalvelut. Kahdella sosiologian tutkijalla oli myös laadullista paperimuotoista aineistoa, jota he säilyttivät omassa työhuoneessaan. Yksi tutkija työskenteli usean yliopiston hankkeessa, jolloin aineistot olivat toisen yliopiston palvelimella. Toisten omistamia aineistoja oli tutkimuksen aikana myös arkistoissa ja tilastokeskuksen palvelimella.

Tutkijoilta kysyttiin, ovatko he kohdanneet ongelmia aineistojen hallinnassa. Laajasti käsitettynä aineistonhallinta tarkoittaa aineiston käsittelyä koko sen elinkaaren ajalta, ja tutkijat vastasivatkin kysymykseen melko erilaisista näkökulmista. Informaatiotutkimuksen tutkija kertoi, että välillä on vaikea muistaa missä tiedostoissa on mitään, jos on tallentanut eri versioita eri paikkoihin. Hän kuvasi aineistonhallinnan kannalta haasteelliseksi tilanteen, jossa ulkomailla työskennellessään ei päässyt tallentamaan normaalisti käyttämäänsä yliopiston verkkokansioon, vaan joutui käyttämään muistitikkaa ja ulkomaisen yliopiston konetta. Kaksi informaatiotutkijan tutkijaa sanoi, että varsinaisesti aineistonhallinnassa he eivät ole kohdanneet ongelmia. Toinen heistä sanoi, että hänen kohtaamansa ongelmat liittyivät lähinnä aineistonkeruuvaiheeseen; esimerkiksi laitteet eivät toimineet tai ihmiset peruivat suostumuksensa osallistua tutkimukseen. Toinen heistä taas kuvasi aineiston jakamisvaiheen välillä haastavaksi, kun aineistojen avoimuutta pitäisi suosia mutta samalla olla hyvin tarkka henkilötietojen suojan kanssa.

Myöskään sosiologian tutkijat eivät kokeneet suuria ongelmia aineistonhallinnassa. Yksi tutkija mainitsi, että hän ei ole varma kaikkien vanhojen yliopistolle varastoitujen paperisten aineistojen sijainnista, mutta ne löytyisivät kyllä tarvittaessa. Toinen tutkija kertoi, että toisten tutkijoiden kanssa hankkeessa työskennellessä oli välillä tullut

viivästyksiä aineistojen saamisessa. Kolmas tutkija nosti esiin sen, että aineistojen jatkokäyttöön tulisi nykyään muistaa pyytää tutkittavien lupa. Hänelle epäselvää oli se, voisiko ennen EU:n uutta tietosuojasetusta kerättyjä haastatteluaineistoja hyödyntää uusissa tutkimuksissa.

Tutkijoilta kysyttiin, mitä he tekevät aineistoilleen tutkimuksen jälkeen, kuinka usein he avaavat aineistonsa ja missä palveluissa. Siihen, mitä aineistoille tehdään tutkimuksen jälkeen, vaikutti suuresti se, minkä tyyppisestä aineistosta on kyse. Yksi informaatiotutkimuksen tutkija kertoi pääsääntöisesti tallentavansa kvantitatiiviset aineistot, kuten kyselyaineistot, avoimesti saataville mutta kvalitatiivisia aineistoja ei. Hän ei kuitenkaan halunnut tuhota mitään aineistojaan, vaan säilytti niitä koneellaan. Toinen informaatiotutkimuksen tutkija taas sanoi pääsääntöisesti tuhoavansa sensitiiviset aineistot. Neutraalimmat aineistot, kuten tietokoneohjelmakoodit, hän sen sijaan jakoi. Sensitiivistä aineistoa kuitenkin jaettiin hänen mukaansa projektin sisällä. Kolmas informaatiotutkimuksen tutkija kertoi, että tutkimuksen jälkeen hän joko tuhoaa tai anonymisoi ja tallentaa sensitiiviset aineistot. Sen sijaan esimerkiksi tietokoneohjelmakoodit hän yleensä tallensi avoimesti saataville. Arkistoaineistot puolestaan ovat jo valmiiksi avoimia. Informaatiotutkimuksen tutkijoiden mainitsemia aineistojen avaamispaikkoja olivat Yhteiskuntatieteellinen tietoarkisto (FSD), Github, Zenodo ja yhteistyöhankkeissa myös muiden yliopistojen data-arkistot.

Sosiologian tutkijoilla ainut mainittu aineistojen avoin tallennuspaikka oli Yhteiskuntatieteellinen tietoarkisto. Yksi heistä sanoi tallentavansa lähes kaikki aineistonsa avoimesti saataville. Ainoastaan jotkut vanhat paperiset laadulliset aineistot olivat jääneet yliopiston arkistoon ja yksittäisiä hyvin sensitiivisiä aineistoja ei avattu. Toinen tutkija puolestaan käytti paljon muiden omistamia aineistoja, jotka täytyi tuhota käytön jälkeen. Itse kerätyistä aineistoista osa jäi tutkimusryhmälle ja osa tallennettiin avoimesti. Kolmas sosiologian tutkija kertoi, että hänen tutkimushankkeessaan osa aineistosta tallennetaan avoimesti ja osasta hän ei ollut varma, mutta ne mahdollisesti tuhoetaan kun kaikki julkaisut on tehty.

7.3 Avoimuuden hyödyt

Tutkijoiden vastauksista välittyi positiivinen suhtautuminen avoimuuteen. Kun tutkijoilta kysyttiin, ovatko he halukkaita jakamaan aineistojaan, lähtökohtaisesti kaikki olivat. Käytännössä päätökseen avoimuudesta kuitenkin vaikuttavat aineistojen laatu ja monet eri tekijät.

Tutkijoita pyydettiin erittelemään, mitä hyötyjä he näkevät aineistojen avoimuudessa. Viisi haastateltua tutkijaa mainitsi ajan ja vaivan säästymisen tai päällekkäisen tiedonkeruun välttämisen. Kaikki haastatellut tutkijat nostivat esiin joitain aineistojen uudelleenkäyttöön liittyviä näkökulmia. Kaksi sosiologian tutkijaa näki aineistojen jäävän tällä hetkellä harmillisen vähälle käytölle, mihin avoimuus on yksi vastaus. Avoimuus mahdollistaisi sen, että toinen tutkija voisi jatkaa analyysia siitä, mihin edellinen tutkimusryhmä jäi. Tutkijat kuvasivat avointen aineistojen mahdollistavan saman aineiston käytön eri näkökulmista ja eri käyttötavoilla. Yksi informaatiotutkija mainitsi avointen aineistojen nostavan esiin uusia tutkimusaiheita. Yksi sosiologian tutkija totesi avoimista aineistoista voivan löytyvän omaa tutkimusaineistoa täydentävää materiaalia.

Yksi tutkija näki aineistojen avoimuudessa myös mahdollisuuksia tieteenalojen väliseen vuorovaikutukseen:

”Aivan älytöntä koota aineistoo kahden kolmen tutkijan käyttöön pelkästään, et se on niin kallista tietoo, että sehän on niin kun suorastaan tuhlausta. Ja sitä paitsi se ei edistä edes tiedettä, että niin kun tieteen edistymisenkin kannalt on mahottoman hyvä se, et me saadaan niin kun erilaisii näkökulmii ja erilaisii käyttötapoja. Vielä ois todella hienoo se, että niin kun psykologien aineistoja tutkis vähä sosiologit ja päinvastoin, että siinä tulis vielä niin kun tämmöstä tieteellistä ristivalotusta, et kun kysymykset vaihtelee, tavallaan niin kun aineistolle esitettävät kysymyksetkin vaihtelee sen

mukaan, että mistä tieteenalasta on kysymys, niin sehän niin kun rikastuttaa sitä". (H4, sosiologia)

Aineistojen ja tutkimusinstrumenttien avoimuudessa nähtiin mahdollisuus samojen tutkimusasetelmien toistamiseen ja pitkittäistutkimuksiin. Yksi tietokoneohjelmia aineistona käyttävä tutkija totesi, että tutkijan asema olisi erittäin hankala, jos näitä aineistoja ei olisi avoimena, vaan ne pitäisi rakentaa alusta. Kolme tutkijaa nosti esiin avointen aineistojen hyödyllisyyden opinnoissa ja opettamisessa, erityisesti oppinäytetöiden tekemisessä. Yksi sosiologian tutkija toi esiin myös sen, että myös graduaineistojen avoimuus voi olla hyödyllistä esimerkiksi jatkettaessa väitöskirjaan.

Kolme tutkijaa nosti haastatteluissa esiin hyödyksi tutkimuksen läpinäkyvyyden ja luotettavuuden. Avointen aineistojen avulla muut tutkijat voivat paremmin nähdä miten tutkimustuloksiin on päädytty ja varmistaa niiden luotettavuuden. Siten aineistojen avoimuus mahdollistaa paremmin tutkimuksen vertaisarvioinnin ja kritiikin.

Haastatteluissa mainittiin myös muita yksittäisiä hyötynäkökulmia. Jos aineistot tallennetaan avoimesti arkistoon, tutkija tietää, että aineistot ovat tallessa ja joku taho huolehtii niiden käytettävyydestä myös tulevaisuudessa. Aineistojen avoimuus voi myös nostaa tutkimuksen laatua, sillä huonosti tehtyjä tai puutteellisia aineistoja ei haluta tallentaa. Aineistojen avoimuus voisi myös johtaa positiiviseen yhteistyöhön tutkijoiden välillä. Avoimuuden kuvattiin myös yleisesti edistävän tiedettä ja toteuttavan avoimen tieteen tavoitteita.

Tutkijoilta kysyttiin, kokevatko he tutkijan saavan henkilökohtaista hyötyä aineistojen jakamisesta. Henkilökohtaisen hyödyn rooli aineistojen jakamisessa vaikutti olevan hyvin pieni – yleensä sitä ei koettu saatavan juuri lainkaan. Hyötynäkökulma ei noussut keskeiseksi, vaan aineistojen jakaminen nähtiin ennemminkin yhtenä tutkimusprosessiin kuuluvana asiana.

"Ehkä tää on enemmän tämmönen ideologinen ajatus, et se hyöty on tämmöstä et sen kuuluu olla avointa, ja ehkä tämmönen oma maine, että pitää itseään avoimena tutkijana, tai avoimen tieteen edistäjänä" (H1, informaatiotutkimus)

Viisi tutkijaa pohti haastattelussa aineistoihin viittaamisen kautta saatavan näkyvyyden mahdollisuutta. Molempien alojen tutkijat kuvasivat tilannetta kuitenkin niin, että käytännössä aineistojen jakaminen ei ainakaan vielä meritoi tutkijaa. Yksi sosiologian tutkija mainitsi sen, että avoimien aineistojen kautta toiset tutkijat voivat löytää aineistoihin liittyviä julkaisuja, ja sitä kautta avoimet aineistot voivat tuoda näkyvyyttä tutkijan työlle. Informaatiotutkimuksen tutkija kuvasi tilannetta näin:

”Se koko aineistoihin viittaaminen ja se sellanen, se on kyllä vielä ainakin meidän alalla ihan alkeellista. Että en mä nyt koe, että mä viel saisin siit mitään varsinaista meriittiä että joku, joku toinen käyttäis jotain mun aineistoo, se ei oo niin semmosta systemaattista se. Että vaikka jos joku viittaa johonki mun julkasuun nii sitte mä hyödyn siitä, mutta jos joku käyttäis mun aineistoo niin mä en ehkä vielä sitä ees sais tietää.” (H2, informaatiotutkimus)

7.4 Avoimuuteen vaikuttavat tekijät, ongelmat ja esteet

Haastatteluista kävi ilmi, että tutkijat suhtautuvat lähtökohtaisesti myönteisesti avoimuuteen, mutta avoimuuden toteutumiseen vaikuttaa käytännössä monet tekijät. Haastateltujen tutkijoiden omaan toimintaan vaikuttavat tekijät ja heidän yleisellä tasolla kuvaamansa avoimuuden esteet ja ongelmat on tässä kappaleessa yhdistetty. Tulosten tarkoitus on kuvata sitä, miksi aineistoa ei avata ja mitä ongelmia ja haasteita avoimuuteen liittyy.

Haastatellut tutkijat käyttivät monipuolisesti erilaisia tutkimusaineistoja, ja avoimuuteen liittyvät haasteet ja kysymykset vaihtelivat käytetyn aineistotyypin mukaan. Yleisesti voi sanoa, että laadullisen ja ihmisiin liittyvän aineiston, kuten havainnointi- ja haastatteluaineistojen, jakamiseen liittyi enemmän ongelmia. Helpommaksi avattaviksi kuvattiin kvantitatiiviset ja neutraaleja aiheita koskevat aineistot, kuten tietokoneohjelmat ja -koodit, testikokoelmat, sanastot ja

kyselyaineistot. Tutkijat käyttivät aineistoina myös esimerkiksi arkistomateriaaleja, sanomalehti- ja kirjeaineistoa, tutkimusjulkaisuja, muiden tahojen kokoamia tilastoja ja lokitietoja. Näihin aineistotyyppisiin liittyvät kysymykset ovat erilaisia.

Kahdessa haastattelussa nousi esiin kysymys aineistojen julkaisuajankohdasta. Informaatiotutkimuksen tutkija sanoi haluavansa saada omat julkaisunsa tehtyä, ennen kuin jakaa aineistot muille. Hän kuvasi tilannetta, jossa aineiston analyysi on viivästynyt ajan puutteen takia, mutta jakamista ei silti halua tehdä ennen kuin on itse julkaissut jotain. Sosiologian tutkija kuvasi aineistojen jakamisessa olevan tyypillisesti ns. embargoaika, eli aineistot jaetaan esimerkiksi hankkeen loppumisen jälkeen.

Kaikki haastatellut informaatiotutkijat ja yksi sosiologian tutkija mainitsivat avoimuuteen vaikuttavaksi tekijäksi sen, katsotaanko aineiston olevan hyödyllinen ja relevantti muille tutkijoille. Jos aineistot ovat esimerkiksi kooltaan pieniä tai aiheeltaan spesifejä, niiden ei nähdä kiinnostavan muita, eikä jakamista tehdä. Yksi tutkija kuvasi tutkijoiden voivan ajatella, että joissain tilanteissa aineistoa ei pystyisi hyödyntämään kukaan muu. Yksi tutkija nosti esiin sen, että jos aineisto on jollain tapaa huonolaatuista, ei sitä haluta jakaa.

”Jos mä vaan itse nään että aineisto olis sen kaltanen että siitä vois olla yleisempää hyötyä. Isot tilastoaineistot tietenkin on sellasia, että ne sopii hyvin jatkokäyttöön ja tietysti jotkut laajat haastatteluaineistotkin. Mutta sitten jos tehdään jotain hanketta varten vaikka pieni määrä laadullisia haastatteluja, vaikka kymmenkunta haastatteluja jotka vaikka sitten niin kun täydentää jonkun hankkeen muita isompia aineistokokonaisuuksia, niin en mä oikein tiedä sit tällais tapauksissa, että en mä lähtis niitä ainakaan tonne tietoaarkistoon kyllä viemään.” (H5, sosiologia)

Aineistojen relevanttiuteen liittyen yksi tutkija nosti ongelmaksi tallennettujen aineistojen vanhenemisen. Hänen mukaansa esimerkiksi monet tieteelliset lehdet näkisivät yli viisi vuotta vanhoista kyselyaineistoista tehdyt julkaisut vanhentuneina. Tämä nostaa esiin kysymyksen siitä, ketä vanhat tallennetut aineistot oikeastaan hyödyttävät.

Kaksi informaatiotutkimuksen tutkijaa mainitsi esteeksi aineistojen avaamisen työläyden ja siihen kuluvan ajan. Aikaa vie erityisesti aineiston anonymisointi ja kuvailu. Toinen tutkija kertoi, että on jättänyt laadullisia aineistojaan tallentamatta, koska prosessi on niin raskas ja hänellä ei ole aikaa. Kyselyaineistojen tallentaminen kuvattiin haastatteluissa laadullista aineistoa helpommaksi. Yksi tutkija kuvasi kyselyaineiston tallentamisen Yhteiskuntatieteelliseen tietoaarkistoon vaivattomaksi. Yhteiskuntatieteellinen tietoaarkisto oli hoitanut kaiken aineiston tallentamiseen liittyvän teknisen puolen. Toisaalta toinen tutkija mainitsi myös kyselyaineistoihin liittyvän muuttujatason kuvailun työlääksi. Yksi tutkija kuvasi aineistojen tallentamista resurssi- ja rahakysymykseksi, eli pitäisi olla rahaa palkata vaikka tutkimusapulainen tekemään siihen liittyvää työtä.

Avoimuuden haasteista eniten käsiteltyjä ja kaikissa haastatteluissa esiintyneitä teemoja olivat eettisyyteen, tietosuojaan ja sensitiivisyyteen liittyvät kysymykset. Nämä kysymykset liittyivät erityisesti laadulliseen ja ihmisiä koskevaan tutkimukseen, joissa aineistoon sisältyy henkilö- tai kontekstietoja.

”Sillon kun me tuodaan siihen kentälle yhtä aikaa se avoimuuden vaatimus ja sitten tää tämmönen ihmisen anonymisoinnin vaatimus, niin siinä on tavallaan jonkinlaista ristiriitaa ikään ku oleellisesti, joittenkin datatyyppien kohdalla” (H3, informaatiotutkimus)

Haastatteluissa kävi ilmi, että monet aineistot koettiin niin sensitiivisiksi, että niitä ei voida jakaa ollenkaan. Tällaisia olivat esimerkiksi videokuva-aineistot, jotkin ihmisistä kerätyt havainnointiaineistot, ihmisten toiminnasta kerätyt lokiaineistot, aihepiiriltään arkaluontoiset haastatteluaineistot, kirjeet ja salaisia asioita kuten yrityssalaisuuksia koskevat aineistot.

” Siis ne oli niin henkilökohtasii ja arkaluontosii ne haastattelut, että ne on tuol, niit ei oo jaettu, pelkästään ihan niin kun haastateltavien, aiheen arkuuden ja haastateltavien niin kun anonymiteetin suojelemiseksi niitä ei oo näytettykään kun tälle tutkijaryhmälle.” (H4, sosiologia)

Jos ihmisiä koskeva tutkimusaineistoa avataan, täytyy se anonymisoida niin, että henkilöt eivät ole tunnistettavissa. Haastatteluissa nousi esiin joitain tähän liittyviä haasteita. Kahdessa haastattelussa anonymisointi mainittiin työlääksi. Kahdessa haastattelussa se kuvattiin haasteelliseksi esimerkiksi siksi, että aineistoista ei saisi millään tavalla yhdistelemälläkään saada henkilöä selville. Tutkijoilla tulisikin olla aiheeseen riittävää osaamista. Mainittiin myös, että tutkijan tulee tapauskohtaisesti miettiä, onko anonymisointi edes mahdollista niin että aineisto pysyy ymmärrettävänä. Yksi tutkija puhui anonymisoinnin vaikutuksesta jatkokäyttöön. Hänen mukaansa erityisesti laadulliset aineistot ovat muutenkin helposti ulkopuoliselle vaikeaselkoisia, ja anonymisointi vaikeuttaa niiden tulkintaa entisestään.

Neljässä haastattelussa mainittiin tietosuojalainsäädäntö, jota tulee noudattaa aineistojen avoimuudessa. Esiin tuotiin EU:n tietosuoja-asetus, jonka myötä suhtautuminen tietosuoja-asioihin on tiukentunut. Tietosuojaan suhtautumista kuvattiin kolmessa haastattelussa tiukaksi, varovaiseksi tai byrokraattiseksi. Esiin tuotiin aiheeseen liittyvän osaamisen tarve ja se, että kaikilta osin käytännöt eivät vakiintuneet. Sosiologian tutkija näki tietosuojaosaamisen kuitenkin hyväksi Suomessa, ja asiantuntijatahojen kuten Yhteiskuntatieteellisen tietoarkiston roolin erittäin merkittäväksi.

Informaatiotutkimuksen tutkija kuvasi tutkijan joutuvan tasapainoilemaan avoimuuden vaatimusten ja henkilötietojen suojaamisen välillä. Tutkijat näkivät yleisesti ihmisten henkilötietojen suojaamisen tärkeäksi. Joistain vastauksista välittyi kuitenkin näkemys siitä, että joskus aiheeseen liitetään liikaakin uhkakuvia:

”Olis hyvä että enemmän avattas ja et tutkijat kokis et se ei ois niin pelottavaa ja ahdistavaa. Et ei tarviis pelätä, että tekee jotain väärin jos jakaa niitä aineistoja. Mä luulen et osa jättää senkin takia ihan semmosia avattavia aineistoja jakamatta, koska pelottaa et se ei oo jotenkin sallittua, tai et sinne on jääny jotain.” (H1, informaatiotutkimus)

Sosiologian tutkija kuvasi tilanteen, jossa hyödyllistä aineistoa ei saanut käyttää tietosuojalainsäädännön takia. Hän kommentoi tilannetta näin:

”Tietolainsäädäntö, tietoturvalainsäädäntö ja yksityisyyttä turvaava lainsäädäntö on tällä hetkellä muuttunu niin tiukaks, että se on kaikkein suurin este ja se on todella iso ongelma --- nää tietosuojakriteerit on minun mielestä este kunnon tieteen edistämiselle.” (H4, sosiologia)

Haastatteluissa mainittiin, että nykyisen EU:n tietosuoja-asetuksen myötä tutkittavilta pitää saada lupa aineistojen jatkokäyttöön. Lupa jatkokäyttöön pitäisi siis huomioida jo tutkimuksen suunnittelussa ja aineiston keruuvaiheessa. Yksi sosiologian tutkija mainitsikin avoimuuden esteeksi huonosti tai puutteellisesti tehdyt tutkimusluvut. Hän kertoi avoimuutta estäneen tutkimuksessaan sen, että tutkittaville tahoille on pitänyt luvata, ettei aineistoa jaeta. Myös havainnointi- ja lokitietoja keräävä informaatiotutkimuksen tutkija sanoi, että hänen tutkimuksissaan ihmiset eivät välttämättä suostuisi tutkittaviksi, jos aineistot laitettaisiin avoimeksi. Toisaalta yksi sosiologian tutkija kertoi hänen tutkimuksissaan ihmisten suhtautuneen aina positiivisesti kysymyksen aineistojen jatkokäytöstä.

Avoimiin aineistoihin liittyvät kysymykset toimivista tallennuspaikoista ja aineistojen löydettävyydestä. Näitä teemoja pohdittiin kahdessa informaatiotutkimuksen tutkijan haastattelussa.

”Sehän on myös ihan hirvee haaste, että kuka niitä aineistoja löytää, miten ne löydetään, miten joku tutkija jossain missä lie saa tietää että tutkija on koonnu tämmösen aineiston ja julkassu sen jossain hitsin suomalaisessa arkistossa, en mä tiedä, mutta tuolla maailmalla voi olla vaikka miten mahtavii tutkimusaineistoja mistä mulla ei oo niin ku hajuakaan, koska mistä niistä tietää, mistä niistä saa tietoo, miten niitä osaa hakee, se on kyllä tosi iso ongelma, mut mä luulen et sekin tulee tässä kehitty.” (H2, informaatiotutkimus)

Toinen informaatiotutkimuksen tutkija sanoikin, että hänen mielestään parhaiten toimisi se, että aineistot saataisiin linkitettyä ja haettavaksi tutkimusjulkaisujen

yhteyteen, jolloin ne olisi helpompi löytää. Hän kuvasi myös aineistojen tallennukseen liittyvää infrastruktuuria puutteelliseksi. Ratkaisut ovat usein yliopistotasoisia, kun taas tutkimus on kansainvälistä. Hän näki, että parhaiten toimisi tieteenalakohtainen kansainvälinen tallennuspaikka. Yksittäisten tallennuspaikkojen ongelmista mainittiin Zenodon huonot hakuominaisuudet ja se, että Yhteiskuntatieteellinen tietoarkisto ei ota kaikkea aineistoa vastaan.

Yksi avoimuutta estävä tekijä on kysymykset aineistojen omistuksesta ja tekijänoikeuksista. Näitä kysymyksiä käsiteltiin yhdessä informaatiotutkimuksen tutkijan ja yhdessä sosiologian tutkijan haastatteluissa. Jos aineiston omistusoikeus on jollain muulla taholla, ei niitä voi julkaista, vaikka tutkija muokkaisi aineistot omaan tutkimukseensa sopivaksi. Esimerkkinä mainittiin tilastokeskuksen omistamat aineistot. Samoin tutkijoiden rikastamia arkistoaineistoja ei ole mahdollista julkaista avoimesti, vaikka arkistoon voi kuka tahansa mennä tutkimaan näitä aineistoja.

Monissa haastatteluissa tulivat esiin aineistojen jakamisen tiedettä edistävät vaikutukset. Yksi informaatiotutkimuksen tutkija kuitenkin toi esiin sen, että aineistojen jakamiseen voi liittyä myös tieteen tekemistä kapeuttavia tai hidastavia uhkakuvia. Uusia aineistoja ei esimerkiksi viitsitä kerätä, jos valmiita on saatavilla, ja tutkimuksen aiheet saattavat rajautua saatavilla olevien aineistojen mukaan.

Tutkijoiden tietoisuus, osaaminen ja asenteet vaikuttavat aineistojen jakamiseen. Yksi sosiologian tutkija totesi, että kaikki tutkijat eivät ole riittävän tietoisia siitä, että aineistot pitäisi jakaa. Toinen sosiologian tutkija mainitsi tietosuojan ja erityisesti EU:n tietosuoja-asetuksen soveltamisen usein haasteellisiksi tutkijoille. Informaatiotutkimuksen tutkija taas nosti esiin sen, että esimerkiksi aineistonhallintasuunnitelman tekeminen ja aineistojen anonymisointi vaativat osaamista, jota kaikilla ei välttämättä ole. Pari tutkijaa korostikin aineistojenhallinnan asiantuntijatahojen merkitystä. Haastatellut tutkijat suhtautuivat itse positiivisesti aineistojen jakamiseen. Kuitenkin kaksi informaatiotutkimuksen tutkijaa mainitsi myös kilpailuedun ja epäreilouden kokemuksen yleisesti mahdollisena jakamista estävinä tekijöinä. Aineistojen kerääminen ja tallentaminen on todella työlästä, ja valmista

aineistoa hyödyntävä pääsee helpommalla. Aineiston pitämisen itsellä mahdollisimman pitkään voi nähdä myös kilpailuetuna.

Yksi sosiologian tutkija käytti aineistoina myös tutkimusjulkaisuja. Niiden saatavuuteen liittyen hän mainitsi mahdolliseksi esteeksi kustantajien ja yliopistojen välisten sopimusten puuttumisen. Jos julkaisuihin ei pääse maksumuurien takia suoraan käsiksi, joutuu ottamaan yhteyttä tutkijoihin aineistojen saamiseksi, mikä on erittäin työlästä.

Tutkijoilta kysyttiin, millaisia haasteita aineistojen jatkokäyttöön voi liittyä. Yksi haastateltava toi esiin sen, että usein aineistot eivät erilaisen näkökulmansa takia sovi suoraan omaan tutkimukseen, tai ne eivät sisällä sitä tietoa mitä olettaisi. Useassa haastattelussa tuotiin esiin se, että joskus toisten keräämien aineistojen ymmärtäminen voi olla hankalaa, kun aineiston keruuta ei ole itse ollut tekemässä ja suunnittelemassa. Laadulliset aineistot, kuten havainnointi- ja haastatteluaineistot ja etnografisilla menetelmillä kerätyt aineistot, nähtiin erityisesti haastavina ulkopuolisen ymmärtää. Aineistojen pitäisikin olla huolellisesti dokumentoitu ja kuvailtu. Yhden tutkijan mukaan aineiston anonymisointi helposti vaikeuttaa sekä laadullisten että määrällisten aineistojen tulkintaa. Kahdessa haastattelussa mainittiin se, että aineistojen laatu ei ole aina riittävää, tai että tutkija olisi itse tehnyt ne toisella tavalla. Yksi informaatiotutkimuksen tutkija mainitsi sen, että aineistojen yhteydessä ei aina ole viittausohjetta, ja voi olla epäselvää, kuinka aineistoihin tulisi viitata. Toisaalta yksi monipuolisesti aineistoja uudelleen käyttänyt sosiologian tutkija näki ongelmaksi lähinnä ajan puutteen – hyviä aineistoja olisi paljon tarjolla, mutta aikaa niiden analysointiin rajallisesti.

7.5 Näkemyksiä avoimuuden tilanteesta

Kaksi informaatiotutkimuksen tutkijoista sanoi kokevansa hankalaksi arvioida, miten yleistä aineistojen jakaminen heidän alallaan on. Informaatiotutkimuksessa jaetaan

erityisesti niitä aineistoja, jotka ovat neutraaleja ja avoimuus on siten luontevaa. Jakamista tehdään paljon esimerkiksi tiedonhaun tutkimuksessa. Paljon jaettuja aineistotyypppejä ovat testikokoelmat, sanastot, siivotut lokiaineistot ja tietokonekoodit. Ihmistutkimuksen aineistot, kuten haastattelut ja havainnointiaineistot, ovat sen sijaan haastavampia, joten niitä jaetaan vähemmän. Informaatiotutkimuksen osalta mainittuja säilytyspaikkoja avoimille aineistoille olivat Github, Zenodo, Yhteiskuntatieteellinen tietoarkisto, yliopistojen palvelimet ja arkistot ja tutkimusryhmien sivut. Lisäksi mainittiin kirjastojen ja arkistojen portaalit, joiden kautta sai avoimia aineistoja. Testikokoelmia jaettiin netissä avoimesti, mutta myös niin että niissä täytyi allekirjoittaa lisenssisopimus ja ne olivat maksullisia. Muutenkin aineistoja jaetaan välillä myös esimerkiksi salasanalla suojattuna, eli niin että ne eivät ole suoraan avoimia

Kaksi informaatiotutkimuksen tutkijoista sanoi, että jakamista on tehty jonkin verran aina. Esimerkiksi testikokoelmia on jaettu jo 30 vuotta sitten. Kaikkien informaatiotutkimuksen tutkijoiden mukaan kuitenkin yleinen trendi on se, että avoimuus on viime vuosina lisääntynyt. Yksi heistä kertoi, että avoimia aineistoja on huomattavasti enemmän saatavilla nyt verrattuna esimerkiksi 80-lukuun. Yksi tutkija sanoi aineistojen avoimuuden olevan uusi asia. Hänen mukaansa avoimesta tieteestä ja aineistojen jakamisesta on puhuttu pitkään, mutta käytännössä asiassa ollaan vielä melko alussa. Rahoittajien ja tieteellisten lehtien vaatimukset ovat lisänneet avoimuutta. Toinen tutkija sanoi, että tietoisuus avoimuudesta on lisääntynyt ja aineistojen jakaminen on tuetumpaa ja systemaattisempaa kuin ennen. Hänen mukaansa myös tietoisuus aineistojen jakamiseen liittyvistä uhista kuten tietosuojariskeistä on lisääntynyt. Hänen lisäksi myös toinen tutkija toi esiin sen, että lainsäädäntö ja tietosuojavaatimukset ovat viime vuosina lisääntyneet. Kaksi informaatiotutkimuksen tutkijaa totesi avoimuuden käytäntöjen olevan osin vielä muotoutumassa. Yksi tutkija pohti koko tutkimuskulttuurin ja tieteen teon olevan muutoksessa.

Yleisesti informaatiotutkimuksen tutkijat näkivät avoimuuden lisääntymisen positiivisena asiana. Yksi tutkija sanoi kannattavansa avoimuutta ja avoimuuteen liittyvän uhkakuvien lisäksi paljon mahdollisuuksia. Kaksi tutkijaa sanoi avoimuuden olevan sinänsä hyvä asia, mutta kaikkea aineistoa ei ole mahdollista jakaa. He kuvasivat tutkijoiden kokevan helposti ristipainetta avoimuuden ja tiukkojen juridisten näkökohtien välillä.

Sosiologian osalta yksi tutkija kuvasi kansainvälistä tilannetta niin, että kehittyneissä maissa aineistot ovat nykyään pitkälti avoimia, mutta on myös maita, joissa avoimuus ei toteudu. Hän mainitsi esimerkkinä EU:n rahoittaman yhteiskuntatieteellisen tietoaarkiston, johon kuuluvat kaikki EU maat. USA:lla ja Venäjällä on myös omat tietoaarkistonsa. Lisäksi OECD-mailla on hänen mukaansa yhteinen arkisto survey-aineistoille. Kansallisella tasolla kaikki sosiologian tutkijat sanoivat Yhteiskuntatieteellisen tietoaarkiston olevan tallennuspaikka Suomessa. Kaksi tutkijaa näki tilanteen aineistojen avoimuuden suhteen hyvänä. Heidän mukaansa Yhteiskuntatieteellisestä tietoaarkistosta oli saatavilla paljon laadukkaita aineistoja. Kehitystä on viime vuosina tapahtunut erityisesti siinä, että myös laadullista aineistoa tallennetaan ja avoimeen tieteeseen ja aineistojen avoimuuteen kannustetaan yhtä enemmän.

”Nykytilanne on mun mielestä pääsääntöisesti hyvä. Et niin kun kaikki isot aineistot tallentuu automaattisesti ja enemmän ja enemmän myöskin kvalitatiivisesta datasta tallentuu automaattisesti, puhumattakaan siitä et mitä tehään digitalisoinnin suhteen kansallisarkistossa ja niin edelleen --- avoimuus on myöskin tän digitalisoinnin kautta lisääntyny ihan valtavii harppauksia, joka on hieno asia.” (H4, sosiologia)

Toinen tutkija sanoi, että sosiologiassa tehdään myös pieniä, spesifejä ja sensitiivisiä aineistoa sisältäviä tutkimuksia, joissa aineistoa ei aina tallenneta. Kolmas tutkija sanoi tutustuneensa Yhteiskuntatieteelliseen tietoaarkistoon, mutta koki tilanteen arvioinnin vaikeaksi. Hän mainitsi tutkimusjulkaisujen osalta tapahtuneen kehitystä, kun yliopistojen ja kustantajien sopimusten kautta pääsee nyt käsiksi julkaisuihin, jotka olivat aiemmin maksumuurin takana.

Tutkimusrahoittajien suositukset aineistoihin liittyen vaihtelevat tutkijoiden vastausten perusteella. Kaikki haastateltavat mainitsivat Suomen Akatemian. Sen kuvattiin vaativan nykyään yksityiskohtaista aineistonhallintasuunnitelmaa ja edellyttävän julkaisujen avoimuutta ja myös aineistojen avoimuutta silloin kun se on mahdollista. Akatemian sanottiin olevan edelläkävijä aineistojen avoimuuden suosituksissa Suomessa, ja yksi tutkija epäili, että Akatemian käytännöt voivat laajeta myöhemmin muille rahoittajille. Akatemian lisäksi mainittiin yleisesti isoimpien tutkimusrahoittajien, kuten monien eurooppalaisten tutkimusrahoittajien, suosittavan avoimuutta. Informaatiotutkimuksen tutkijoista kaksi sanoi, että säätiöillä sen sijaan ei ole avoimuuden suhteen toiveita ja yksi ei ollut asiasta varma. Myös kaksi sosiologian tutkijaa sanoi, että kaikilla pienemmillä rahoittajilla ei ole suosituksia aineistojen hallintaan tai avoimuuteen liittyen. Toisaalta yksi sosiologian tutkija sanoi, että kaikki keskeiset säätiöt, Akademia, työsuojelurahasto ja eurooppalaiset tutkimusrahoittajat edellyttävät aineistojen avoimuutta.

Informaatiotutkimuksen tutkijan mukaan rahoittajien vaatima avoimuusperiaate vaikuttaa toimintaan niin, että avoimuus tulee huomioitua enemmän. Toinen informaatiotutkimuksen tutkija sanoi, että rahoittajien takia tulee tehtyä aineistonhallintasuunnitelma, mutta jakamiseen rahoittajat eivät vaikuta, sillä hän haluaa muutenkin jakaa aineistonsa. Kolmas informaatiotutkimuksen tutkija kuvasi, että rahoittajia varten tulee tehtyä yksityiskohtainen aineistonhallintasuunnitelma, jossa kuvataan aineiston käsittelyä koko sen elinkaaren ajalta. Toisaalta aineiston kanssa toimisi hänen mukaansa melko samoin ilman suunnitelmaakin. Myös sosiologian tutkija kuvasi tilannetta samoin, eli rahoittajien vaatimukset vaikuttavat lähinnä niin, että tulee laadittua yksityiskohtainen kuvaus aineiston käsittelystä. Yhden sosiologian tutkijan mukaan avoimuus on tutkijoilla jo niin itsestään selvä käytäntö, että rahoittajien vaatimukset eivät enää vaikuta.

Tutkijoiden mukaan yliopisto ei aseta vaatimuksia aineistojen hallintaan tai avoimuuteen. Yliopistolla kuitenkin mainittiin olevan ohjeistuksia, koulutuksia ja tietoisuuden lisäämistä aiheeseen liittyen. Pari tutkijaa mainitsi yliopiston

datapolitiikan, joka suosittelee avoimuutta, mutta osa ei ollut kuullut yliopiston taholta mitään suosituksia aineistoihin liittyen. Julkaisujen avoimuuteen sen sijaan mainittiin olevan suosituksia ja myös kannustimia. Kaikki tutkijat olivat jossain määrin tietoisia yliopiston ja kirjaston tutkimusaineistoihin liittyvien palvelujen olemassaolosta. Esimerkiksi yliopistojen yhteinen DMPTuuli -palvelu, Yhteiskuntatieteellisen tietoarkiston ohjeistukset, kirjaston palvelut ja yliopiston datapalvelukeskus saivat mainintoja. Kaksi sosiologian tutkijaa eivät olleet itse käyttäneet näitä palveluja. Kaksi informaatiotutkimuksen tutkijaa sanoi itse käyttäneensä yliopistojen yhteistä DMPTuuli -palvelua ja Yhteiskuntatieteellisen tietoarkiston aineistoihin liittyviä ohjeistuksia. Yksi informaatiotutkimuksen tutkija ja yksi sosiologian tutkija sanoivat osallistuneensa aineistonhallintaa ja tietosuojaa käsitteleviin koulutuksiin, mutta muuten eivät ole tarvinneet palveluja. Tutkijat kokivat yleisesti yliopiston ja kirjaston palvelut hyvinä.

8 JOHTOPÄÄTÖKSET JA POHDINTA

Tämän tutkimuksen tarkoituksena oli selvittää informaatiotutkimuksen ja sosiologian tutkijoiden käytäntöjä tutkimusaineistojen hallinnassa ja näkemyksiä tutkimusaineistojen avoimuudesta. Ensimmäinen tutkimuskysymys oli, mitä käytäntöjä tutkijoilla on aineistojen hallinnassa. Toinen tutkimuskysymys oli, miten tutkijat suhtautuvat aineistojen avoimuuteen ja mitä kannustimia ja esteitä avoimuudessa on. Kolmas tutkimuskysymys oli, mikä on yleinen tilannekuvaa aineistojen avoimuuden suhteen tutkijoiden edustamilla tieteenaloilla. Tässä luvussa vastataan tutkimuskysymyksiin ja tarkastellaan tutkimuksen tuloksia suhteessa aiempaan tutkimuskirjallisuuteen.

Tämä tutkimus on luonteeltaan kuvaileva ja tarkastelee kuuden tutkijan käytäntöjä ja näkemyksiä. Tutkijat ovat tehneet erilaisia tutkimuksia, käyttäneet erilaisia aineistoja ja heidän työuransa ovat eripituisia. Niinpä yhteenvetoa aineistojen käytöstä oli hieman haastavaa tehdä. Jotkut tutkijat kokivat alan yleistä tilannetta koskevat kysymykset hankaliksi vastata. Tämä tutkimus kuvaakin haastateltavien käsitystä alan tilanteesta. Avoimuuden hyötyjä ja haasteita nousi haastatteluissa esiin runsaasti.

Informaatiotutkimuksen ja sosiologian tutkijoiden vastauksia verrataan toisiinsa niiltä osin, kuin niistä löytyy eroavaisuuksia ja vertailu on pieni otanta huomioiden mielekästä. Tieteenaloja merkittävämmäksi erottavaksi tekijäksi nousi aineistotyyppi. Yleisesti voi sanoa, että laadulliseen ja ihmistutkimukseen liittyi enemmän haasteita kuin kvantitatiivisiin ja neutraaleja aiheita käsitteleviin tutkimuksiin. Tuloksia tarkastellessa on syytä huomioida, että kaikki haastateltavat tutkijat ovat käyttäneet kvalitatiivisia ja neljä tutkijaa myös kvantitatiivisia aineistoja.

Tuloksia tarkastellessa täytyy huomioida myös se, että informaatiotutkimuksen tutkijoiden valinnassa oli harkinnanvaraisuutta ja osa sosiologian tutkijoista ei vastannut haastattelupyyntöön. Niinpä on mahdollista, että haastateltavat olivat

keskimääräistä tietoisempia aiheesta. Pienen ja harkinnanvaraisen otannan takia tämän tutkimuksen perusteella ei voida tehdä yleistyksiä koskemaan kaikkia sosiologian ja informaatiotutkimuksen tutkijoita.

Tutkimusaineistojen avoimuus edellyttää suunnitelmallisuutta ja hyvää aineistojen hallintaa (Cox ja Verbaan 2018, 53; Tenopir ym. 2020, 2). Tulosten mukaan tutkijat säilyttivät tutkimuksen aikana aineistonsa yleisimmin yliopiston verkkokansioissa. Suuria ongelmia aineistojen hallinnassa ei yleensä koettu. Jos tutkijat tekivät aineistohallintasuunnitelman, oli se yleensä rahoittajien vaatimuksien takia tehty. Ilman kirjattua suunnitelmaakin aineistohallintaa kuitenkin suunnitellaan. Suunnitelmallisuudessa nähtiin hyötyjä: esimerkiksi luvat jatkokäyttöön tulee muistaa kysyä tutkimuksen alussa tutkittavilta. Osa aineistosta tallennetaan ja osaa ei; päätökseen vaikutti esimerkiksi aineiston sensitiivisyys, omistusoikeudet ja aineiston relevanssi muille tutkijoille. Niiden aineistojen osalta, joita ei avattu, tutkijoiden välillä oli eroja sen suhteen, tuhotaanko aineistot vai säilyttääkö tutkija ne itsellään.

Useissa aikaisemmissa tutkimuksissa on todettu tutkijoiden suhtautuvan positiivisesti aineistojen avoimuuteen, mutta aineistojen avaaminen ei ole niin yleistä kuin asenteet antaisivat olettaa (Tenopir ym. 2020, 11, 14; Berghman ym. 2017, 20–21). Tämä tutkimus vahvistaa tätä käsitystä. Kaikki haastatellut tutkijat suhtautuivat lähtökohtaisesti positiivisesti aineistojen jakamiseen. Kuitenkin siihen käytännössä vaikuttavat monet eri tekijät.

Tämä tutkimus tukee näkemyksiä, joiden mukaan aineistojen jakaminen edistää tiedettä (mm. Chawinga ja Zinn 2019, 113) ja toteuttaa avoimen tieteen tavoitteita. Merkittävimmit aineistojen jakamisen hyödyiksi nousivat mahdollisuudet aineistojen uudelleenkäyttöön eri yhteyksissä ja tutkijoiden ajan ja vaivan säästyminen. Nämä hyödyt on todettu myös aiemmissa tutkimuksissa (Pampell ja Dallmeir-Tiessen 2014, 214; Borg ja Kuula 2007, 43; Patel 2016, 227). Puolet tutkijoista nosti esiin myös aiemmissa tutkimuksissa mainitun näkemyksen, jonka mukaan aineistojen jakamisen lisää tutkimuksen läpinäkyvyyttä ja todennettavuutta (Patel 2016, 227; Pampell ja Dallmeir-Tiessen 2014, 214).

Tässä tutkimuksessa esiin nousi avointen aineistojen hyöty tutkimuksen lisäksi opetuksessa. Kursseilla ja opinnäytetöitä tehdessä opiskelijoilla ei välttämättä ole mahdollisuutta kerätä omaa aineistoa, joten he voivat hyödyntää avoimesti tallennettuja aineistoja.

Henkilökohtainen hyöty ei noussut merkittäväksi jakamiseen kannustavaksi tekijäksi, vaan jakamisen nähtiin yleisesti kuuluvaan tieteenteon prosessiin. Patelin (2016, 227) mukaan tutkijat saavat aineistojen jakamisesta hyötyä viittausten kautta, toisaalta Jomierin (2017, 365) mukaan avoimeen tieteeseen liittyä haasteita tunnustuksen saamisessa tieteellisestä työstä. Tämän tutkimuksen pohjalta voidaan todeta, että informaatiotutkimuksessa ja sosiologiassa tutkijat eivät saa aineistojen julkaisemisesta vastaavaa meriittiä kuin tutkimusjulkaisuista. Informaatiotutkimuksen osalta esiin tuli myös se, että käytännöt aineistoihin viittaamiseen eivät aina ole selvät. Tämä herättääkin kysymyksen siitä, voiko tutkimuskulttuuri kehittyä tulevaisuudessa siihen suuntaan, että myös aineistoihin on selvät viittauskäytännöt ja tutkija saa niistä samaan tapaan tunnustusta kuin tutkimusjulkaisuista. Tälläkin hetkellä tutkijoiden julkaisut voivat kuitenkin mahdollisesti saada tunnettavuutta aineistojen kautta, ja aineistojen jakaminen voi johtaa positiiviseen yhteistyöhön tutkijoiden välillä.

Kaikki haastatellut tutkijat tekivät myös laadullista ihmistutkimusta, ja tämän tutkimuksen keskeisimmäksi avoimuuden haasteeksi nousikin tutkittavien tietosuojat. Vastauksista kävi ilmi, että on olemassa aineistotyyppisiä, joita ei niiden sensitiivisyyden takia voi jakaa ollenkaan, kuten aiemmassakin tutkimuksessa on todettu (Chawinga ja Zinn 2019, 118). Aineistojen anonymisointi voi olla työlästä, vaatia osaamista ja hankaloittaa aineistojen jatkokäyttöä. Avoimuuden ja tietosuojakäytännöiden välillä tasapainoilu voi olla tutkijoille haastavaa sekä tämän että aikaisempien tutkimusten mukaan (Strasser 2015, 11). Vastauksista voi päätellä, että tutkijat näkevät tutkittaviensa tietosuojaan tärkeänä, mutta toisaalta tietosuojakäytännöihin suhtautuminen nähdään hyvin tiukkana. Tämän tutkimuksen perusteella voidaankin todeta, että tietosuojaan ja esimerkiksi EU:n tietosuoja-

asetuksen soveltamiseen liittyvät ohjeistukset ja koulutukset ovat tärkeitä tutkimusorganisaatiossa.

Keskeiseksi avoimuuteen vaikuttavaksi tekijäksi nousi kysymys aineistojen relevanssista. Vastauksista voi päätellä, että tutkijat eivät halua jakaa aineistoa vain jakamisen takia, vaan he punnitsevat sitä, onko aineisto kiinnostava muille tutkijoille. Voidaankin miettiä, onko kaiken aineiston jakaminen edes mielekäästä.

Tutkijoiden asenteet, osaaminen ja tietoisuus asiasta vaikuttavat aineistojen jakamiseen. Haastatellut tutkijat suhtautuivat positiivisesti aineistojen jakamiseen, mutta esiin tuotiin se, että kaikki tutkijat eivät ole tarpeeksi tietoisia aineistojen jakamisesta. Aikaisemmissa tutkimuksissa (Chawinga ja Zinn 2019, 110, 114; Tenopir ym. 2020, 18, 23) on tullut esiin avoimuutta hidastavana tekijänä tutkijoiden halu julkaista itse ensin ennen aineistojen jakamista. Tämä tuli esiin myös tässä tutkimuksessa. Tutkijat näkivät asiantuntijatahot, kuten Yhteiskuntatieteellisen tietoarkiston, merkityksellisinä.

Aiemmissä tutkimuksissa (Chawinga ja Zinn 2019, 110, 114; Tenopir ym. 2020, 18, 23) mainittu tutkijoiden ajan puute mainittiin avoimuutta estävänä tekijänä myös tässä tutkimuksessa informaatiotutkimuksen tutkijoiden osalta. Aineistojen avaamisen voisi päätellä olevan osin resurssikysymys: tutkijoilla tulisi olla riittävästi työaika aineistojen tallentamiseen ja kuvailuun tai vaihtoehtoisesti mahdollisuus hyödyntää vaikka tutkimusavustajaa tekemään tätä työtä.

Tutkijat sanoivat myös toisten tutkijoiden aineistojen tulkinnan ja jatkokäytön olevan välillä haasteellista. Tämä korostuu laadullisessa tutkimuksessa. Tämän pohjalta voidaan todeta aineiston huolellisen dokumentoinnin olevan jatkokäytön kannalta tärkeää.

Sosiologian osalta välittyi kuva siitä, että aineistojen jakaminen on erittäin yleistä sekä Suomessa että kansainvälisesti kaikissa kehittyneissä maissa. Tutkijat näkivät tilanteen hyvänä. Erityisesti suuria tilastoaineistoja jaetaan paljon. Viime vuosina kehitystä on tapahtunut siinä, että yhä isompi osa myös hankalammin tallennettavasta laadullisesta

aineistosta tallennetaan. Kuitenkin sosiologiassa tehdään myös sensitiivisiä, pieniä tai spesifejä tutkimuksia, joissa aineistoa ei välttämättä jaeta. Sosiologiassa Suomessa aineistojen tallennus on keskittynyt yhteen paikkaan eli Yhteiskuntatieteelliseen tietoaarkistoon.

Informaatiotutkimus poikkeaa sosiologiasta siinä, että siinä aineistojen tallennuspaikkoja on useita aineistotyyppistä riippuen. Informaatiotutkimus on Suomessa pieni ala ja tutkimus on hyvin kansainvälistä. Haasteeksi nousikin aineistojen löydettävyys, kun yhtä kansainvälistä tieteenalakohtaista tallennuspaikkaa ei ole. Tämä herättääkin kysymyksen siitä, voisiko infrastruktuuri parantua ja esimerkiksi aineistojen linkitys tutkimusjulkaisujen yhteyteen lisääntyä tulevaisuudessa, jolloin aineistot olisivat helpommin löydettävissä. Informaatiotutkimuksessa neutraaleja aineistoja, kuten tietokoneohjelmakoodeja, on jaettu jo pitkään. Toisaalta taas sensitiivisiä ja ihmistutkimusaineistoja jaetaan vähemmän. Myös informaatiotutkimuksessa puhe aineistojen hallinnasta ja jakamisesta on viime vuosina lisääntynyt, ja tutkijat näkivät kehityksen positiivisena.

Tutkimusaineistoina voidaan käyttää myös arkistoaineistoja tai muita toisten omistamia aineistoja, jolloin oikeutta jakaa aineistoja eteenpäin ei välttämättä ole. Avoimuudessa on tapahtunut edistystä myös arkistoaineistojen digitalisoinnin kautta viime vuosina.

Tutkimusrahoittajista suurimmat, kuten Suomen Akatemia, suosittivat aineistojen avoimuutta ja vaativat aineistonhallintasuunnitelmaa. Kaikilla pienemmillä rahoittajilla kuten säätiöillä vastaavia vaatimuksia ei ollut. Tutkijat olivat tietoisia yliopiston ja kirjaston tarjoamista tutkimusaineistoihin liittyvistä palveluista, ja heidän käsityksensä niistä oli positiivinen. Kuitenkaan Tampereen yliopiston (n.d) avoimen tieteen linjauksen suositukset siitä, että aineistot avataan mahdollisuuksien mukaan, eivät olleet kaikille tutkijoille tuttuja.

Aikaisemmassa tutkimuksessa on todettu, että tutkimusaineistoon jakamiseen liittyviä kysymyksiä ei ole täysin ymmärretty, ja että tieteenalakohtaisia avoimuuden

kannustimia ja esteitä olisi syytä tutkia (Chawinga ja Zinn 2019, 110; Fecher ym. 2015, 20). Tässä pienessä laadullisessa tutkimuksessa tuotiin esiin Tampereen yliopiston informaatiotutkimuksen ja sosiologian tutkijoiden käytäntöjä ja näkemyksiä aiheesta. Laajemman kuvan saamiseksi näiltä aloilta aiheesta voisi tehdä esimerkiksi laajemman joukon tavoittaman kyselytutkimuksen. Tilannetta voisi tarkastella myös joiltain muilta tieteenaloilta. Esimerkiksi humanististen tai luonnontieteellisten alojen käyttämät tutkimusaineistot ja tilanne aineistojen jakamisen suhteen voivat olla hyvin erilaisia kuin nyt tarkastelun kohteena olleilla aloilla. Aihetta voisi tutkia laajemman kuvan saamiseksi kyselytutkimuksena, tai tietyn tieteenalan tilanteeseen syvällisemmin pureutuvana haastattelututkimuksena.

LÄHTEET:

- Alakyynty, J., Korhonen, T. & Roinila, M. (2017). Tutkimusdatan avaamisen esteet: haastattelututkimus Helsingin yliopistossa. *Signum*, 49(4), 25–29.
<https://journal.fi/signum/article/view/69198>
- Arza, V. & Fressoli, M. (2017). Systematizing benefits of open science practices. *Information services & use*, 37(4), 463–474. <https://doi.org/10.3233/ISU-170861>
- Avoimen tieteen kansallinen koordinaatio (2019a). *Avoim tiede*. Avoim tiede. Viitattu 12.2.2020. <https://avointiede.fi/fi/avoin-tiede>
- Avoimen tieteen kansallinen koordinaatio (2019b). *Koordinaatio*. Avoim tiede. Viitattu 16.2.2020. <https://avointiede.fi/fi/koordinaatio>
- Avoimen tieteen kansallinen koordinaatio (2019c). *Linjaukset ja aineistot*. Avoim tiede. Viitattu 16.2.2020. <https://avointiede.fi/fi/linjaukset-ja-aineistot>
- Bartling, S. & Friesike, S. (2014). Towards another scientific revolution. Teoksessa Bartling, S. & Friesike, S. (toim.) *Opening science - The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing*. (3–15.) Springer. <https://link.springer.com/book/10.1007%2F978-3-319-00026-8>
- Berghman, S., Cousijn, H., Deakin, G., Meijer, I., Mulligan, A., Plume, A., Rushforth, A., Rijke, S., Tatum, C., Tobi, S., Leeuwen, T., Waltman, L. (2017). *Open data. The researcher perspective*.
https://openaccess.leidenuniv.nl/bitstream/handle/1887/72152/Elsevier_CWTS_Brochure_v10_LO.pdf?sequence=1
- Borg, S. & Kuula, A. (2007) *Julkisrahoitteisen tutkimusdatan avoin saatavuus ja elinkaari. Valmisteluraportti OECD:n datasuosituksen toimeenpanomahdollisuuksista Suomessa*. Yhteiskuntatieteellinen tietoarkisto (FSD).
- Borgman, C. (2008). Data, disciplines and scholarly publishing. *Learned Publishing*, 21(1), 29–38. <https://doi.org/10.1087/095315108X254476>
- Candela, L., Castelli, D., Manghi, P., Tani, A. (2015). Data journals: A Survey. *Journal of the Association for Information Science and Technology*, 66(9), 1747–1762.
<https://doi-org.libproxy.tuni.fi/10.1002/asi.23358>

- Chawinga, D. C., Zinn, S. (2019). Global perspectives of research data sharing: A systematic literature review. *Library & Information Science Research*, 41(2), 109–122. <https://doi.org/10.1016/j.lisr.2019.04.004>
- Corti, L. (2007). Re-using archived qualitative data - where, how, why? *Archival Science*, 77(1), 37–54. <https://doi.org/10.1007/s10502-006-9038-y>
- Cox, A. & Verbaan, E. (2018). *Exploring research data management*. Facet Publishing. <http://web.a.ebscohost.com.libproxy.tuni.fi/ehost/detail/detail?vid=0&sid=64c1204e-4c9f-4de5-b408-efda4c10dcd7%40sdc-v-sessmgr03&bdata=JkF1dGhUeXBIPWNvb2tpZSxpcCx1aWQmc2l0ZT1laG9zdC1saXZlJnNjb3BIPXNpdGU%3d#AN=1817094&db=nlebk>
- Dietrich, N. & Wiebe, A. (2013). Definition of research data. Teoksessa Guibault, L. & Wiebe, A. (toim.) *Safe to be open. Study on the protection of research data and recommendations for access and usage*. (17). <https://doi.org/10.17875/gup2013-160>
- Eisfeld-Reschke, J., Herb, U. & Wenzlaff, K. (2014). Research Funding in Open Science. Teoksessa Bartling, S. & Friesike, S. (toim.) *Opening science - The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing*. (237–255.) Springer. <https://link.springer.com/book/10.1007%2F978-3-319-00026-8>
- Enwald, H. (2018). Tutkimusaineiston avoin jakaminen - tutkimusorganisaatioiden jäsenten käsityksiä, kokemuksia ja mielipiteitä. *Informaatiotutkimus*, 37(4), 26–47. <https://doi.org/10.23978/inf.77411>
- Enwald, H., Huuskonen, S., Kuusniemi, M. E. & Söderholm M. (2016). Aineistohallintaan suunnitelmallisuutta - Tuuli-projekti. *Informaatiotutkimus*, 35(1–2), 29–32. <https://journal.fi/inf/article/view/58981>
- Erola, J. & Räsänen, P. (2014a). Sosiologinen lähestymistapa. Teoksessa Erola, J. & Räsänen, P. (toim.) *Johdatus sosiologian perusteisiin*. (11–27.) Gaudeamus.
- Erola, J. & Räsänen, P. (2014b). Sosiologinen tutkimus ja menetelmät. Teoksessa Erola, J. & Räsänen, P. (toim.) *Johdatus sosiologian perusteisiin*. (60–78.) Gaudeamus.

- European Science Foundation (2020). *About*. PlanS. Viitattu 16.2.2020.
<https://www.coalition-s.org/about/>
- Fairdata.fi (n.d). *FAIR-periaatteet*. Fairdata.fi. Viitattu 9.2.2020.
<https://www.fairdata.fi/miksi-fairdata/fair-periaatteet/>
- Fecher B., Friesike S. & Hebing, M. (2015). What Drives Academic Data Sharing? *PLoS ONE*, 10(2). <https://doi.org/10.1371/journal.pone.0118053>
- Force 11. (2017). *The Fair data principles*. Viitattu 9.2.2020.
<https://www.force11.org/group/fairgroup/fairprinciples>
- Forsström, P. & Kutilainen, T. (2019). *Kohti yllättäviä löytöjä ja luovia oivalluksia - Avoin tiede ja tutkimus -hankkeen loppuraportti*. Avoin tiede ja tutkimus -hanke.
<https://avointiede.fi/fi/koordinaatio/hankearkistot/att-hankkeen-arkisto>
- Forsström, P., Kutilainen, T. & Räisänen, S. (2015). Avoin tiede ja tutkimus vahvistaa suomalaista tutkimusta. *Signum*,3. <https://journal.fi/signum/article/view/53289>
- Fält, K. & Okuloff, A. (2016). *Avoin tiede ja tutkimusaineistot -kysely 2015. Raportti terveystieteiden ja humanististen alojen tutkijoille suunnatun kyselyn tuloksista*. Yhteiskuntatieteellinen tietoaarkisto.
<https://trepo.tuni.fi/bitstream/handle/10024/99129/978-952-03-0107-1.pdf?sequence=1&isAllowed=y>
- Goertzen, M. (2017). Introduction to Quantitative Research and Data. *Library technology reports*, 35(4), 12–18. https://search-proquest-com.libproxy.tuni.fi/docview/1903876597?rfr_id=info%3Axri%2Fsid%3Aprimo
- Guedj, D. & Ramjoué, C. (2015). European Commission Policy on Open-Access to Scientific Publications and Research Data in Horizon 2020. *Biomed Data J*, 1(1), 11–14. <https://doaj.org/article/97f2e8bb71d6439aa3c05c2ebd876074>
- Higman, R., Bangert, D. & Jones, S. (2019). Three camps, one destination: the intersections of research data, management, FAIR and Open. *The UKSG Journal*, 32(1). <https://insights.uksg.org/articles/10.1629/uksg.468/>
- Hirsjärvi, S. & Hurme, H.(2008). Tutkimushaastattelu. Teemahaastattelun teoria ja käytäntö. Gaudeamus.

- Holopainen, M. & Koskinen, K. (2017). Avoin tiede ja tiedejulkaisemisen uudet tuulet. *Tieteessä tapahtuu*, 35(1), 43–46. <http://journal.fi/tt/article/view/60789>
- Huuskonen, S. & Toikko, T. (2018). Tampereen yliopiston tutkimusaineistosiselvitys. *Informaatiotutkimus*, 37(3), 47–50. <https://doi.org/10.23978/inf.76075>
- Jomier, J. (2017). Open science – towards reproducible research. *Information Services & Use*, 37, 361-367. <https://doi.org/10.3233/ISU-170846>
- Kim, J., Schuler, E. & Pechenina, A. (2017). Predictors of Data Sharing and Reuse Behavior in Academic Communities. *Knowledge Discovery and Data Design Innovation*, 1–25. https://doi.org/10.1142/9789813234482_0001
- Krishnan, A. (2009). *What are academic disciplines? Some observations on the Disclipinarity vs. Indisclipinarity Debate*. ECRS National Centre for Research Methods. Working paper 03/09. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.467.3004&rep=rep1&type=pdf>
- Ketola, A. (2014). Tiedollinen itsemääräämisoikeus ja laaja suostumus ihmistieteellisessä tutkimuksessa. Teoksessa Ketola, A. & Lahti, R. (toim.) *Ihmistieteellisten tutkimusaineistojen jatkokäyttö ja tietosuoja*. Helsingin yliopiston oikeustieteellisen tiedekunnan julkaisut.
- Late, E. & Talja, S. (2016). Tutkimusdata-palvelut eurooppalaisissa Yliopistokirjastoissa. *Informaatiotutkimus*, 35(3), 45–47. <https://journal.fi/inf/article/view/59435>
- Lilja, J. (2017). Tieteelliset lehdet ja tutkimusdata. *Informaatiotutkimus*, 36(1), 1–7. <https://doi.org/10.23978/inf.63189>
- Mäkinen, I. (2010). Informaatiotutkimuksen tie. Teoksessa Serola, S. (toim.), *Ote informaatiosta. Johdatus informaatiotutkimukseen ja interaktiiviseen mediaan*. BTJ.
- OECD (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding*. OECD Publishing. https://www.oecd-ilibrary.org/science-and-technology/oecd-principles-and-guidelines-for-access-to-research-data-from-public-funding_9789264034020-en-fr

- Pampel, H. & Dallmeier-Tiessen, S. (2014). Open Research Data: From Vision to Practice. Teoksessa Bartling, S. & Friesike, S. (toim.), *Opening science – The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing*. (213–224.)
Springer. <https://link.springer.com/book/10.1007%2F978-3-319-00026-8>
- Patel, D. (2016). Research data management: a conceptual framework. *Library Review*, 65(4/5), 226–241. <https://doi.org/10.1108/LR-01-2016-0001>
- Piwovar, H. & Vision, T. (2013). Data reuse and the open data citation advantage. *Dryad*. <https://doi.org/10.5061/dryad.781pv>
- Rentier, B. (2016). Open Science – A Revolution in sign? *Interlending & Document Supply*, 44(4), 155–160.
<https://www.emerald.com/insight/content/doi/10.1108/ILDS-06-2016-0020/full/html>
- Sitek, D. & Bertelmann, R. (2014). Open Access: A State of the art. Teoksessa Bartling, S. & Friesike, S. (toim.) *Opening science – The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing*. (139–153.)
Springer. <https://link.springer.com/book/10.1007%2F978-3-319-00026-8>
- Strasser, C. (2015). *Research data management*. National information standards organisation (NISO). <https://wiki.lib.sun.ac.za/images/2/24/PrimerRDM-2015-0727.pdf>
- Suomen akatemia. (2019, huhtikuu 1). *Avoin tiede: avoin saatavuus ja aineistojen avaaminen*. Suomen akatemia. Viitattu 16.2.2020.
https://www.aka.fi/avoin_tiede
- Tampereen Yliopisto (n.d). *Avoin tiede*. Tampereen Yliopisto. Viitattu 16.2.2020. <https://www.tuni.fi/fi/tutkimus/vastuullinen-tiede/avoin-tiede>
- Tampereen Yliopisto (2017). *Avoin tiede -kyselyn tuloksia*. https://www.uta.fi/sites/default/files/inline-files/Avoin_tiedekysely_%20tulokset.pdf

- Tampereen Yliopisto (2018a, kesäkuu 11). *Avoin tiede ja tutkimus Tampereen korkeakouluyhteisössä*. <https://www.tuni.fi/sites/default/files/2019-02/avoin-tiede-tampereen-korkeakouluyhteisossa.pdf>
- Tampereen Yliopisto (2018b, joulukuu 11). *Tampereen yliopiston avoimen tieteen ja tutkimuksen toimenpideohjelma*. <https://www.tuni.fi/sites/default/files/2019-02/tau-avoimen-tieteen-toimenpideohjelma.pdf>
- Tenopir, C., Hughes, D., Allard, S., Frame, M., Birch, B., Baird, L., Sandusky, R., Langseth, M., Lundeem, A. (2015). Research Data Services in Academic Libraries: Data Intensive Roles for the Future? *Journal of e-science Librarianship*, 4(2), 1–21. <http://doi.org/10.7191/jeslib.2015.1085>
- Tenopir, C., Talja, S., Horstmann, W., Late, E., Hughes, D., Pollock, D., Schmidt, B., Baird, L., Sandusky, R., Allard, S. (2017). Research Data Services in European Academic Research Libraries. *LIBER Quarterly*, 27(1), 23–44. <http://doi.org/10.18352/lq.10180>
- Tenopir, C., Rice, N. M., Allard, S., Baird, L., Borycz, J., Christian, L., Grant, B., Olendor, R. & Sandusky, R. J. (2020). Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLoS ONE*, 15(3), 1–26. <https://doi.org/10.1371/journal.pone.0229003>
- The National academies of Sciences, Engineering & Medicine (2018). *Open Science by design – Realizing a Vision for 21st Century Research*. The National Academies Press. <https://www.nap.edu/catalog/25116/open-science-by-design-realizing-a-vision-for-21st-century>
- Tuomi, J. & Sarajärvi, A. (2013). Laadullinen tutkimus ja sisällönanalyysi. Tammi.
- Yhteiskuntatieteellinen tietoaarkisto (2017). *Miksi aineistonhallintaa ja jatkokäyttöä?* Viitattu 10.3.2020. <https://www.fsd.tuni.fi/aineistonhallinta/fi/miksi-aineistonhallintaa-ja-jatkokayttoa.html>
- Yhteiskuntatieteellinen tietoaarkisto (2016, tammikuu 18). *Sopimukset ja oikeudet*. Aineistonhallinnan käsikirja. Viitattu 6.4.2020. <https://www.fsd.tuni.fi/aineistonhallinta/fi/sopimukset-ja-oikeudet.html>

- Yhteiskuntatieteellinen tietoaarkisto (2019, huhtikuu 28). *Tunnisteellisuus ja anonymisointi*. Aineistonhallinnan käsikirja. Viitattu 27.3.2020.
<https://www.fsd.tuni.fi/aineistonhallinta/fi/tunnisteellisuus-ja-anonymisointi.html>
- Yhteiskuntatieteellinen tietoaarkisto (n.d.). *Tietoaarkisto organisaationa*. Viitattu 20.3.2020. <https://www.fsd.tuni.fi/fi/tietoaarkisto/#toimintaperiaatteet>
- Zenk-Möltgen, W., Akdeniz, E., Katsanidou, A., Naßhoven, V. & Balaban, E. (2018). Factors influencing the data sharing behavior of researchers in sociology and political science. *Journal of Documentation*, 74(5), 1053–1073.
<https://doi.org/10.1108/JD-09-2017-0126>
- Zuiderwijk, A. & Spiers, H. (2019). Sharing and re-using open data: A case study of motivations in astrophysics . *International journal of information management*. 49, 228–241. <https://doi.org/10.1016/j.ijinfomgt.2019.05.024>

LIITE 1

Teemahaastattelun kysymykset

Tutkimusaineistot

Kauanko olet työskennellyt tutkijana?

Minkälaista tutkimusta teet?

Millaista tutkimusaineistoa tutkimuksessasi syntyy / millaisia aineistoja käytät?

Ovatko tutkimusaineistosi kvalitatiivisia vai kvantitatiivisia?

Miten hankit aineistoa?

Aineistonhallinta:

Teetkö tutkimuksen alussa aineistonhallintasuunnitelmaa?

- Oletko kokenut sen hyödyllisenä?

Missä säilytät aineistoja tutkimuksen aikana?

Mitä teet aineistoille tutkimuksen jälkeen?

Oletko kohdannut ongelmia aineistojen hallinnassa?

Aineistojen avoimuus

Oletko halukas jakamaan aineistojasi muiden tutkijoiden käytettäväksi?

Miten suuren osan aineistoistasi tallennat avoimesti saataville? Minne ja miten avoimesti?

Mitkä tekijät vaikuttavat päätökseen aineistojen avoimuudesta?

Tutkimuksissa, joissa et avaa aineistoa, mitkä ovat keskeiset syyt?

Mitä hyötyjä näet aineistojen avoimuudessa yleisesti ja itsellesi?

Mitä ongelmia ja esteitä aineistojen avoimuudessa on?

Miten yleistä aineistojen avoimuus on alallasi Suomessa ja kansainvälisesti?

Missä avoimia aineistoja alallasi säilytetään?

Miten tilanne aineistojen avoimuuden suhteen on kehittynyt?

Mitä mieltä olet nykytilanteesta?

Yliopisto ja tutkimusrahoittajat

Asettavatko tutkimusrahoittajat vaatimuksia aineiston hallintaan ja avoimuuteen?

- Onko tämä vaikuttanut käytäntöihisi?

Tuleeko yliopistolta suosituksia aineistojen hallintaan ja avoimuuteen?

Oletko käyttänyt yliopiston/kirjaston tarjoamia palveluja aineistojen hallintaan?

- Mitä mieltä olet näistä palveluista?

Aineistojen jatkokäyttö

Oletko käyttänyt muiden tutkimusaineistoja omassa tutkimustyössäsi?

Millaisia kokemuksia sinulla on tästä?

Millaisia ongelmia tai haasteita aineistojen jatkokäytössä voi olla?