

## Tutkimusaineistot tutkimusmetriikan kohteina

Korkeakouluhallinnon ja -johtamisen  
maisteriohjelma  
Pro gradu -tutkielma  
Maaliskuu 2018  
Ohjaajat: Jussi Kivistö & Reetta Muhonen

Joonas Nikkanen  
<https://orcid.org/0000-0002-5036-6444>

<b>1. Johdanto</b>	<b>1</b>
1.1. Tutkimuskysymys	5
1.2. Työn rakenne	6
<b>2. Tutkimusmetriikka ja tutkimusaineistot</b>	<b>7</b>
2.1. Tutkimusmetriikan hyödyntäminen	12
2.2. Tutkimusmetriikan dysfunktiot	14
2.3. Tutkimusaineiston määritelmä	17
2.4. Tutkimusaineistot osana tutkimustoimintaa	18
2.5. Tutkimusaineistot tutkimusmetriikan kohteina	21
<b>3. Metasynthese</b>	<b>22</b>
3.1. Aineiston haku ja rajaus	24
3.2. Aineiston arviointi ja mahdolliset rajoitteet	26
3.3. Aineiston muodostavat raportit	29
<b>4. Analyysi</b>	<b>35</b>
4.1. Tutkimusaineistojen ominaisuudet	35
4.2. Julkaiseminen	39
4.3. Tekniset valmiudet	43
4.4. Tutkijan toiminta	46
4.5. Arvottaminen	49
4.6. Luokittelu	55
<b>5. Yhteenveto</b>	<b>59</b>
5.1. Tulosten tarkastelu ja tutkimusmetriikan dysfunktiot	59
5.2. Johtopäätökset	63
<b>Lähteet</b>	<b>66</b>
<b>Liitteet</b>	<b>73</b>

## TIIVISTELMÄ

Tampereen yliopisto Johtamiskorkeakoulu, hallintotiede,  
korkeakouluhallinnon ja -johtamisen maisteriohjelma (KOHAMA)

Tekijä: Joonas Nikkanen

Tutkielman nimi: Tutkimusaineistot tutkimusmetriikan kohteina

Pro gradu -tutkielma: 72 sivua + 2 liitettä (3 sivua)

Ohjaajat: Jussi Kivistö & Reetta Muhonen

Aika: Maaliskuu 2018

Avainsanat: tutkimusmetriikka, tutkimusaineisto, tutkimuksen arviointi

---

Tieteellisen tutkimuksen keskeisenä välittäjänä on pitkään toiminut tutkimustulosten julkaiseminen artikkeleina erilaisissa tieteellisissä julkaisusarjoissa. Tieteellinen tutkimus ei kuitenkaan tuota pelkästään tutkimusjulkaisuja, vaan tutkijoiden tekemiä tutkimustuotoksia saattavat olla tutkimusjulkaisujen lisäksi esimerkiksi tutkimuksen pohjana toimivat tutkimusaineistot, tutkimusta ohjaavat tutkimusmenetelmät tai muut merkittävästi tutkijan työtä vaatineet tuotokset.

Tutkimuksen arvioinnissa laajasti hyödynnetty tutkimusmetriikka huomioi tällä hetkellä kattavasti tutkimustuotoksista pääasiassa tutkimusjulkaisut. Tutkimusmetriikan pohjana toimivat matemaattis-tilastolliset analyysit soveltuvatkin tutkimustoiminnan arvioinnissa lähinnä tutkimusjulkaisujen analysointiin. Sen sijaan tutkimusaineistojen tarkastelu tutkimusmetriikan keinoin on vähintäänkin haastavaa, sillä tutkimusmetriikan soveltamisesta tutkimusaineistoille ei ole olemassa vakiintuneita tapoja ja tutkimusaineistojen julkaiseminen poikkeaa tutkimusjulkaisujen perinteisemmistä käytännöistä.

Pro gradu -tutkielmassa tarkastellaan tutkimusaineistoja tutkimusmetriikan kohteina. Tutkielmassa keskitytään erityisesti siihen, onko tutkimusaineistoja mahdollista tarkastella tutkimusmetriikan näkökulmasta ja mitkä tutkimusaineistojen ominaisuudet tai muut tutkimusaineistoihin liittyvät tekijät asettavat haasteita tutkimusmetriikan toteuttamisen kannalta. Tutkielmassa tarkastellaan lisäksi tutkimusaineistojen tutkimusmetriikan mahdollisia dysfunktioita, jotka ovat usein läsnä tuloksellisuuden mittaamisessa.

Tutkielma tuo uutta tietoa paitsi tutkimusmetriikan alalle, mutta laajemmin myös tutkimuksen arviointia toteuttaville tahoille, kuten korkeakouluille, tutkimusrahoittajille ja muille tutkimusjärjestelmän toimijoille. Tutkielman tulosten myötä tutkimusaineistojen tutkimusmetriikkaa koskeviin haasteisiin on tarjolla eri kokonaisuudet havainnoiva luokittelu ja mittaamisen mahdolliset dysfunktiot ja näiden todennäköisyydet on selvitetty.

Tutkielman tutkimusmetodinä toimii metasynteesi, joka on eräs systemaattisen kirjallisuuskatsauksen muodoista. Aineistona tutkielmassa toimivat tutkimusaineistojen tutkimusmetriikkaa uutena ilmiönä pohtivat kansainväliset asiantuntijatyöryhmien raportit ja selvitykset.

# 1. Johdanto

Akateemisessa maailmassa tutkimusjulkaisujen tuottaminen on vakiintunut käytäntö, jolla yksittäiset tutkijat voivat tuoda työtään näkyväksi, kommunikoida muun akateemisen yhteisön kanssa, meritoitua ja kerätä näin arvostusta tutkimusalueensa piirissä. Tieteellinen julkaiseminen on ollut tieteen teon käytäntönä aina 1600-luvulta saakka, jolloin kokeellinen tiede vakiinnutti asemaansa ja ensimmäiset tieteelliset aikakauslehdet, aikansa ainoat julkaisusarjat painettujen kokonaisten kirjojen lisäksi, ryhtyivät julkaisemaan lyhyitä artikkeleita tutkimustuloksista (Lilja 2013).

Siitä lähtien ja katkotta tutkimusjulkaisu, usein vertaisarvioitu artikkeli, on ollut tieteellisen tutkimuksen keskeinen välittämisen mahdollistaja ja tutkimuksen pääasiallinen tuote. Tästä syystä ne kiinnostavat kaikkia tieteellisen tutkimuksen toimijoita: korkeakouluja ja tutkimuslaitoksia, tieteellisiä rahoittajia, poliittisia päättäjiä ja erityisesti yksittäistä tutkijaa - akateeminen meriitti on tyyppillistä yhdistää julkaisuluettelon pituuteen.

Tieteellinen tutkimus ympäri maailmaa on tuottanut viime vuosina hieman arvioista riippuen miljoonasta aina kahteen ja puoleen miljoonaa tutkimusjulkaisua vuodessa (Neylon toim. 2017) ja Suomen osuus tästä tieteellisestä julkaisutoiminnasta on hieman yli 15000 julkaisua vuosittain, mikäli lasketaan vain vertaisarvioidut tieteellisissä julkaisusarjoissa julkaistut artikkelit (julkaisutyypin A1, Vipunen 2017).

Koska julkaisumäärät kipuavat jo miljooniin, on luonnollisena jatkumona syntynyt nimenomaan tutkimusjulkaisujen mittaamiseen ja arvottamiseen erikoistunut suuntaus bibliometriikka, jolla pyritään analysoimaan yhä kasvavaa julkaisujen määrää ja esimerkiksi näissä esiintyviä viittauksia. Bibliometriikkaa on hyödynnetty Suomessa esimerkiksi kansallisesti tutkimuksen arvioinnissa Suomen Akatemian Tieteen tila - katsauksissa (Suomen Akatemia 2014, 2016) sekä julkaisumääriä osana Opetus- ja kulttuuriministeriön vuosittaisen valtionrahoituksen jakamisperusteita korkeakouluille (Opetus- ja kulttuuriministeriö 2015). Toisaalta bibliometriikkaa on sovellettu myös yksittäisten korkeakoulujen sisäisessä rahanjaossa ottaen mallia kansallisesta rahoitusmallista, niin Suomessa kuin ulkomaillakin (Pölönen & Wahlfors 2016, Aagaard 2015).

Akateeminen maailma ja tieteellinen tutkimus eivät kuitenkaan tuota pelkästään tutkimusjulkaisuja, vaan tutkimustoiminnan tuloksena syntyy huomattava määrä myös muuta materiaalia, käytäntöjä ja työkaluja (Neylon toim. 2017). Viime vuosikymmenien kehitys tietotekniikassa ja digitaalisten materiaalien sekä työkalujen saatavuudessa on mahdollistanut tutkijoiden työssä sen, että muiden tutkijoiden tekemää tutkimusta ja niiden tuloksia, aineistoja ja metodeita voidaan helpommin hyödyntämään myös omassa tutkimuksessa. Tutkimusjulkaisujen sijaan olisikin syytä puhua tutkimustuotoksista, sillä tieteenalasta ja tutkimuksen luonteesta riippuen tutkijan tekemien julkaisujen lisäksi akateemisen yhteisön näkökulmasta arvokkaita tuotoksia saattavatkin olla erilaiset tutkimusaineistot tai näiden tulkintaa ohjaavat tutkimusmenetelmät. Lisäksi tutkimustoiminnassa käytetyt ohjelmistot, näiden pohjana oleva lähdekoodi tai muut komponentit voivat olla tutkijayhteisöjen sisällä arvostettuja tuotoksia, jotka eivät kuitenkaan istu perinteistä tutkimusjulkaisuja mittaavaan bibliometriikkaan. Tutkija saattaa myös tutkimustoimintansa ohella toimia erilaisissa yhteisöissä tai perinteisen akateemisen maailman ulkopuolisissa organisaatioissa asiantuntijatehtävissä, joihin nykyinen tutkimusmetriikka ei välttämättä taivu ja tutkijan meriitti jää näin saamatta, vaikka tehtävä on usein oleellinen osa tutkimustoiminnan ydintä - uuden tiedon tuottamista ja tämän tiedon välittämistä.

Tutkimusaineistojen merkittävyyttä suhteessa tutkimusjulkaisuihin on viime vuosina pyritty useammassakin yhteydessä korostamaan sekä kansainvälisesti että Suomessa. Muun muassa Euroopan komission linjauksissa (2010, 7-8; 2011, 2012) viitataan tutkimusaineistojen tulevan muuttamaan tieteellisen tutkimuksen tekoa muuttumattomasti, mikäli vain tutkimusaineistojen saatavuuden, jakamisen ja näiden perustana olevan infrastruktuurin haasteet saadaan ratkaistua. Linjauksissa viitataan jopa tieteellisen tutkimuksen neljänteen paradigmaan, jonka mahdollistajana ovat nimenomaan tutkimusaineistot ja niiden pohjalle perustuva tutkimus (Tolle, Tansley & Hey 2009).

Suomessa tutkimusaineistojen asemaa edistää ja tukee muun muassa Tiedeakatemiain neuvottelukunnan kansalliskomiteaverkoston uusi jäsen, Data-asian kansalliskomitea. Opetus- ja kulttuuriministeriön Data-asian neuvottelukunnalta tilaamassa projektissa (Data-asian kansalliskomitea 2017) on tavoitteena tuottaa suosituksia tutkimusaineistojen viittauskäytäntöihin suomalaisen tutkimusjärjestelmän kontekstissa. Toisaalta Opetus- ja kulttuuriministeriö edistänyt tutkimusaineistojen asemaa myös asettamissaan kansallisissa hankkeissa. Tutkimuksen tietoaaineisto -hanke vuosina 2011-2013 pyrki saattamaan tietovarantoja paremmin tieteen, päätöksenteon ja

yhteiskunnan saataville ja hyödynnettäväksi. Sitä seurannut Avoin tiede ja tutkimus -hanke vuosina 2014-2017 taas keskittyi laajemmin avoimen tieteen ja tutkimuksen edistämiseen, jossa tutkimusaineistot edustavat merkittävää osaa (Avoin tiede ja tutkimus -hanke 2017).

Tutkimustoiminnan laajentuminen on vaikuttanut väistämättä myös korkeakouluihin ja huomioitu siellä muun muassa tutkimushallinnossa. Useissa korkeakouluissa onkin otettu viime vuosikymmenen aikana käyttöön tutkimustietojärjestelmiä, joiden avulla pyritään seuraamaan ja luetteloimaan organisaatiossa tehtyjen tutkimusjulkaisujen lisäksi esimerkiksi tutkimusprojekteja, rahoitusta ja erilaisia tutkijoiden aktiviteetteja kuten konferenssivierailuja, palkintoja ja yhteistyöverkostoihin osallistumisia (Ilva 2014). Myös kansalliset ja kansainväliset tutkimustietoja kokoavat palvelut ovat kehittyneet viime vuosina ja niiden eteen ollaan tekemässä huomattavia panostuksia esimerkiksi kansallisen tutkimustietovarannon käynnistämällä, jonka toteuttaminen aloitettiin 2017 opetus- ja kulttuuriministeriön toimeksiannosta (Opetus- ja kulttuuriministeriö 2017).

Tutkimusmetriikka ja sen hyödyntäminen ovat korkeakoulujen ja korkeakoulututkimuksen näkökulmasta erittäin ajankohtaisia useammastakin eri syystä. Ensinnäkin tieteen teon tutkiminen on korkeakoulututkimuksessa aliedustettuna siihen nähden, että tutkimustoiminta edustaa korkeakoulujen kokonaismenoista huomattavaa osaa yhdessä koulutustoiminnan kanssa (Tight 2012). Myös yliopistojen rahoitusmallissa tutkimustoiminnan perusteella myönnetään 33 prosenttia kaikesta yliopistojen saamasta perusrahoituksesta (Opetus- ja kulttuuriministeriö 2016). Kuten Suomessa, myös useammassa muussa maassa yliopistojen rahoitusmalliin sisältyy jonkinlainen tutkimustoiminnan tuottavuuteen perustuva osa, jossa tutkimusta tarkastellaan ennalta asetettujen indikaattoreiden avulla (Auranen & Nieminen 2010; Hicks 2012).

Tutkimustoiminnan arviointia tukevaa tutkimusmetriikkaa hyödynnetään lisäksi korkeakoulujen sisäisessä ohjauksessa ja strategisessa päätöksenteossa, sillä näin mahdollistetaan tutkimustoiminnan kokonaisvaltaisempi hahmottaminen (Auranen & Nuutinen 2016; Mustajoki 2014). Useassa yliopistossa on lisäksi viime vuosina suoritettu tutkimusmetriikkaan pohjautuvia tutkimuksen kokonaisarviointeja. Tutkimusmetriikkaa hyödynnetään myös laajasti eri korkeakouluja järjestykseen asettavissa ranking-listauksissa (Mustajoki 2010). Korkeakoulujen johdon ja etenkin tutkimushallinnon piirissä on siis vähintäänkin tärkeää ymmärtää ja seurata tutkimusmetriikan kehitystä ja

sen sovelluksia, sillä näillä arvotetaan enenevissä määrin korkeakoulujen ydintoiminnan yhtä kulmakiveä - tutkimustoimintaa.

Korkeakoulututkimuksen perusteoksessa *Researching Higher Education* Malcolm Tight (2012) luokittelee tiedon ja tutkimuksen erääksi kahdeksasta eri korkeakoulututkimuksen vakiintuneesta tutkimusalueesta. Näistä kahdeksasta tutkimusalueesta nimenomaan tiedon ja tutkimuksen tutkimuksen (*researching knowledge and research*) Tight määrittelee monella tapaa fundamentaaleimmaksi ja toisaalta vähiten tutkituksi tutkimusalueeksi korkeakoulututkimuksen kentällä. Vaikka Tight ei suoraan mainitsekaan bibliometriikkaa tai tutkimusmetriikkaa, on hänen esittämänsä peruste tutkimusalueen tärkeydestä kuitenkin sovellettavissa myös tutkimusmetriikalle. Uuden tiedon tuottaminen ja tutkimus ovat korkeakoulujen toiminnan ytimessä sekä akateemisen henkilöstön, tutkijoiden ja opettajien, että opiskelijoiden näkökulmasta ja siksi tämän korkeakoulujen näkökulmasta universaalien toiminnan tutkiminen on välttämätöntä. Tutkimusmetriikka pystyy osin vastaamaan kysymykseen siitä, miten tutkimustietoa hyödynnetään ja millaista näkyvyyttä sillä on. Tightin määrittelemänä tutkimusalalla onkin ensisijaista keskittyä tiedon ja tutkimuksen saatavuuteen, tulkintaan, välittämiseen sekä hyödyntämiseen, ja nämä neljä seikkaa ovat hänen mukaansa korkeakoulujen toiminnan keskiössä.

Tutkimusmetriikka nivoutuu myös tiiviisti osaksi korkeakoulujen muuttumista managerialistisiksi organisaatioiksi, joissa on omaksuttu New Public Management - ajattelun oppeja osaksi organisaatioiden toimintaa (Koski 2002). Vaikka managerialistisella organisaatiolla on tutkimuskirjallisuudessa useita määritelmiä, yhdistää niitä kaikkia tulostavuuksisuuden kasvu osana toimintaa (Teelken 2012, 272; Suomessa managerialismista erit. Lähdesmäki 2003; Salminen 2003). Näin myös korkeakoulut ovat aiempaa kiinnostuneempia panostuksilleen saamista tuotoksista ja Teichler (2003, 177) määrittelee erääksi managerialistisen korkeakoulujen viidestä piirteestä arviointien, raportoinnin ja muiden vastaavien toimenpiteiden roolin kasvun ja niiden määrän lisääntymisen. Nimenomaan raportoinnin ja arvioinnin näkökulmasta tutkimusmetriikka tarjoaakin olennaista tukea sekä korkeakouluhallinnolle että korkeakoulujen ympärillä toimivien toimijoiden päätöksentekoon.

Korkeakoulujen näkökulmasta tutkimusmetriikassa on kyse myös tutkimustoiminnan tuloksellisuuden mittaamisesta. Tuloksellisuutta ja sen suhdetta erityisesti julkisten toimijoiden sekä korkeakoulujen näkökulmasta ovat pohtineet Meklin ja Vakkuri (Vakkuri 1998; Vakkuri & Meklin 2003, 2006). Tuloksellisuuden tarkastelussa on tapana

tarkastella mittaamisen mahdollisia dysfunktioita, joita on tutkimuskirjallisuudessa (esim. Vakkuri & Meklin 2006, 243-244; alkup. Smith 1996, 200-212) esitetty yhteensä seitsemän: tunnelinäkemys, osaoptimointi, lyhytnäköisyys, konvergenssi, tylsistyminen, pelaaminen ja luova laskentatoimi. Tunnelinäkemyksellä viitataan tilanteeseen, jossa tarkastelu kohdistuu kokonaisuuden kannalta vain kapeaan osa-alueeseen; osaoptimoinnissa organisaatioiden ja yksittäisten henkilöiden tavoitteet eivät kohtaa; lyhytnäköisyydessä tarkasteluaika on liian lyhyt suhteessa suoritukseen ja sen mittaamiseen; konvergenssissa organisaatioiden tavoitteet lähentyvät liiaksi muiden organisaatioiden tavoitteita; tylsistymisessä liian jäykkä mittaaminen johtaa luovuuden katoamiseen; pelaamisessa tavoitteet asetetaan liian mataliksi tai niitä täytetään vain raportoinnin takia ja luovassa laskentatoimessa itse raportointia tarkoituksenmukaisesti silotetaan. Tutkimusmetriikan tarkastelun näkökulmasta on oleellista huomioida myös nämä mittaamisesta mahdollisesti juontuvat dysfunktiot.

## 1.1. Tutkimuskysymys

Pro gradu -työni tutkimuskysymys on: "Mitkä tutkimusaineistojen ominaisuudet tai piirteet vaikuttavat tutkimusaineistojen tutkimusmetriikkaan?"

Tutkimuskysymykseen vastatessani tulen käymään läpi tutkimusaineiston erityispiirteitä muihin tutkimustuotoksiin, erityisesti tutkimusjulkaisuihin, verrattuna ja pyrin tunnistamaan sekä luokittelemaan niitä tekijöitä ja ilmiöitä, jotka tällä hetkellä muodostavat haasteita tutkimusaineistojen tutkimusmetriikkaan liittyen. Tavoitteena on luokittelun kautta yhdistellä tutkimusaineistoihin liittyviä käsitteitä ja tunnistaa ilmiöön liittyviä yhteisiä aihealueita tutkimusaineistojen tutkimusmetriikassa.

Tutkimuskysymyksen lisäksi tutkielmassa tarkastelen tuloksellisuuden mittaamisen mahdollisia dysfunktioita, kun tutkimusmetriikan kohteena ovat tutkimusaineistot. Pyrin tutkimuskysymykseen vastatessani huomioimaan tutkimusaineistojen eri ominaisuuksien tai piirteiden vaikutuksia mahdollisiin mittaamisen dysfunktioihin ja havainnoimaan sitä, kuinka todennäköisiä eri dysfunktiot tosiasiassa ovat.

Olen asemoinut tutkimukseni niin, että työni jättää käsittelemättä tutkimusaineistoihin liittyvän tutkimusmetriikan ja sen yhteyden esimerkiksi tutkimusrahoitukseen tai -arviointeihin. En myöskään sisällytä siihen tutkimustoiminnan ohjauksen aspektia. Tarkastelun kohteina ovat siis nimenomaan tutkimusaineistot ja ne huomioiva tutkimusmetriikka, eikä niinkään niistä juontuvat tarpeet eri tutkimusjärjestelmän



toimijoiden toiminnassa. Tutkimus kuitenkin tarjoaa tietoa välillisesti kaikkien edellä mainittujen toimijoiden tarpeisiin mikäli tutkimusmetriikan soveltaminen tutkimusaineistoihin käytäntönä vakiintuu.

Tutkimuskysymykseen vastaaminen tuo kuitenkin merkittävää ja olennaista tietoa erityisesti tutkimusmetriikan alalle, jossa monipuolistuva tutkimustoiminta asettaa yhä enenevässä määrin haasteita kunnollisen ja oikeudenmukaisen metriikan toteuttamisessa. Tutkimusaineistojen osalta on jo otettu ensiaskeleita esimerkiksi indikaattoreiden ja sitaattien huomioimisessa, jonka takia tutkimuskysymykseen vastaaminen on ajankohtaista jo keskeisten käsitteiden ja ilmiöiden tunnistamisen takia. Tutkimusmetriikan näkökulmasta aiheesta ei kuitenkaan ole saatavilla juurikaan tutkimuskirjallisuutta, minkä takia ilmiöön liittyvää tutkimusta on syytä tehdä.

## 1.2. Työn rakenne

Pro gradu -työ käsittelee luvun 2 alkupuolella tutkimusmetriikkaa ja sen sovellutuksia Suomessa sekä kansainvälisesti. Tuon esille ajankohtaista tutkimuskirjallisuutta ja foorumeita, joissa aihetta on käsitelty ja pyrin tätä kautta myös kartoittamaan niitä tutkimusjärjestelmän toimijoita, jotka tavalla tai toisella hyödyntävät tutkimusmetriikkaa toiminnassaan tai sen ohjaamisessa. Luvun 2 loppupuolella tarkastelen tutkimusaineistoja yhtenä tutkimustoiminnan tuotoksista ja asetan niille tämän työn kannalta soveltuvan määritelmän, jotta aiheen käsittely rajautuu ja sitä kautta helpottuu. Luvussa 3 esittelen työn tutkimusmetodina toimivan metasynteessin ja sen valintaan johtaneet perustelut valitun aiheen näkökulmasta. Käyn läpi metasynteessin prosessin ja pro gradu -työtä varten analysoitavaksi valikoituneen aineiston.

Luvussa 4 esittelen, aineiston arvioituani ja vertailun tehtyäni, myös metasynteessin tuottaman synteessin ja tämän työn keskeiset tulokset. Pyrin lisäksi tuomaan esiin niitä tutkimusaineistojen ominaispiirteitä ja rajoituksia jotka asettavat haasteita tutkimusaineistoja koskevaa tutkimusmetriikkaa ajatellen. Luku 5 toimii yhteenvetona ja nivon sen ensimmäisessä alaluvussa yhteen analyysistä ilmenneitä seikkoja tutkimusaineistojen tutkimusmetriikkaan liittyen ja jälkimmäisessä alaluvussa esittelen työn keskeiset johtopäätökset sekä mahdollisia jatkokysymyksiä ilmiön tarkasteluun myöhemmän tutkimuksen näkökulmasta.

## 2. Tutkimusmetriikka ja tutkimusaineistot

On tärkeää huomioida, että esimerkiksi tutkimusaineisto käsitteenä voidaan eri yhteyksissä määritellä hyvinkin eri tavalla riippuen siitä, tarkastellaanko sitä nimenomaan tietyn tieteenalan tai tutkijan näkökulmasta vai yleisemmin akateemisen tutkimuksen tuotoksena. Olen seuraavassa määritellyt tämän työn kannalta keskeisiä käsitteitä ja niiden ulottuvuuksia nimenomaan oman tutkimukseni näkökulmasta tukeutuen kuitenkin aihepiiriin tutkimuskirjallisuuteen sekä aineistonhallinnan asiantuntijoiden tulkintoihin.

Altmetriikka <i>altmetrics</i>	Muistuttaa menetelmiltään bibliometriikkaa, mutta kattaa tutkimusjulkaisujen lisäksi myös monia muita digitaalisia tutkimustuotoksia. Ottaa huomioon esimerkiksi tuotoksen katseluiden määrän, latausmäärät, suositukset, arvioinnit ja keskustelut eri alustoilla, esimerkiksi sosiaalisessa ja perinteisessä mediassa sekä eri yhteistyöalustoilla. (Englund & Forsman 2013; TUHA-tietomallityöryhmä 2017)
Bibliometriikka <i>bibliometrics</i>	Tieteellisen julkaisutoiminnan arviointia tilastollisin menetelmin, jossa tutkimusjulkaisuja arvioidaan esimerkiksi niiden viittausten määrän, vaikuttavuuden tai näkyvyyden kautta. (Andrés 2009; Forsman 2016; Puuska 2014; TUHA-tietomallityöryhmä 2017)
Dysfunktio <i>dysfunction</i>	Tuloksellisuuden mittaamisen yhteydessä käytetty termi, jolla viitataan mahdollisiin arvioinnissa tai mittaamisessa muodostuviin riskeihin tai negatiivisiin vaikutuksiin. (Vakkuri 1998; Vakkuri & Meklin 2003, 2006).
Indikaattori <i>indicator</i>	Luku, jota käytetään kuvaamaan esimerkiksi määrää, vaikuttavuutta tai muuta ominaisuutta. Tutkimusmetriikan kontekstissa tällä viitataan esimerkiksi viittaussmääriin, indekseihin tai muihin lukuihin, jotka pyrkivät kuvaamaan jotakin tutkimustuotoksen ominaisuutta, esimerkiksi vaikuttavuutta tai laatua. (Andrés 2009; Puuska 2014; TUHA-tietomallityöryhmä 2017)

Tutkimusaineisto <i>research data</i>	Tutkimustoiminnassa tuotettu tai käytetty materiaali, jota voidaan käyttää esimerkiksi vastaamaan asetettuun tutkimuskysymykseen. On jalostetumpaa kuin tutkimusdata, joka voi olla esimerkiksi raakadataa mittauslaitteesta tai haastattelun litteroimatonta äänitettä, eikä ole siis tutkimusaineisto sinänsä. (Yhteiskuntatieteellinen tietoaarkisto 2018)
Tutkimusmetriikka <i>scientometrics</i>	Tutkimustoiminnan mittaamista ja arviointia erilaisin, usein tilastollisin tai matemaattisin, menetelmin. Laaja käsite, jonka osaksi voidaan katsoa esimerkiksi bibliometriikka ja altmetriikka. Vakiintunut heikosti suomenkieliseen tutkimuskirjallisuuteen. (Garfield 2009; Suomen yliopistokirjastojen verkosto 2016; Hicks ym. 2015)
Tutkimustoiminta <i>scientific/academic research</i>	Systemaattista ja luovaa toimintaa, joka tähtää olemassa olevan tiedon lisäämiseen. Tämän työn yhteydessä viitataan nimenomaan tutkimustoimintaan, joka tapahtuu akateemisissa tutkimusorganisaatioissa. (TUHA-tietomallityöryhmä 2017)
Tutkimustuotos <i>research output</i>	Tutkimustoiminnan myötä syntynyt tuotos, joka voi olla esimerkiksi tieteellinen artikkeli, monografia, tutkimusaineisto, patentti, ammatillinen teos tai ohjelmisto. (TUHA-tietomallityöryhmä 2017)
Vertaisarviointi <i>peer review</i>	Vakiintunut tiedeyhteisön sisäinen prosessi, jossa tutkijat arvioivat tutkimustuotoksia, usein julkaisuja, oman tieteenalansa riippumattomina asiantuntijoina. (TUHA-tietomallityöryhmä 2017)

Tutkimusmetriikka viittaa “tutkimuksen tutkimukseen” (*science of science*), jonka pääasiallisena tarkoituksena on havainnoida ja analysoida tutkimustoiminnassa syntyneiden tuotosten välisiä suhteita ja niiden kehitystä. Englanninkielinen käsite *scientometrics*, johon tutkimusmetriikalla viitataan, on oikeammin käännös venäläisestä termistä *naukometrija*, jonka matemaatikot Nalimov ja Mulchenko viittasivat olevan

“kvantitatiivisten metodien soveltamista analysoitaessa tiedettä informaatioprosessina” (Andrés 2009, 1-6; Garfield 2009; Wouters 1999, 79-106). Tutkimusmetriikka käsitteenä voidaan ymmärtää laajasti sisältävän kaiken tieteen ja tutkimuksen mittaamisen ja analysoinnin. Vaikka esimerkiksi bibliometriikka ja altmetriikkaa voidaan nähdä sisältyvän tutkimusmetriikan alle, näitä käsitellään hyvin usein suomenkielisessä kontekstissa omina käsitteinä. Tässä työssä tutkimusmetriikalla viitataan kuitenkin laiveammin koko tieteen ja tutkimuksen mittaamiseen.

Metriikka itsessään juontaa juurensa matematiikkaan, jossa mitattavat kohteet ovat joko kvantitatiivisia tai approksimaatioita jostakin kvalitatiivisesta ominaisuudesta. Niiden yhteydessä puhutaan usein indikaattoreista, jotka ovat tutkimusmetriikan tapauksessa yksinkertaisimmillaan kokonaislukuja, esimerkiksi viittausten lukumäärä, mutta monimutkaisimmillaan algoritmeja, joiden avulla isoista julkaisutietokannoista ajetaan läpi analyyseja ja tuotetaan yksittäisiä indikaattoreita päätöksenteon tueksi. Varhaisimmat tutkimukset silloin vielä tutkimusalana tuntemattoman tutkimusmetriikan alalta ovat peräisin 1920-luvulta, jolloin julkaisujen tieteellistä tuottavuutta mallintanut Lotkan laki sai määritelmän Alfred J. Lotkan tutkimuksen tuloksena. Lotkan lain mukaan vain pieni osa tutkijoista tuottaa suurimman osan kaikista tutkimusjulkaisuista. Laajempi tutkimusmetriikan hyödyntäminen mahdollistui kuitenkin vasta 1960-luvulta eteenpäin, jolloin alkunsa sai ensimmäinen viittaustietokanta *Science Citation Index* (Garfield 2009; Puuska 2014, 36-38; Forsman 2016).

Klassisena esimerkkinä tutkimusmetriikasta toimii bibliometriikan viittausanalyysi (*citation analysis*). Näissä analyyseissa aineistona toimivat laajat kansainväliset viittaustietokannat, merkittävimpinä Clarivaten Web of Science, Elsevierin Scopus sekä Googlen Google Scholar, joihin on pyritty keräämään mahdollisimman paljon tutkimusjulkaisuissa ilmeneviä viittaustietoja (Andrés 2009, 141-148; Ellegaard & Wallin 2015; Opetus- ja kulttuuriministeriö 2013). Näihin viittaustietokantoihin kerätään siis yksinkertaisimmillaan tietoja siitä, paljonko yhteen tutkimusjulkaisuun on viitattu toisista tutkimusjulkaisuista ja ketkä näitä viittauksia ovat tehneet. Viittaustietokannan tietoja analysoimalla voidaan muodostaa esimerkiksi yhden tutkijan kaikkien tutkimusjulkaisujen saamien viittausten lukumäärä ja havainnoida näin kyseisten tutkimusjulkaisujen saamaa huomiota. Toisaalta voidaan kiinnostuksesta riippuen myös verrata yksittäisten tutkijoiden sijaan tieteellisten julkaisujen joukkoja osittamalla ne ensin maiden ja tieteenalojen mukaan ja selvittää näin esimerkiksi Top 10 -indeksi, jossa edellä osoitetulle julkaisutoiminnalle annetaan indeksiluku sen perusteella, kuinka suuri osuus keskimäärin ositetusta julkaisutoiminnasta kuuluu tieteenalansa eniten viitattuun

10 prosenttiin (TUHA-tietomallityöryhmä 2017). Tätä nimenomaista indikaattoria käytetään runsaasti esimerkiksi Suomen Akatemian julkaisemissa Tieteen tila -raporteissa (Suomen Akatemia 2014, 2016).

Viittausanalyysi ja muut bibliometriikan sovellutukset ovat olleet käytössä jo 1960-luvulta saakka ja ne ovat kohdanneet ymmärrettävästi myös runsaasti kritiikkiä aina käytetystä lähdeaineistosta ja niiden puutteista (tietokannat), mitattavaan ilmiöön ja sen käyttöön (viittaukset) että sen pohjalta tehtyihin analyyseihin (arvioinnit, ranking-listat). Suomalaisen tutkimuksen kontekstissa bibliometriikkaa ja sen ilmiötä on kuvailtu esimerkiksi Forsman (2016 52-81) sekä summanut yhteen Auranen ja Nuutinen (2016). Toisaalta vastaavat ilmiöt koskevat myös laajemmin koko akateemista tutkimusta, jonka takia myös kansainvälisessä tutkimuskirjallisuudessa on aiheeseen paneuduttu syvällisesti (esim. Andrés 2009; van Drooge ym. 2013; Wallin 2005).

Ongelmaksi ovat muodostuneet lisäksi tieteenalojen väliset erot viittausanalyysien perustana toimivissa tietokannoissa, joissa sosiaalitieteiden ja humanististen tieteiden julkaisut ovat usein aliedustettuina siinä missä luonnon- ja lääketieteiden julkaisut ovat hyvin kattavasti saatu kerättyä tietokantoihin (Andrés 2009, 143-146; Auranen & Nuutinen 2016, 49-50; Opetus- ja kulttuuriministeriö 2013). Tieteenalojen välisiä eroja huomioidessa tulee kuitenkin huomioida myös erot tieteellisen julkaisemisen tavoissa. Humanistisilla aloilla tutkimuksen julkaiseminen monografiassa tai laajemmassa teoksessa on huomattavasti yleisempää kuin luonnontieteissä, joissa julkaiseminen on painottunut nimenomaan julkaisusarjojen artikkeleihin. Ilmiötä on Suomen näkökulmasta pohtinut laajemmin muun muassa Puuska väitöskirjassaan vuodelta 2014 (Puuska 2014, 38-41) sekä erillisessä opetusministeriön tilaamassa tutkimuksessa (Puuska & Miettinen 2008, 65-67, 78-81). Kansainvälisesti eri tieteenalojen välisiä käytäntöjä ja tapoja on pohtinut erityisesti Becher (1989; Becher & Trowler 2001, 1-22), jonka teos *Academic Tribes and Territories* pureutuu nimenomaan tieteelliseen julkaisemiseen ja sen moniin ulottuvuuksiin osana akateemista tutkimusta.

Toinen tietokannoista kumpuava kansallinen ongelma koskee tutkimusjulkaisujen kieltä ja pienten kieliryhmien huomiointia, minkä seurauksena suomalaisten julkaisema englanninkielinen julkaisutoiminta tai mahdolliset yhteisjulkaisut kansainvälisten kumppaneiden kanssa korostuvat viittauksissa usein suomenkielisten julkaisujen rinnalla (Auranen & Nuutinen 2016, 49-50; Glänzel 2001; Opetus- ja kulttuuriministeriö 2013, 33). Tähän syynä on jo pelkästään se, että suomenkielisellä tutkimusjulkaisulla on ylipäätään vähemmän mahdollisuuksia saada viittauksia, sillä potentiaalisia

suomenkieliseen julkaisuun viittaavia tutkijoita on koko maailman mittakaavassa vähän verrattuna englanninkielisen julkaisun mahdollisiin hyödyntäjiin (Markkanen 2006, 37-38; Taavitsainen 2004, 13-18; Puuska & Miettinen 2008, 26-27). Tutkimusrahoituksen näkökulmasta ongelmaan on puututtu Julkaisufoorumi-luokituksella, jossa pyritään tasapäistämään julkaisujen arvoa julkaisukielestä riippumatta. Tutkimuksen ja siitä tehtävien julkaisujen väliseen kielten epätasaiseen kohteluun on kuitenkin aivan viime aikoina puututtu aina Helsingin Sanomia myöten, jossa ongelman laajuutta on sekä korostettu (Helsingin Sanomat 2018a, 2018c) että vähätelty (Helsingin Sanomat 2018b, 2018d).

Näistä tunnistetuista puutteista huolimatta bibliometrisiä analyyseja käytetään laajasti joko yhtenä tutkimuksen arvioinnin menetelmistä tai harvoissa tapauksissa myös ainoana menetelmänä (Forsman 2016, 119-125; Mustajoki 2014).

Tieteellinen tutkimustoiminta ei monipuolisuudestaan johtuen kuitenkaan istu ongelmitta määrällisiin tutkimusmetriikan indikaattoreihin. Tästä syystä esimerkiksi bibliometriikan soveltaminen pelkkiin tieteellisiin julkaisuihin "saattaa olla vaarallista, ja pahimmillaan johtaa väärin tulkintoihin ja johtopäätöksiin tieteellisen tutkimuksen arvioinneissa esimerkiksi rahoituspäätöksiä koskien" kuten Leidenin manifestossa kuvataan (Hicks ym. 2015). Vaarassa ovat erityisesti uudet ja perinteisestä tieteenalajaottelusta pesäeroa tehneet tutkimusalat, jotka eivät välttämättä istu vakiintuneeseen, yksittäisiä julkaisuja suosivaan, tutkimusjulkaisumalliin. Tällaisten tutkimusalojen tutkijoilla tuotokset saattavat olla enenevässä määrin tutkimusaineistoja, innovaatioita, ohjelmistoja ja koodia sekä tärkeimmät yhteistyökumppanit huomattavasti lähempänä elinkeinoelämää kuin perinteikästä akateemista yhteisöä (Neylon toim. 2017).

Tutkimusmetriikkaan liittyen on Suomessa keskitytty tutkimuskirjallisuudessa erityisesti 2010-luvun puolella, jolloin aiheeseen liittyen on ilmestynyt useampi väitöskirja (Auranen 2014; Puuska 2014) ja lukuisia artikkeleita eri julkaisusarjoissa, joista suuri osa toimii tämänkin työn lähteinä. Vuodesta 2014 lähtien on myös järjestetty vuosittaista Bibliometriikkaseminaaria, johon ovat kokoontuneet niin bibliometriikkapalveluita tarjoavat toimijat, kuin niiden hyödyntäjät sekä yliopistoissa että muissa organisaatioissa. Nimestään huolimatta bibliometriikkaseminaareissa on keskitytty laajemmin koko tutkimusmetriikan kirjoon, muun muassa altmetriikkaan useampana vuonna.

## 2.1. Tutkimusmetriikan hyödyntäminen

Alun perin bibliometrinen tutkimusten kohdeyleisön määrittelyyn tarkoitettu Glänzelin (2003) kolmiportainen malli soveltuu myös laajemmin tutkimusmetriikasta saatavan tiedon hyödyntäjien määrittelyyn. Ensimmäisellä tasolla ovat Glänzelin mukaan bibliometriikan soveltamiseen liittyvät kysymykset, jossa tarkastelu koskee bibliometrisiä metodeita ja matematiikkaa niiden takana. Toisella tasolla ovat tieteen ja tutkimusalojen kehityksestä kiinnostuneet, jotka pyrkivät selvittämään eri alojen syntyä ja kehitystä sekä kartoittamaan tieteen leviämistä. Kolmannella, ja Glänzelin mukaan myös tärkeimmällä, tasolla ovat tieteellisen tutkimuksen ohjaukseen ja tiedepolitiikkaan liittyvät tahot (Glänzel 2003). Siinä missä kaksi ensimmäistä tasoa omaavat suhteellisen pienen, lähinnä tutkimusalan sisäisen, kohdeyleisön on kolmannessa tasossa laajasti tutkimuksen arvioinnista ja ohjauksesta kiinnostuneita toimijoita sekä korkeakouluissa että niiden ulkopuolella (Puuska 2014, 37-38). Seuraavassa esittelen tutkimusmetriikan hyödyntämistä nimenomaan tämän kolmannen tason näkökulmasta.

Tutkimusmetristen analyysien käyttö on kasvanut viime vuosikymmeninä ja tähän on eittämättä vaikuttanut useat korkeakouluja ja julkista toimintaa koskevat pidemmän aikajänteen muutokset sekä organisaatioiden sisällä, että niiden toimintaympäristössä. Oleellisimpana muutoksen kannalta on ollut organisaatioiden siirtyminen kohti managerialismia ja vaatimus toimintaan sijoitettujen panostusten tuotoksista ja erityisesti näiden raportoinnista eri tahoille. Raportoinnin lisääntymisen lisäksi toiminnan läpinäkyvyyttä edellytetään sekä yhteiskunnan että tutkimustoiminnan pääasiallisen rahoittajan, valtion, suunnalta (Lähdesmäki 2003; Salminen 2003). Esimerkkinä toimii hyvin nykyinen korkeakoulujen rahoitusmalli, jossa tutkimuksen perusteella jaetaan huomattava osa yliopistojen perusrahoituksesta (Opetus- ja kulttuuriministeriö 2015). Tulostavuuksien myötä myös korkeakouluhallintoon raportoidaan aiempaa enemmän tutkimustoimintaan liittyviä tunnuslukuja, sillä ne edustavat merkittävää osaa korkeakoulujen ydintoiminnasta. Juuri raportoinnin näkökulmasta tutkimusmetriikka tarjoaa useita mahdollisuuksia tutkimustoiminnan hahmottamiseen, analysoimiseen sekä mahdolliseen arvottamiseen (Auranen & Nuutinen 2016; Mustajoki 2014; Puuska 2014).

Tutkimusmetriikan hyödyntämistä on vauhdittanut eittämättä myös tutkimusorganisaatioiden lisääntyvä tarve ohjata toimintaansa ja kehittää tutkimustoimintansa tueksi soveltuvaa strategiaa, jolloin indikaattorit tarjoavat korkeakoulujen johdolle mahdollisuuksia tehdä aiempaa enemmän tietoon johdettavia ja

siihen perustuvia päätöksiä. Tällaisia ovat esimerkiksi yliopistojen tuottamat tutkimuksen kokonaisarviointit, joiden yhtenä osiona ovat nimenomaan bibliometriset analyysit organisaation tutkimusjulkaisuista. Tutkimuksen kokonaisarviointeja on tehty viime vuosina useimmissa suuremmissa suomalaisissa yliopistoissa (esim. Aalto-yliopisto 2009; Helsingin yliopisto 2012; Oulun yliopisto 2013; Tampereen teknillinen yliopisto 2017; Turun yliopisto 2015).

Korkeakoulut ovat myös lisääntyvissä määrin kiinnostuneita käyttämään tutkimustoimintansa arvioinnin tukena altmetriikan tarjoamia menetelmiä. Altmetriikassa tarkastellaan perinteisempien tutkimusjulkaisujen sijaan esimerkiksi tutkijan sosiaalisessa mediassa tapahtuvaa viestintää ja sen määrää, aktiviteetteja Wikipedian tai muun vastaavan yhteisöpalvelun päivittämisessä sekä eri uutispalveluissa tehtyjä viittauksia tutkijan työhön. Avoin tieteiden toimintakulttuuria arvioineen raportin (Avoin tiede ja tutkimus -hanke 2016) mukaan suomalaisista yliopistoista yhtä lukuunottamatta kaikki ja ammattikorkeakouluistakin suurempi osa aktiivisesti käyttää tai on suunnittelemassa altmetriikkaa tarjoavien palveluiden käyttöä oman organisaationsa tutkimustoimintansa seuraamisessa.

Osin tutkimusmetriikkaan liittyvästä kiinnostuksesta, osin lisääntyneestä raportoinnin tarpeesta johtuen tutkimustoiminnan hallintaan ja arviointiin on kehitetty vuosien saatossa yhä kehittyneempiä tutkimustietojärjestelmiä (*current research information system*). Ne ovat käytännössä yksittäisten korkeakoulujen ylläpitämiä organisaatiokohtaisia tietokantoja tai -varantoja, joihin lisätään jatkuvasti tietoja mahdollisimman monesta kyseisen korkeakoulun tutkimukseen juontuvasta toiminnasta. Yleisimpiä lisättyjä tietoja ovat perinteikkäästi tutkimusjulkaisut, mutta tutkimustietojärjestelmiin saatetaan tallentaa tietoja myös tutkijan osallistumisista eri konferensseihin, tutkijan toimimisesta osana eri tutkimusryhmiä tai projekteja sekä tietoja tutkijan saamasta tutkimusrahoituksesta. Suomessa tutkimustietojärjestelmiä on käytössä pääasiassa yliopistoissa, joissa käytettyjä tuotteita ovat joko Elsevierin Pure (käytössä esimerkiksi Helsingin yliopistossa ja Oulun yliopistossa), Clarivaten Converis (Jyväskylän yliopisto) tai kotimainen SoleCris (Tampereen yliopisto) (Ilva 2014).

Kansallisella tasolla tutkimustietoon on suunnitteilla valtakunnallinen tutkimustietovaranto, jonka pääasiallisena tavoitteena on koota yhteiseen tietovarantoon tietoa tutkimustoiminnasta ja -tuotoksista Suomessa (Opetus- ja kulttuuriministeriö 2017). Työ tutkimustietovarannon kehittämiseen on alkanut vuonna 2017 opetus- ja kulttuuriministeriön toimeksiannosta. Sen toteutuksesta vastaa CSC - Tieteen



tietotekniikan keskus Oy, joka ylläpitää kansallisella tasolla jo esimerkiksi VIRTJA-julkaisutietopalvelua, jonne on kerätty vuodesta 2010 lähtien suomalaisten tutkimusorganisaatioiden tutkimusjulkaisujen metatietoja. Kansalliset tutkimustiedon tietovarannot tarjoavat laajemmat mahdollisuudet tutkimuksen analysointiin ja tutkimusmetriikkaan paitsi kansallisesta näkökulmasta, myös vertailun eri tutkimusalojen ja -organisaatioiden kesken.

Tutkimusmetriikan tarjoamien indikaattoreiden käyttöä suositaan myös kansainvälisellä tasolla korkeakouluja vertailevissa niin sanotuissa ranking-listoissa, joiden yhtenä osana tutkimustoimintaa tarkastellaan yleensä muutamien bibliometrinen avainlukujen kautta (Wilsdon ym. 2015). Ranking-listoista muun muassa Shanghai-lista käyttää huomattavaa määrää nimenomaan tutkimusjulkaisuihin liittyviä bibliometrisia indikaattoreita, joissa suositaan aivan tutkimusalan huipulla tapahtuvaa julkaisemista. Toisaalta myös Times Higher Education -lista sekä siitä erkaantunut QS-lista perustavat organisaatioiden arvottamisen laajasti bibliometrisiin analyyseihin tutkimuksen vaikuttavuudesta. Tunnetuista ranking-listoista ainoastaan Taiwanin lista sekä Leidenin tuottamat arvioinnit perustavat tutkimuksen vaikuttavuuden arvioinnissa myös laadullisiin arvioihin organisaatiossa tuotetusta tieteellisestä tutkimuksesta (ranking-listoista tarkemmin ks. Mustajoki 2010, 2014).

Tutkimusmetriikasta ja sen käyttämistä indikaattoreista ei siis olla luopumassa, vaan käytön voidaan katsoa lisääntyvän enenevässä määrin yhä helpommin saatavilla olevien tietovarantojen ja lähestyttävempien ohjelmistojen myötä, jotka ovat vaivanneet tutkimusmetriikan omaksumista laajemmassa mittakaavassa (Muhonen & Puuska 2014; Puuska 2014). Tutkimustoiminnan suuresta volyyymistä johtuen indikaattoreista luopuminen ei myöskään ole mielekäästä, sillä tutkimuskirjallisuudessa usein todetaan että tieteellisen tutkimuksen arviointiin tarvitaan joka tapauksessa kokoavia indikaattoreita (Neylon toim. 2017). Laajamittaisen ja läpinäkyvän määrällisten indikaattoreiden käytön nähdään myös vahvistavan muutosta kohti avoimempaa tutkimusjärjestelmää. Tällöin indikaattorien tulisi tukea, ei ohjata tai syrjäyttää, laadullista tutkimustoiminnan arviointia (Auranen & Nuutinen 2016; Wallin 2005; Wilsdon ym. 2015).

## 2.2. Tutkimusmetriikan dysfunktiot

Vaikka tutkimusmetriikka seuraa ja tarkastelee pitkälti tutkimustoiminnan tuloksia, on kyseessä laajemmin nähtynä tapa tarkastella tutkimustoiminnan tuloksellisuutta osana tutkimusjärjestelmää, tutkimusorganisaatiota tai yksittäistä tutkimusryhmää, jossa

tutkimusta tehdään. Tutkimusmetriikan sovellutuksissa, esimerkiksi tutkimuksen kokonaisarvioinneissa tai ranking -listoissa, kiinnitetään nimittäin usein huomiota tutkimuksella tuotettujen tuotosten lähtötilanteeseen eli tuloksellisuuden näkökulmasta tutkimukseen asetettuihin panoksiin. Tutkielman ja asetetun tutkimuskysymyksen kannalta onkin tarpeen huomioida tutkimusmetriikan mahdollisia dysfunktioita yleisesti tutkimusmetriikan tasolla. Seuraava jaottelu noudattaa tutkimuskirjallisuudessa vakiintunutta tuloksellisuuden tarkastelun jakoa seitsemään mahdolliseen dysfunktiioon, joihin tuloksellisuuden mittaaminen voi johtaa (Vakkuri & Meklin 2006, 243-244; Smith 1996, 200-212).

Dysfunktioista ensimmäinen, tunnelinäkemys, viittaa tilanteeseen, jossa tarkastelu ja mittarit asetetaan sen mukaan, mistä on saatavilla eniten tai helpoiten tietoa tarkastelun tai mittarien tueksi. Vaarana on tällöin se, että tuloksellisuutta arvostellaan kokonaisuuden kannalta epäolennaisia asioita tai ilmiöitä. Tutkimusmetriikan kannalta nimenomaan tunnelinäkemys dysfunktiona on todennäköinen, sillä bibliometriikan perustana olevat viittausanalyysit ovat saaneet kritiikkiä nimenomaan siitä, että tarkastelun perustana ovat usein tutkimusjulkaisujen väliset viittaukset, joiden perusteella ei voida tehdä suuremmin johtopäätöksiä siitä, onko viittaus tehty esimerkiksi yhteisymmärryksen merkiksi vai kritisoidaanko viittauksen kohteena olevaa tutkimusjulkaisua (tarkemmin kts. Auranen & Nuutinen 2016).

Osoptimoinnissa on kyse siitä, että arvioinnin toteuttaja ohjaa toimijoita kohti asetettuja mittareita, eikä lopputuloksessa välttämättä toteudu koko järjestelmän etu. Tutkimusmetriikan näkökulmasta dysfunktiioon voidaan korkeakoulujen tapauksessa törmätä esimerkiksi siinä tilanteessa, että asetettu rahoitusmalli tai tieto arviointikriteereistä johtaa julkaisuaktiivisuuden kasvuun julkaisujen laadun kärsiessä. Toistaiseksi tällaista toimintaa ei ole Suomessa havaittu (Pölönen & Auranen 2017), mutta mahdollisuus tähän dysfunktiioon on kuitenkin läsnä (esim. Karvonen & Kortelainen & Saarti 2014).

Lyhytnäköisyyden riskiä lisää se, että asetettu tarkastelujakso on liian lyhyt suhteessa tulosten saavuttamisen kannalta. Tutkimusmetriikassa tilanne voi syntyä esimerkiksi sellaisessa tapauksessa, että tutkimusjulkaisuja mahdollisille viittauksille ei suoda tarpeeksi aikaa kertyä, vaan analyysit tehdään hyvinkin tuoreista tutkimusjulkaisuista. Esimerkiksi bibliometriikan näkökulmasta tuorein saatavilla oleva aineisto on analysoitaessa jo vuoden tai kaksi vanhaa, jolloin merkittävä osa julkaisun saamista viittauksista on usein saatu. Tilanne kuitenkin vaihtelee merkittävästi eri tieteenaloilla,

eikä ole lainkaan poikkeuksellista, että tutkimusjulkaisu kerää vielä viidenkin vuoden päästä merkittävästi viittauksia (Andrés 2009; van Drooge yms. 2013).

Eräs tutkimusmetriikan näkökulmasta monitulkintainen dysfunktio on konvergenssi, jolla viitataan tilanteeseen, jossa tuloksellisuuden arviointi ohjaa toimijoita yhteneväiseen tekemiseen muiden vastaavien toimijoiden kanssa. Toisaalta tutkimusmetriikan perusideana on tarkastella kohteena olevaa tutkimustoimintaa osana laajempaa kokonaisuutta, jolloin konvergenssia ilmiönä voi olla vaikeaa havaita tai tunnistaa. Toisaalta on mahdollista, että tutkimuksen arvioinneissa esimerkiksi englannin kielen pääsääntöisesti laajempi hyödyntäjien joukko ohjaa suomalaisia tutkijoita julkaisemaan tutkimuksensa etupäässä englanniksi. Viitteitä tällaisesta toiminnasta on olemassa (esim. Markkanen 2006; Taavitsainen 2004), joten tästä näkökulmasta konvergenssi dysfunktiona on läsnä myös tieteellisessä tutkimuksessa.

Tylsistymisellä viitataan tilanteeseen, jossa organisaatiossa tuloksellisuuden mittaamisella tehdään toiminnasta liian jäykkää, eikä esimerkiksi luovalle toiminnalle jätetä sijaa. Yleistä tälle dysfunktiolle on se, että toiminnassa ei enää kyetä toimimaan ketterästi, vaan tehdään mittaamisen mukaista toimintaa vakiintuneiden käytäntöjen mukaisesti. Tähän nimenomaiseen asiaan on ottanut kantaa muun muassa Mustajoki (2014) tarkastellessaan tutkimusmetriikkaa ja sen vaikutuksia tutkijan työhön. Laajemmin katsottuna dysfunktio voi olla läsnä myös laajemmassa mittakaavassa, jolloin esimerkiksi perinteikkäästä tieteellisen julkaisemisen mallista ei päästä eteenpäin, sillä tutkimusmetriikka sekä tutkijoiden rahoitus ovat siinä määrin kiinnittyneitä vakiintuneisiin tapoihin arvioida tieteellistä tutkimusta.

Bibliometriikan ja laajemmin tutkimusmetriikan alalla on pohdittu nimenomaan pelaamista eräänä mittaamisen dysfunktioista. Dysfunktiossa on kyse siitä, että asetettu mittaaminen ohjaa toimijoita kohti kokonaisuuden kannalta epäedullista toimintaa. Tilanteeseen, jossa tutkimustuloksia pyritään pilkkomaan mahdollisimman moneen tutkimusjulkaisuun viitataan termillä *salami slicing*, jossa tutkijan tarkoituksena on jakaa omat tutkimustulokset pienimpään mahdolliseen julkaisuyksikköön (*least publishable unit*) tutkimusmetriikan ja tutkimusrahoituksen arvioinnin ohjaamana (Tammi 2014).

Luovalla laskentatoimella viitataan tilanteeseen, jossa raportoinnissa tai tulosten ilmoittamisessa tarkoituksenmukaisesti parannellaan asetettujen mittareiden tai arvioinnin mukaan. Luovan laskentatoimen mahdollisuus dysfunktiona on

tutkimustoiminnan sekä tutkimusmetriikan näkökulmasta verrattain pieni. Laajasti ajateltuna esimerkiksi tiedevilppi voi kuitenkin asettua luovan laskentatoimen piiriin, jossa tutkimuksen tuloksia tietoisesti muokkaamalla saavutetaan etuja omalle tutkimukselle tai tutkimusryhmälle.

Tuloksellisuuden mittaamisen eri dysfunktiot soveltuvat pääasiassa hyvin myös tieteellisen tutkimuksen arviointiin sekä tutkimusmetriikkaan laajemmin. Niiden avulla pystytään havainnoimaan ja luokittelemaan mahdollisia mittaamisen myötä ilmenneitä negatiivisia vaikutuksia sekä tutkimustoiminnan arvioinnissa että itse tutkimustoiminnassa. Tutkielmassa hyödynnän näitä dysfunktioita metasynteesin muodostamien tulosten luokittelussa sekä tarkastelen tutkimusaineistoja koskevan tutkimusmetriikan eri ilmiöitä niiden kautta.

### 2.3. Tutkimusaineiston määritelmä

Tutkimusaineistolle ei varsinaisesti voi asettaa yksiselitteistä määritelmää joka rajaisi sen tarkasti omaksi entiteetiksi irralleen muusta tutkimustoiminnasta tai sen tuotoksista. TUHA-tietomallityöryhmä on määritellyt tutkimusaineiston seuraavasti: “tutkimuksessa tuotettava ja/tai käytettävä materiaali, jota analysoimalla pyritään vastaamaan asetettuun tutkimuskysymykseen” (TUHA-tietomallityöryhmä 2017). Määritelmässä korostuu tutkimusaineiston käyttö tutkimuksessa ja tarve asettaa se analysoitavaksi tutkimuskysymykseen vastattaessa. Yhteiskuntatieteellinen tietoarkisto korostaa toisaalta tutkimusaineiston yhteydessä kuvailevan tiedon (metatieto tai -data) merkitystä, ilman sitä tutkimusaineistot tutkimusaineisto “on vain merkityksetön kokoelma yksittäisiä tiedostoja, numeroita ja merkkejä” (Yhteiskuntatieteellinen tietoarkisto 2018).

Käsitteen määritelmästä hankalan tekee myös Suomen kieli. Englanniksi “*research data*” voi tarkoittaa hyvin suurta kirjoa erilaista materiaalia aina mittauslaitteen raakadatasta laajoihin ja asiallisesti kuvattuihin aikasarja-aineistoihin niitä varten perustetussa arkistoissa. Suomen kielessä tutkimusaineisto taas usein viittaa jalostusasteeltaan nimenomaan sellaisiin aineistoihin, joita on jo analysoitu ja valmisteltu esimerkiksi kuvailun kautta valmiiksi jatkokäyttöä ajatellen. Prosessoimattomista ja vähemmän jalostetuista materiaaleista puhutaankin usein raakadatana tai tutkimusdatana (Yhteiskuntatieteellinen tietoarkisto 2018).

Tässä työssä tutkimusaineistolla tarkoitetaan pääasiallisesti niitä julkaistuja tutkimusaineistoja, joihin on jo liitetty kuvailevaa tietoa esimerkiksi omistajuudesta tai

aineistonkeruusta ja jotka on valmisteltu jatkokäyttöä ajatellen saattamalla ne saataville johonkin tutkimusaineistojen varten perustettuun arkistoon. Nämä seikat ovat myös välttämättömiä tutkimusmetriikan näkökulmasta, ilman tietoa tutkimusaineiston tekijyydestä tai pääsyä tutkimusaineiston äärelle ei voida aidosti varmistaa tutkimusmetriikan kannalta olennaisia tietoja nimenomaan tutkimusmetristen analyysien kannalta.

## 2.4. Tutkimusaineistot osana tutkimustoimintaa

Tutkimusaineistot ovat tutkimustoiminnan kannalta merkittäviä, sillä niiden pohjalta tutkija usein tekee asetetun tutkimuskysymyksen kannalta merkittäviä johtopäätöksiä tutkittavasta ilmiöstä tai asiasta. Tutkimusaineistot tarjoavat tutkimusjulkaisujen tavoin muille, esimerkiksi alan tutkijoille, mahdollisuuden hyödyntää sekä jatkokäyttää jo kerättyjä tutkimusaineistoja ja näin voidaan mahdollisesti välttää osin päällekkäistä tutkimustyötä. Tutkimusaineistojen elinkaari on myös verrattain pitkä, sillä moni tutkimus alkaa tutkimusaineiston keruulla ja sen käsittelyllä sekä kuvailulla (Yhteiskuntatieteellinen tietoarkisto 2018). Vasta tämän jälkeen tutkimukseen liittyvä tutkimusjulkaisu voidaan alkaa perusteellisemmin kirjoittamaan sekä tekemään analyysia tai yhteenvetoa tuloksista. Tutkimushankkeen loppumisen jälkeen tutkimusaineisto on myös parhaassa tapauksessa hyödynnettävissä pitkään varsinaisen tutkimushankkeen päätyttyä, mikäli aineistonhallinta on toteutettu asiallisesti ja tutkimusaineisto avattu tai arkistoitu (Avoin tiede ja tutkimus -hanke 2015b).

Tutkimusaineistojen jatkokäyttö edellyttää asiallista aineistonhallintaa, jolla tarkoitetaan tutkimusaineistojen sekä niihin liittyvän kuvailevan tiedon luomista ja järjestämistä. Aineistonhallinnalla varmistetaan, että tutkimusaineisto pysyy käytettävänä ja luotettavana jatkokäyttöä ajatellen sekä pidentää näin huomattavasti tutkimusaineiston elinkaarta. Aineistonhallintasuunnitelmalla varmistetaan myös, että tutkimuksessa tuotetuille tutkimusaineistoille on mietittynä niille soveltuva arkistointitapa aineistojen jatkokäyttöä ajatellen, mikä mahdollistaa esimerkiksi viittaukset tutkimusaineistoihin. (Avoin tiede ja tutkimus -hanke 2015b; Yhteiskuntatieteellinen tietoarkisto 2018) Tarkemmin tutkimusaineistoja ja niiden käytäntöjä tutkijan näkökulmasta esimerkiksi julkaisemiseen liittyen on esitetty Liitteessä 1.

Suomessa aineistonhallinnan edistämiseen liittyvää työtä on tehty esimerkiksi Yhteiskuntatieteellisessä tietoarkistossa (Yhteiskuntatieteellinen tietoarkisto 2017) sekä osana Avoin tiede ja tutkimus -hanketta (Avoin tiede ja tutkimus -hanke 2015b).

Saatavilla on myös kansallinen aineistohallintatyökalu DMPTuuli, jota tutkijat voivat hyödyntää aineistohallintasuunnitelman laadinnassa esimerkiksi tutkimusrahoitusta hakiessaan. Yhä useampi tutkimusrahoittaja edellyttääkin aineistohallintasuunnitelmaa osana rahoitushakemuksia, merkittävimpinä näistä kansallisesti Suomen Akatemia (Suomen Akatemia 2017) sekä Euroopassa Horizon 2020 tutkimusrahoitus (Euroopan komissio 2017). Kansainvälisesti tutkimusrahoittajista tutkimusaineistojen arkistointia ja täten myös aineistohallintasuunnitelmaa edellyttää tai suosittaa tätä tietoa seuraavan Sherpa Juliet -palvelun tietojen mukaan 42 prosenttia kaikista tutkimusrahoittajista vuonna 2017 (Sherpa Juliet 2017).

Myös korkeakoulut suhtautuvat, osin opetus- ja kulttuuriministeriön asettamien tavoitteiden mukaisesti (Opetus- ja kulttuuriministeriö 2014), tutkimusaineistoihin aiempaa suuremmalla mielenkiinnolla. Hieman yli puolella suomalaisista yliopistoista on asetettu jonkinlainen vaatimus tutkimusaineistojen julkaisemiseen osana tutkimustoimintaa (Avoin tiede ja tutkimus -hanke, 2016), ja esimerkiksi Helsingin yliopiston tutkimusdatapolitiikassa mainitaan tutkimusaineistoihin liittyen seuraavaa: "Helsingin yliopiston piirissä tuotettu, julkaistuihin tutkimustuloksiin liitetty tutkimusdata on lähtökohtaisesti yhteiskäyttöistä ja avointa" (Helsingin yliopisto 2015). Näiden suhteen tulee kuitenkin huomioida, että korkeakouluilla on usein vähän tai ei lainkaan mahdollisuuksia seurata itse asettamiaan vaatimuksia tutkimusaineistojen julkaisemiseen liittyen (Avoin tiede ja tutkimus -hanke 2015a; Nykyri 2017). Näin ollen asetetut politiikat saattavat jäädä vaille kytköstä käytännön tutkimustoimintaan organisaatioissa.

On myös hyvä huomata, että tutkimusaineistot ja niiden julkaiseminen liittyvät tiukasti osaksi hyvää tieteellistä käytäntöä, sillä ne ovat merkittävässä osassa usean tutkimuksen toteuttamista. Suomessa hyvää tieteellistä käytäntöä ohjaa Tutkimuseettinen neuvottelukunta, jonka ohjeissa (TENK 2012) todetaan, että tutkimuksen tulee paitsi soveltaa tieteellisen kriteerien mukaisia ja eettisesti kestäviä tiedonhankinta- ja tutkimusmenetelmiä, mutta niissä myös edellytetään tutkimuksessa syntyneiden tietoaineistojen tallennusta tieteelliselle tiedolle asetettujen vaatimusten mukaisesti. Hyvää tieteellistä käytäntöä noudatettaessa tutkimusaineistojen aineistohallinnasta tulisi siis huolehtia samalla tarkkuudella kuin muustakin tutkimukseen liittyvästä toiminnasta.

Kun sekä tutkimusrahoittajat, korkeakoulut että hyvän tieteellisen käytännön ohjeistus velvoittavat tutkimusaineistojen julkaisemista, mitä mahdollisuuksia tutkijalla on toimia näiden velvoitteiden mukaisesti ja miten tilanne eroaa esimerkiksi tutkimusjulkaisujen osalta?

Tutkimusaineistojen julkaisemisen kehitys on suuntautunut kahteen tutkimusjulkaisujen julkaisemisesta osittain perittyyn malliin. Molemmat mallit ovat askel eteenpäin tilanteesta, jossa tutkimusaineistoihin viitattiin tutkimusjulkaisussa vain mainitsemalla aineiston nimi ja esittelemällä siitä tunnuslukuja tai tuloksia esimerkiksi taulukkojen muodossa. Tällä tavoin yhteys tutkimusjulkaisun perustana olleeseen tutkimusaineistoon jää kuvailun tasolle, eikä tutkimusaineistoa näin ollen pääse esimerkiksi hyödyntämään tai jatkokäyttämään uudessa tutkimuksessa.

Lähimpänä tutkimusjulkaisujen perinteikstä julkaisemista ovat tutkimusaineistojen julkaisemiseen keskittyneet tieteelliset julkaisusarjat (*scientific data journal*), joista esimerkkeinä *Data in Brief* Elsevieriltä tai *Scientific Data* Naturelta. Näihin, usein hyvin tieteenalakohtaisiin, julkaisusarjoihin tutkijan on mahdollista lähettää oma tutkimusaineistonsa ja sitä kuvaava lyhyt tutkimusaineiston keruuseen, rajoitteisiin tai muuhun seikkaan liittyvä teksti. Julkaisusarjasta riippuen julkaisun oheen on rakennettu erilaisin palveluin esimerkiksi mahdollisuus aineiston tarkempaan tutustumiseen ja jatkokäyttöön. Näin toimien tutkimusaineistoista muodostuu perinteikkään artikkelin muotoinen julkaisu, johon voidaan viitata tavanomaisen tutkimusjulkaisun tapaan.

Toinen tapa julkaista tutkimusaineistonsa on sijoittaa ne johonkin tutkimusaineistoille tarkoitettuihin julkaisuarkistoihin (*data repository* tai *data archive*), esimerkkinä kansainvälinen Zenodo tai kansallinen Yhteiskuntatieteellinen tietoarkisto. Tällöin kuvaileva tieto liitetään tutkimusaineiston yhteyteen ja tutkimusaineisto on suoraan jatkokäytettävissä julkaisuarkistosta esimerkiksi mahdollisuuten ladata se suoraan omaan käyttöön. Näiden arkistojen yhteyteen on usein rakennettu muitakin jatkokäyttöä paremmin palvelevia toimintoja esimerkiksi aineiston versiointiin, kommentointiin ja muutoksiin liittyen. Tässä mallissa viittaus tutkimusaineistoon tapahtuu digitaalisen tunnusteen (DOI, *digital object identifier*) avulla eikä erillistä tutkimusjulkaisua tarvita tutkimusaineiston julkaisemisen tueksi.

## 2.5. Tutkimusaineistot tutkimusmetriikan kohteina

Tutkimusmetriikan kannalta juuri tutkimusaineistojen julkaiseminen ja mahdollinen avaaminen ovat kriittisiä, sillä ilman niitä tutkimusaineistosta ei välity jälkeä tietokantoihin. Ei ole lainkaan harvinaista, että tutkija on tallentanut tutkimusaineistonsa lähinnä omalle tai tutkimusryhmän tietokoneelle, ulkoiselle kiintolevyllä tai muistitikulle. Päin vastoin, Helsingin yliopiston tutkijoille suunnatun kyselyn perusteella (Kuusniemi & Salmi 2017) vajaa kolmannes on jättänyt tutkimusaineistonsa julkaisematta missään tutkimusaineistolle suunnatussa arkistossa. Vastaavia lukuja on saatu myös kansainvälisissä tutkimuksissa, kun on tarkasteltu tutkijoiden aktiivisuutta tutkimusaineistojen avaamisessa ja julkaisemisessa (Tenopir ym. 2011). Tämä siitäkin huolimatta, että avoimesti julkaistujen tutkimusaineistojen on todettu monessa yhteydessä tuottavan huomattavasti enemmän viittauksia kuin pelkän aiheeseen liittyvän perinteisen tutkimusjulkaisun julkaisemisen (Peters ym. 2016; Piwowar & Vision 2013)

Nimenomaan tutkimusaineistojen julkaisemisen puutteellisista käytännöistä johtuen tutkimusaineistojen viittaustietokannat, jotka tutkimusmetriikan pohjana usein toimivat, ovat vielä hyvin vaillinaisia. Esimerkiksi Thomson Reutersin *Data Citation Index* viittaustietokantaan päätyy yhä hyvin vähän tunnisteellisten tutkimusaineistojen viittaustietoja. Erään tutkimuksen perusteella jopa 85 prosenttia kaikista viittaustietokannan tutkimusaineistoista on yhä vailla kunnollista viittausta. Joskin muutamat tutkimusaineistotyytit, erityisesti kyselyaineistot, aggregoidut tutkimusaineistot sekä sekvenssiaineistot, erottuvat muita tutkimusaineistoja viitatuimmaksi (Peters ym. 2016).

Tutkimusaineistojen tutkimusmetriikan voidaankin katsoa olevan pitkälti samassa tilanteessa kuin oltiin tutkimusjulkaisujen metriikan osalta 2000-luvun puolivälissä. Kun tällöin tutkimusjulkaisujen varten vaaditut tekniset ratkaisut, aihepiiriä pohtinut tutkimuskirjallisuus, tutkimusjärjestelmän asettamat poliittiset linjaukset ja ohjaus sekä käytännön työn tekevät tutkijat olivat pääpiirteissään etenemässä samaan suuntaan, havahduttiin tutkimusmetriikan osalta siihen, että tutkimusjulkaisuihin liittyen paikkaansa pitävää aineistoa oli hyvin vähän tarjolla kattavien analyysien tekoon (Puuska 2014). Tutkimusaineistojen kohdalla työ kaikkien edellä mainittujen toimijoiden suhteen on vielä vahvasti käynnissä, joten kattavia tutkimusmetrisiä analyysieja tutkimusaineistoista saataneen odotella vielä vuosia.



### 3. Metasynteesi

Pro gradu -työn metodina toimii metasynteesi, joka on eräs kirjallisuuskatsauksen muodoista. Metasynteesi pyrkii paitsi yhdistämään aihealueen tutkimuksen, myös ymmärtämään ja selittämään tutkittavia ilmiöitä näiden yhtäläisyyksien ja erojen kautta. Salminen (2011, 12-13) on koostanut metasynteessin keskeiset elementit sitä käyttävän tutkimuskirjallisuuden pohjalta (esim. Walsh & Downe 2005) ja oheinen listaus noudattelee pääpiirteissään Salmisen esittelemää metodin vaiheistusta:

- Aihealueen rajaaminen - tutkimuskysymykset ja tavoitteet
- Aineiston haku - relevantti tutkimuskirjallisuus
- Aineiston rajaaminen - mitä ottaa mukaan
- Arviointi - aineiston läpikäynti
- Vertailu - yhtäläisyyksien ja erojen etsiminen, niin sanottu vastavuoroinen käänös
- Synteesi

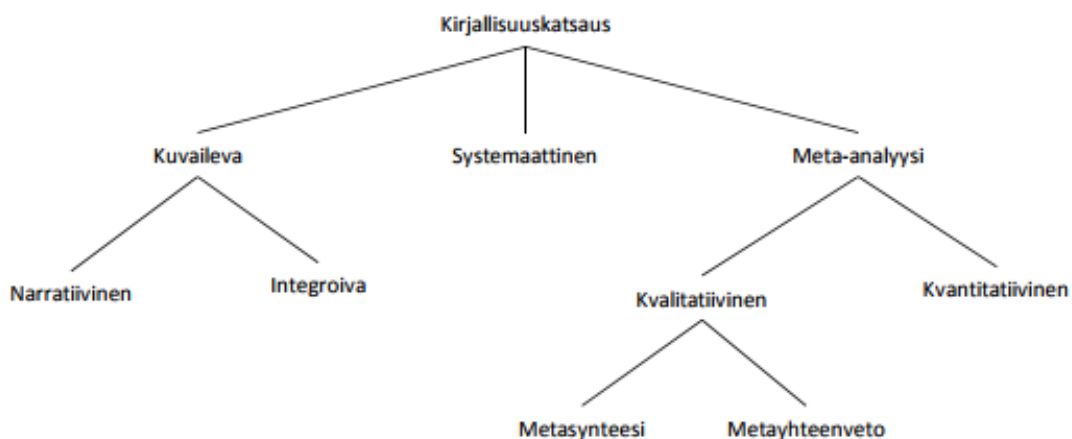
Metodin kannalta oleellista on eri metaforien, fraasien, ideoiden ja käsitteiden vertailu keskenään ja näiden niputtaminen toisiinsa, jolloin voidaan muodostaa käsiteluokitteluja sekä löytää tutkimusten tuloksista yhtenäisyyttä ja toisaalta myös eroavaisuuksia. Kun luokittelu on suoritettu, voidaan suorittaa vastavuoroinen käänös (*reciprocal translation*), jossa yhtenäiset piirteet yhdistetään mahdollisuuksien mukaan eri käsitteiden alle (Salminen 2011, 13). Vastavuoroisessa käänöksessä on kyse samasta vaiheesta joka on läsnä monessa muussakin laadullisessa tutkimusmetodissa, tutkija tekee aineiston ja oman ajattelutyön perusteella aiheeseen liittyvää eri asioiden yhdistelyä ja muodostaa täten lopuksi tutkimustuloksena synteessin.

Metasynteesi lasketaan kvalitatiivisten metodien joukkoon, joskin laajasti nähtynä siinä on myös elementtejä kvantitatiivisista metodeista, esimerkiksi aineiston keruun ja tulkinnan kautta (Salminen 2011, 13; Walsh & Downe 2005).

Kirjallisuuskatsaus tutkimusmetodina saatetaan toisinaan tulkita suhteellisen kapeana lähtökohtana tutkimukselle, mutta käytännössä valitulla kirjallisuuskatsauksen tyypillä on suuri vaikutus metodin haastavuuden tasoon tutkimusta tehtäessä ja toisaalta metodin kautta saavutettujen tutkimustulosten arvoon. Kirjallisuuskatsaus voidaan toteuttaa ilman tarkkoja rajoituksia, jolloin se toimii lähinnä taustoittavassa tai kuvailevassa roolissa

esimerkiksi tutkimuksen alkupuolella jonkin toisen metodin ollessa kuitenkin niin sanottu pääasiallinen tutkimusmetodi (Metsämuuronen 2009, 212-252). Toista kirjallisuuskatsauksen ääripäätä edustaa esimerkiksi kvantitatiivinen meta-analyysi, joka menetelmänä muistuttaa paljon tilastotieteessä käytettyjä metodeja. Sen pohjana on laaja kirjallisuudesta muodostuva aineisto, johon koodaamiseen jälkeen ajetaan laskennallisia analyysejä, jonka jälkeen voidaan muodostaa kvantitatiiviseen aineistoon perustuva tutkimustulos. (Salminen 2011, 6-15)

Salminen (2011, 6-15) on jakanut kirjallisuuskatsauksen metodina kolmeen perustyyppiin, jotka ovat kuvaileva kirjallisuuskatsaus, systemaattinen kirjallisuuskatsaus sekä meta-analyysi, jälkimmäisen ollessa jaettuna vielä kvalitatiiviseen tai kvantitatiiviseen tyyppiin. Salminen (2011) asettaa metasynteesin osaksi kvalitatiivista meta-analyysia. Kirjallisuuskatsauksen eri tyypit on nähtävillä Kuviossa 1.



Kuvio 1: Kirjallisuuskatsauksen tyypit (Salminen 2011, Kuvio 1).

Vaikka metasynteesi on suhteellisen tuore metodi (Walsh & Downe 2005), on sitä käytetty vaihtelevasti eri tieteenalojen tutkimuksissa. Suosituinta metodin käyttö on lääke- ja terveystieteiden tutkimuksessa (esim. Beck 2002; Clemmens 2003; Paterson yms. 1998), mutta hallintotieteissä metodia on sovellettu muun muassa johtamisen tutkimuksessa (Pielstick 1998) sekä eri hallintomallien tarkastelussa (Lee 2010).

Metasynteesin valinta pro gradu -työn metodiksi on perusteltua useasta eri syystä. Ensinnäkin tutkimusaineistot sekä niihin kohdistuva tutkimusmetriikka ovat uudehko ilmiö, jonka käsittelyyn ei tutkimuskirjallisuus tällä hetkellä tarjoa vakiintuneita käytäntöjä

tai käsitteitä. Tämän takia aihepiirin ympärillä on syytä tehdä käsitteitä luokittelevaa ja näkökulmia yhdistävää tutkimusta. Toiseksi aiheen systemaattinen ja laaja-alainen lähestyminen tarjoaa mahdollisuuden havainnoida mahdollisia ristiriitaisuuksia aihepiiriä käsittelevässä tutkimuksessa ja toisaalta tunnistaa yleisimpiä keskustelun lähtökohtia, jotka jo jossain määrin yhdistävät tutkimusta ilmiön ympärillä. Pro gradu -työn lähtökohtana metasynteesi mahdollistaa myös tutkimusaineistoihin liittyvään tutkimusmetriikkaan perehtymisen laajemmin, kuin mitä jonkin kapeamman näkökulman tai tutkimuskysymyksen ympärille rakentunut lähestymistapa olisi mahdollistanut. Tämä johtuu siitä, että aineistoa rajatessa ja valikoidessa tulee väistämättä käytyä läpi myös aihetta laajasti eri näkökulmista tarkasteltavia teoksia, vaikka nämä eivät välttämättä työn aineistoon lopulta päädykään.

Lähtökohdiltaan metasynteesi sopii myös paremmin ilmiön tarkasteluun korkeakouluhallinnon- ja johtamisen näkökulmasta, sillä näin työ ei kohdistu liikaa nimenomaan tutkimusmetriikkaan ja sen sovellutuksiin, vaan pääpaino tutkimuksessa pysyy tutkimusaineistoissa ja niiden ominaisuuksissa. Muodostuvan synteessin kautta ilmiön tarkastelu onnistuu siis ilman laajempaa perehtymistä aihepiirin teknisempiin yksityiskohtiin esimerkiksi tutkimusinfrastruktuurien tai tutkimusmetriikan käytäntöjen ja menetelmien näkökulmasta.

### 3.1. Aineiston haku ja rajaus

Aineiston rajaus ja sen avaaminen on oleellinen osa metasynteesia, sillä se ohjaa merkittävästi tätä nimenomaista tutkimusmetodia ja lopullista analyysia. Tutkimuskysymykseen perustuva tarkka rajaus tutkimuksen alkuvaiheessa mahdollistaa uskottavan analyysin ja synteessin muodostamisen, sillä näin pystytään pitkälti välttämään tilanne, jossa aineisto muodostaa vinoumia tutkijan antaessa aineiston johdatella lähteestä toiseen (Salminen 2011,13). Rajauksella vältytään myös kasvattamasta aineistoa tutkimukseen nähden liian laajaksi, mikä on pro gradu -työn kannalta välttämätöntä.

Aineiston haun kannalta merkittävässä asemassa olivat sekä Tampereen yliopiston Andor-hakupalvelu, jossa haut tehtiin myös kirjaston kokoelmien ulkopuolelle, sekä Google että Google Scholar -hakupalvelut. Hakusanoina toimivat *research data* sekä *data metrics* sekä näiden tukena *evaluation* ja *indicators*. Näiden tuottamia hakuja tarkennettiin tarvittaessa, sillä erityisesti Googlen hakupalvelut tuottivat paljon hakutuloksia, joissa viitattiin erilaisiin kaupallisiin datametriikkaa tarjoaviin

palveluntuottajiin. Kun hakuosumat eivät enää tuottaneet relevantteja hakutuloksia, koossa oli hieman vajaa 40 julkaisuksi luokiteltavaa dokumenttia tai kirjoitusta ilmiöön liittyen. Koska määrä oli huomattavasti suurempi, kuin mitä pro gradu -työn puitteissa on mahdollista käydä läpi, jouduin tekemään tähän määrään rajauksia.

Ajanjakson osalta rajausta työssä on asetettu vuosien 2010-2017 välillä julkaistuihin asiantuntijaryhmien raportteihin. Tätä valintaa ohjaa muutamakin seikka joista merkittävimpänä on se, että tutkimusaineistoihin liittyvä tutkimusmetriikka on siinä määrin uusi ilmiö, ettei sitä käsitteleviä raportteja tai kirjallisuutta yleensä ole juuri julkaistu ennen vuotta 2010. Näissä harvoissa tapauksissa ilmiötä on myös käsitelty pitkälti spekulatiivisesti tulevasta tai niin pienenä osana muuta tutkimusmetriikkaa, usein bibliometriikkaa, että ne eivät olisi tarjonneet analyysia varten tarpeeksi kattavasti materiaalia. Toinen valittua ajanjaksoa perusteleva seikka juontaa juurensa itse tutkimusaineistoihin. Vielä 2000-luvulla tutkimusaineistojen arkistointiin ja tallentamiseen oli käytössä tutkijoilla verrattain vähän vaihtoehtoja. 2010-luvulla erilaisten palveluiden ja tutkimusinfrastruktuurien myötä tilanne on huomattavasti parantunut ja samalla myös tutkimusmetriikan näkökulmasta ilmiön tarkastelu on tullut ajankohtaisemmaksi.

Aineiston kieleksi valikoitui englanti, sillä suomenkielisiä tutkimusaineistoihin tai niiden metriikkaan liittyviä julkaisuja ei juuri ole saatavilla ja nämä on mahdollisuuksien mukaan huomioitu pro gradu -työssä. Kielen suhteen haasteelliseksi osoittautuivat kuitenkin esimerkiksi käsitteet, jotka ovat kansainvälisessä kirjallisuudessa toisinaan limittäisiä ja toisaalta Suomen kielessä ei välttämättä löydy suoraa vastinetta englanninkielisille käsitteille. Jälkimmäisestä hyvänä esimerkkinä mainittakoon työn teemaan tiukasti liittyvä käsite tutkimusaineisto, jonka lähin englanninkielinen vastine *research data*, joka kattaa kaiken yksittäisistä mittaustuloksista aina isoihin tilastoaineistoihin saakka. Suomen kielessä tutkimusaineisto tarkoittaa kuitenkin jo jalostetumpaa tutkimusmateriaalia, jolloin mittalaitteen muodostama data ei ole suoranaisesti tutkimusaineisto, vaikka siitä sellainen voikin muodostua datan jalostamisen myötä (Yhteiskuntatieteellinen tietoaarkisto 2018).

Rajauksena aineistolle käytän julkaisun tyyppiä. Ilmiön tarkasteluun sopivimmaksi materiaaliksi valikoitui erilaiset raportit ja selvitykset, joiden taustalla on usein vaihtelevasti joko organisaatioita tai kansallisia sekä kansainvälisiä yhteistyöverkostoja. Syy raporttien valintaan juontuu siitä, että tutkimusaineistoihin liittyvää tutkimusmetriikkaa käsitellään ilmiön tuoreudesta johtuen juuri erilaisissa työryhmissä ja asiantuntijaverkostoissa, joiden työtä kokoavia yhteenvetoja useat aineiston

muodostavat raportit ovat. Tämän ensimmäisen rajauksen jälkeen aineiston harkinnassa korostui se, käsitelläänkö raportissa tai selvityksessä tutkimusmetriikkaa. Aineistosta karsiutui tällöin sellaiset julkaisut, joiden näkökulma oli puhtaasti tutkimusaineistojen luonteeseen ja niiden ominaisuuksiin liittyviä, usein teknisiä, selvityksiä. Samassa yhteydessä poistuivat aineistoista myös ne selvitykset, joiden näkökulmana oli tarkastella tutkimusaineistoja tutkijatasolta esimerkiksi tutkimusaineistojen julkaisemisen, käytön tai hyödyntämisen näkökulmasta. Kolmas merkittävä rajausta tehtiin sen suhteen, käsitelläänkö raportissa nimenomaan tutkimusaineistoja koskevaa tutkimusmetriikkaa. Tässä rajauksessa poistuivat ne selvitykset, joiden kohteena oli ainoastaan bibliometriikka tai altmetriikka jonkin toisen tutkimustuotoksen, usein tutkimusjulkaisun, näkökulmasta.

Tutkimuskirjallisuus nimenomaan tutkimusaineistojen tutkimusmetriikkaan liittyen on yhä hajanaista ja niiden valinta pro gradu -työn aineistoksi olisi johtanut liian suppeaan aineistoon.

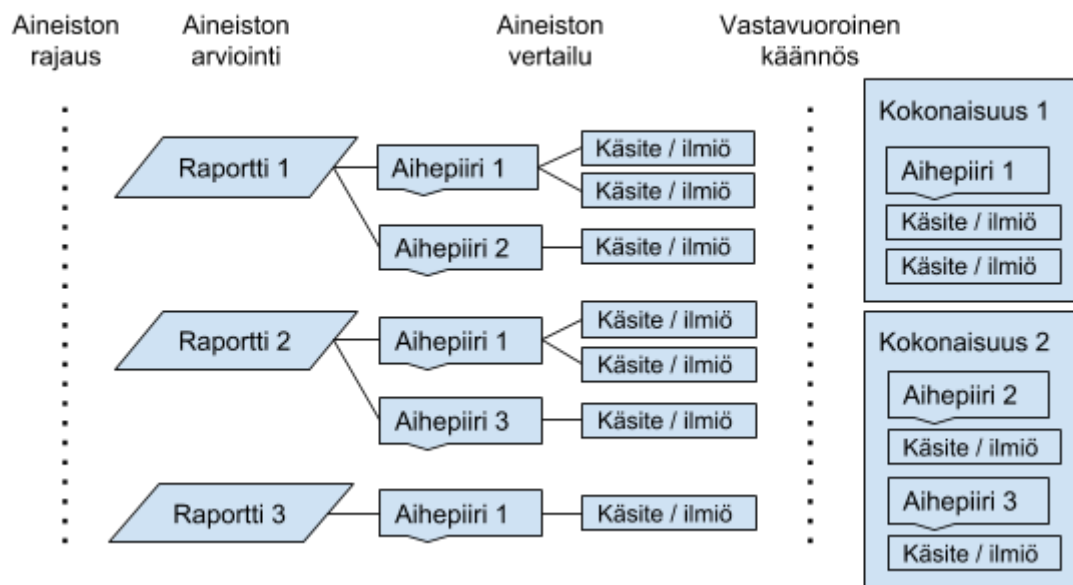
Edellä kuvattujen rajausten perusteella sain kavennettua hakuvaiheen vajaan 40 julkaisun yhteensä 13 eri selvitykseen ja raporttiin, jotka muodostavat työni aineiston. Rajausten myötä valikoitunut aineisto tarjoaa ajankohtaisen ja tarkasti tutkimusaineistoihin kohdistuvan lähtökohdan asetettuun tutkimuskysymykseen vastaamiseen. Aineiston muodostavat selvitykset ja raportit on kuvattu tarkemmin luvussa 3.3.

## 3.2. Aineiston arviointi ja mahdolliset rajoitteet

Aineiston rajausta seurasi sen arviointi ja läpikäynti tutkimuskysymyksen kannalta. Tätä vaihetta tukeakseni muodostin jokaisesta raportista ja selvityksestä lyhyitä yhteenvetoja niiden käsittelemien aiheiden mukaisesti. Varsin vaihtelevista raporteista johtuen tämä selkeytti huomattavasti kokonaisuuden hahmottamista ja jokaisen raportin tulokulmaa aihepiiriin. Näin tehden pääsin myös paremmin ilmiön kannalta keskeisten huomioiden äärelle uuteen raporttiin tai selvitykseen siirryttäessä, sillä tukena olivat yhteenvedot muista raporteista.

Aineiston arvioinnin perusteella saatoin suorittaa aineiston vertailun, jossa pyrin aineistoista kumpuavien käsitteiden ja yhtäläisyyksien kautta muodostamaan laajempia kokonaisuuksia tutkimusaineistojen tutkimusmetriikkaan liittyen. Näiden kokonaisuuksien alle ryhmittelin kustakin raportista ja selvityksestä löytyviä aihepiirejä ja

pyrin näiden aihepiirien sisällä vielä yhtenäistämään niitä mahdollisia käsitteitä, jotka erosivat hieman raportista riippuen, mutta jotka käsittelivät kuitenkin lähestulkoon samaa asiaa. Suoritin arvioinnin loppuun tekemällä niin sanotun vastavuoroisen käännöksen, jossa yhtenäistin vielä kunkin kokonaisuuden alle sijoitetut aihepiirit soveltuvien osin ja muodostin analyysin määrittelevän ja sitä ohjaavan luokittelun. Luokittelun myötä myös kokonaisuudet saivat niistä koostuvien aihepiirien osalta soveltuvat nimet. Nämä tutkimuksen kannalta oleelliset vaiheet on esitetty Kuviossa 2.



Kuvio 2: Metasynteessin vaiheet tutkimuksessa (soveltaen tähän tutkielmaan Salminen 2011, 12-13).

Tutkimusmetodina metasynteesiin ja yleisesti myös kirjallisuuskatsaukseen liittyy kuitenkin joitain riskejä. Näistä riskeistä merkittävimäksi Salminen (2011, 12-15) on esittänyt tutkimuksen ajautumisen liaksi kuvailevaan suuntaan, jolloin tutkimuksen myötä tehdyt tutkimustulokset saattavat jäädä laihoiksi tai johtopäätökset perustua pääasiassa vain tutkimuksen kohteena olleeseen aineistoon ja siitä tehtyihin tulkintoihin. Tämän välttääkseni olen työssäni kontekstoinut tulkintaa vahvasti myös aineiston ulkopuoliseen tutkimuskirjallisuuteen, jotta työn tulokset eivät tukeudu ainoastaan aineistosta muodostettaviin tulkintoihin. Toisaalta tutkimustuloksena syntyvän synteessin vakuuttavalla esittämistavalla ja aineistoista juontuvilla perusteluilla voidaan varmistaa, että tutkimustulokset eivät jää vain aineiston kuvailun tasolle.

Tämän työn kannalta metasynteessin käyttämiseen tutkimusmetodina liittyy myös valitusta aineistosta kumpuava haaste. Ilmiön tarkasteluun valitut raportit ja työryhmien työstä muodostetut yhteenvedot ovat keskenään varsin heterogeenisiä kirjallisia

tuotoksia. Tämä käy ilmi paitsi raporttien vaihtelevasta pituudesta ja rakenteesta, myös erityisesti siitä, kuinka aihetta on niissä lähestytty ja käsitelty. Osassa aineiston muodostavista raporteista on pureuduttu kokonaiskuvan kannalta hyvin pieniin yksityiskohtiin, kun osa taas pyrkii tarkastelemaan ilmiötä laajemmasta näkökulmasta ja ilmiötä abstraktoimalla. Tämä hankaloittaa aineiston pohjalta tehtävää synteesiä, sillä erilaisista tarkastelutasoista johtuen käsitteiden ja termien ryhmittely hankaloituu. Työssäni olen huomionut raporttien eri tarkastelutasot luokittelemalla kirjallisuuskatsaukseen valikoitunutta aineistoa eri kategorioihin esimerkiksi tarkasteltavan tutkimusmetriikan osa-alueen tai näkökulman perusteella. Näin toimimalla synteessin muodostaminen helpottuu, sillä tarkastelutasot eivät sekoita aineiston tulkintaa siinä määrin, kuin ilman minkäänlaista luokittelua olisi vaarana käydä.

Koska aineiston muodostavat eri työryhmien raportit ja selvitykset, ovat ne usein työryhmien työn lopputulemia. Raporttien kautta ei siis ole mahdollista huomioida kaikkea työryhmien "näkökenttää" työtä, jota on tehty raporttia muodostettaessa. Tästä syystä aineistossa on tutkimuskirjallisuudesta poiketen vahvasti toteavampi ote eri ilmiöihin ja kontekstointi sekä keskustelu muuhun tutkimuskirjallisuuteen on täten vähäistä. Tästä syystä oman työni uskottavuuden kannalta on tärkeää, että analyysin tuottamassa yhteenvedossa palataan vielä aihepiirin tutkimuskirjallisuuteen ja siellä käytyyn keskusteluun.

Toinen merkittävä seikka tulosten reliabiliteetin näkökulmasta on se, että aineiston muodostavat raportit saattavat sisältää eri organisaatioiden omia tarkoitusperiä ajavia avauksia. Tämän työn aineiston muodostavissa selvityksissä on taustalla esimerkiksi huomattava määrä tutkimusinfrastruktuurista vastaavien organisaatioiden edustajia, jolloin painotus nimenomaan teknisestä näkökulmasta saattaa välittyä myös selvityksissä ja niiden johtopäätöksissä. Toisaalta metodina metasynteessissä tämä on jo huomioitu lähtökohtaisesti hyvin, sillä nimenomaan luokittelun ja yhdistelyjen sekä erojen hakemisen kautta aineistosta erottuvat hyvin ne julkaisut, joissa keskitytään tarkastelemaan vain tiettyjä ilmiön osa-alueita (Salminen 2011, 12-15). Tämän lisäksi olen työssäni kuvannut raportteja ja niiden taustalla toimivia organisaatioita tarkemmin luvussa 3.3. joka osaltaan lisää aineiston pohjalta laaditun analyysin tulosten reliabiliteettia ja tuo ilmi mahdollisia raporttien tarkoitusperiä.

Eräs tämän työn kannalta tiedostettu haaste liittyy myös aineistoksi valikoituneen kirjallisuuden kieleen. Ilmiön kannalta suurin osa relevantista materiaalista on saatavilla ainoastaan englanninkielisenä, jonka seurauksena tässä työssä termien ja käsitteiden

vakiintuneiden suomenkielisten vastineiden valintaan on käytetty erityistä tarkkaavaisuutta. Käsitteiden tulkinnassa ja valinnassa on siksi tukeuduttu TUHA-tietomallityöryhmän tutkimushallinnon alueen käsitteiden määrittelyyn ja valmiisiin määritelmiin (TUHA-tietomallityöryhmä, 2017). Työssä on hyödynnetty myös ilmiöön liittyvää suomalaista tutkimuskirjallisuutta sekä eri organisaatioiden tutkimusaineistojen hallinnan eteen tehtyä työtä mahdollisuuksien mukaan käsitteiden ymmärtämisessä ja valinnassa. Tästä huolimatta yhtenäisen käsitteistön muodostaminen on tutkimusaiheena olevan ilmiön nopean muutoksen takia vähintäänkin haastavaa, eikä käsitteiden pysyvyydestä voi tästä syystä antaa takeita. Lohduttavasti ongelma vaivaa kuitenkin kaikkea aihepiirin tutkimuskirjallisuutta ja se ilmenee välittömästi, mikäli selailee mitä tahansa tutkimusaineistoja käsittelevää tutkimuskirjallisuutta vaikkapa 1990-luvulta, jolloin käytännöt, esiteltyt työkalut ja infrastruktuurit ovat nykyajan näkökulmasta väistämättä jo vanhentuneita.

### 3.3. Aineiston muodostavat raportit

Aineiston rajauksen perusteella tutkimukseen valitut 13 raporttia ja selvitystä on seuraavassa esitelty niin, että niistä selviää muun muassa taustalla vaikuttavat organisaatiot, raportista tai selvityksestä vastuussa oleva mahdollinen työryhmä ja työlle asetetut tavoitteet. Esittelyistä selviää myös se, millaista aineistoa raportin laadinnassa on hyödynnetty ja millaisia tahoja raportin valmistelussa on ollut mukana. Jotta viittaaminen raportteihin olisi selkeämpää, kunkin raportin viite analyysissä on lyhennetty esimerkiksi seuraavasti: RDA 2016. Käytetyt lyhenteet on ilmoitettu ensimmäisenä kunkin raportin esittelyssä ja täydelliset viitetiedot löytyvät lähdeluettelosta. Liite 1 kuvaa tarkemmin raportteja ja selvityksiä muun muassa niiden pohjana olevan aineiston, tieteenalojen sekä muiden lähtökohtien näkökulmasta.

#### **CODATA 2013 - Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data - Socha, Y. ed. (2013) - 76 s.**

Raportti on laadittu osana CODATA-ICSTI Task Group on Data Citation Standards and Practices työryhmän työtä, joka asetettiin vuonna 2010 selvittämään tutkimusaineistojen viittauskäytäntöjä. Sekä CODATA (Committee on Data for Science and Technology) että ICSTI (International Council for Scientific and Technical information) ovat osa kansainvälistä sateenvarjo-organisaatiota ICSU:a (International Council for Science). Tavoitteeksi työryhmän työlle on asetettu tutkimusaineistoihin viittaamisen ja arvottamisen käytäntöjen selvittäminen ja mahdollisten trendien, onnistumisten ja



haasteiden tunnistaminen. Aihepiiriä lähestytään paitsi tutkimuksen infrastruktuurien tutkimusaineistoihin liittyvän valmiuden tarkastelulla, mutta myös tutkimusaineistojen julkaisemisen ja siitä saatavien hyötyjen kautta.

Raportti koostaa työryhmän työtä yhteisen tiedon kasaamisessa vuodesta 2011 lähtien ja raportin aineisto muodostuu paitsi katsauksesta aihepiiriin tutkimuskirjallisuuteen, myös 62 asiantuntijahaastattelusta. Haastateltavina on ollut tutkimusrahoittajien, tiedekustantajien, julkaisuarkistojen ja kirjastojen edustajien lisäksi myös kourallinen tutkijoita.

**DCC 2015 - How to Track the Impact of Research Data with Metrics - Ball, A., & Duke, M. (2015) - 16 s.**

Raportti muodostettu osana DCC:n (Digital Curation Centre) toimintaa, joka on Edinburghin, Glasgown ja Bathin yliopistojen yhteinen konsortio edistämään tutkimuksen digitaalisten aineistojen jatkokäyttöä. Raportissa käydään läpi tutkimusaineistoihin liittyvää tutkimusmetriikkaa eri tasoilla: millaisia käsitteitä ilmiöön liittyy, mitä palveluita tutkimusaineistojen tarkasteluun tutkimusmetriikan näkökulmasta on nyt saatavilla, millaisia ongelmia aiheeseen liittyy ja kuinka ala ehkä kehittyä jatkossa.

Raportin pääasiallisena aineistona toimii tutkimuskirjallisuus, mutta siinä on hyödynnetty myös konsortion jäsenyliopistojen käytännön toiminnasta kumpuavia ilmiötä liittyen tutkimusaineistoihin sekä niiden metriikkaan.

**EC 2017 - Next-generation metrics: Responsible metrics and evaluation for open science - Wilsdon, J. et al. (2017) - 26 s.**

Euroopan komission asettaman Expert Group on Altmetrics -työryhmän työtä kokoavassa raportissa on keskitytty selvittämään Euroopan komission ajaman avoimen tieteen agendan näkökulmasta eri tutkimusmetriikan alojen soveltumista tutkimuksen arviointiin. Raportissa on keskitytty pääasiassa altmetriikan eri muotoihin ja nykytilanteeseen, mutta samalla sivuttu vahvasti myös tutkimusaineistojen asemaa yhtenä oleellisena tutkimustuotoksena ja siihen liittyvää tutkimusmetriikkaa. Raportin tulokset tiivistyvät viiteen pääteemaan sekä 12 suositukseen, jotka on kohdistettu usealle eri tutkimuskentän toimijalle aina tutkimusinfrastruktuurin tuottajista yhteiskuntaan saakka.

Aineiston muodostaa paitsi tutkimuskirjallisuuteen tehty kirjallisuuskatsaus, myös usean jäsenmaan keskeisille toimijoille lähetetty selvityspyyntö. Tässä joukossa on sekä tiedekustantajia, yliopistoja, konsortioita sekä muita voittoa tavoittelemattomia järjestöjä Euroopasta. Vastauksia selvityspyyntöön tuli yhteensä 19 eri organisaatiolta.

**FORCE11 2016 - The FAIR Guiding Principles for scientific data management and stewardship - Wilkinson, M. D. et al. (2016) - 9 s.**

Yhteistyöorganisaatio FORCE11:a (The Future of Research Communications and e-Scholarship) alla toimivan FAIR työryhmän julkaisemat tutkimustuotosten hallintaan ja tallentamiseen liittyvät FAIR-periaatteet perustuvat usean vuoden asiantuntija- ja työryhmätyöhön eri tutkimusjärjestelmän toimijoiden keskuudessa. Mukana työssä on ollut tutkijoiden lisäksi edustajia kustantajien, rahoittajien, kirjastojen ja arkistojen puolelta. Julkaistuissa periaatteissa käsitellään tutkimustuotoksia pääasiassa infrastruktuurin näkökulmasta, mutta esillä on myös tutkimusmetriikan kannalta mielenkiintoisia näkökulmia tutkimusaineistoihin liittyen.

Julkaisun perustana ovat vuonna 2014 Leidenissä järjestetystä työpajasta “Jointly Designing a Data Fairport” jalostuneet periaatteet, joita on työryhmän toimesta kehitetty eteenpäin.

**HEFCE 2015a - The Metric Tide - Wilsdon, J. et al. (2015) - 176 s.**

Raportin taustalla on HEFCE (Higher Education Funding Council for England), joka jakaa Iso-Britanniassa kansallisesti tutkimusrahoitusta yliopistoille ja muille korkeakouluille. Raportin työstämiseen osallistunut työryhmä on koottu tutkimuskentän eri toimijoista niin tutkimusmetriikan, tutkimusrahoituksen, tiedekustantajien kuin korkeakouluhallinnonkin edustajista. Päämääränä raportilla on 1) koota tietoa tutkimusmetriikan eri muotojen soveltuvuudesta eri tutkimusalojen ja käytäntöjen piirissä, 2) havainnoida tutkimustiedon hallinnan kehitystä organisaatioissa, 3) selvittää kasvavaa määrällisten indikaattoreiden käyttöä ja niiden vaikutusta tutkimustoiminnan eri osissa hallinnon näkökulmasta aina indikaattoreiden “pelaamiseen” (*gaming*) saakka. Raportti tarkastelee myös kuinka tutkimusrahoittajat Iso-Britanniassa jo hyödyntävät määrällisiä indikaattoreita ja hahmottelee tulevia kehityssuuntia niihin liittyen. Tutkimusaineistot raportti huomioi osana laajenevaa tutkimustuotosten kirjoa ja tarjoaa myös niihin liittyen hahmottelua tutkimusmetriikan näkökulmasta.

Aineistona toimii paitsi laajan toimijakentän muodostama työryhmä, myös tutkimuskirjallisuuteen liittyvä laaja kirjallisuuskatsaus (HEFCE, 2015b) sekä työtä tukevat, eri tutkimuksen toimijoille järjestetyt, työryhmät ja niistä saadut huomiot ja kommentit.

**HEFCE 2015b - The Metric Tide Literature Review Supplementary Report I - Wouters, P. et al. (2015) - 200 s.**

Edellä esitellyn HEFCE:n asettaman työryhmän raportin kirjallisuuskatsauksen muodostava itsenäinen osuus, joka paneutuu tarkemmin tutkimusmetriikkaan ja sen sovellutuksiin eri tutkimustuotoksiin, myös tutkimusaineistoihin, liittyen. Kolme raportin teemaa ovat 1) bibliometriikka ja indikaattoreiden käyttö, 2) vertaisarviointi ja bibliometriikka sekä 3) vaihtoehtoiset tutkimusmetriikan muodot. Näiden pohjalta raportti antaa joitakin suosituksia tutkimuksen eri toimijoille liittyen indikaattoreiden ja tutkimusmetriikan käyttöön yleisesti tutkimuksen arvioinnissa.

Aineistona toimii vaikuttava määrä tutkimusmetriikkaan liittyvää tutkimuskirjallisuutta, johon tukeutuen kirjallisuuskatsaus on tehty.

**JISC 2016 - Jisc Briefing Document on Data Citations - Neylon, C. (2016) - 9 s.**

Iso-Britannian keskeiselle tutkimusinfrastruktuurien toimijalle Jisc:lle toimitettu lyhyt dokumentti, jossa käydään läpi tutkimusaineistoihin liittyviä viittauskäytäntöjä, näihin liittyvää harmonisointia infrastruktuurien kannalta ja mahdollisia tulevaisuuden hyviä käytäntöjä tutkimusaineistoihin liittyen. Tarjoaa muutamia suosituksia liittyen tutkimusaineistojen määrälliseen analysointiin sekä niiden jatkokäytön arviointiin. Varsinaisesti tutkimusmetriikkaa ei kuitenkaan käsitellä, vaan keskiössä ovat tutkimusaineistojen viittauskäytännöt ja näiden hyödyntäminen määrällisessä mielessä.

Aineiston muodostaa aihepiirin tutkimuskirjallisuus, johon on kerätty tutkimusaineistojen infrastruktuurien kannalta olennaisia julkaisuja.

**KE 2013a - Making data count - Knowledge Exchange (2013) - 8 s.**

Kaksipäiväisen työpajan pohjalta muodostettu raportti KE:n (Knowledge Exchange), eurooppalaisen tutkimusinfrastruktuurien yhteistyöorganisaation, toimesta jossa tarkastellaan tutkimusaineistojen viemistä osaksi tutkimustoiminnan arviointeja ja

tutkimusmetriikkaa. Keskustelujen pohjalta muodostetussa yhteenvedossa käydään läpi sekä tutkimusaineistojen arviointia että tutkimusaineistojen viittauskäytäntöjä ja esitetään toiveita tutkimuksen eri toimijoille näihin liittyen.

**KE 2013b - The Value of Research Data - Costas, R., Meijer, I., Zahedi, Z. & Wouters, P. (2013) - 48 s.**

Raportti on muodostettu KE:n jäsenorganisaation jäsenistä muodostetun työryhmän työn tuloksena ja sen tavoitteena on perehtyä tutkimusaineistoihin liittyvään tutkimusmetriikkaan, sen nykytilanteeseen ja kehitykseen ja toisaalta käytäntöjen puutteeseen ja ongelmakohtiin. Raportti tarjoaa tutkimusaineistojen jakamiseen ja tutkimusmetriikkaan liittyen suosituksia ja kehityskohteita tutkimuksen eri toimijoille. Tutkimusmetriikan lisäksi raportti paneutuu myös hieman tutkijan näkökulmaan tutkimusaineistojen jakamiseen liittyen.

Aineiston muodostaa paitsi ilmiöön liittyvä tutkimuskirjallisuus, myös yhdeksän tutkimusorganisaation asiantuntijan haastattelut. Raportti sisältää myös empiirisen analyysin erilaisista tutkimusaineistoille saataville olevista arkistoista erityisesti tutkimusmetriikan näkökulmasta.

**NISO 2016 - Outputs of the NISO Alternative Assessment Metrics Project - NISO (2016) - 86 s.**

Raportti koostaa yhteen vuonna 2013 aloitetun vaihtoehtoisen tutkimuksen arvioinnin projektin työtä osana NISO:n (National Information Standards Organization) muuta toimintaa. NISO on kansainvälinen kustantajien, kirjastojen ja infrastruktuurin yhteenliittymä, jossa pyritään muodostamaan yhteisiä standardeja digitaaliseen tietoon liittyen. Raportin pohjana olevan projektin työ jakautui kolmen näkökulman tarkasteluun: 1) altmetriikan hyödyntäminen eri toimijoiden näkökulmasta, 2) dokumentaation koostaminen, joka auttaa organisaatioita parantamaan tutkimusaineistoihin ja tutkimusmetriikkaan liittyvää käyttöä ja 3) altmetriikan ja muiden tutkimusmetriikan analyysien pohjalla käytettyjen aineistojen laatuun perehtyminen ja mitä niille tulisi tehdä paremman hyödyntämisen mahdollistamiseksi.

Raportissa on hyödynnetty projektin järjestämien työpajojen aiheisiin liittyviä keskusteluita ja kommentteja. Raporttia varten on myös tehty kirjallisuuskatsaus aiheiden kannalta relevantteihin tutkimusjulkaisuihin.

**MDC 2015a - Make Data Count - Unit 1 Final Report - Strasser, C., Kratz, J., Lin, J. (2015) - 11 s.**

Raportti on osa MDC -projektia (Make Data Count), jossa yhteistyössä PLOS (Public Library of Science), UC3 (University of California Curation Center) sekä DataONE (Data Observation Network for Earth) pyrkivät muodostamaan uusia tutkimusaineistojen käyttöön liittyviä indikaattoreita eri toimijoiden käyttöön. Projekti lähestyy aihetta tutkijoiden näkökulmasta ja pyrkii kartoittamaan tutkijoiden itsensä mielipiteitä ja toivomuksia tutkimusaineistojen laadulliseen ja määrälliseen tarkasteluun sekä arviointiin liittyen.

Raportissa on käytetty pohjana paitsi tutkimuskirjallisuutta, laadittu myös empiirinen aineisto kyselytutkimuksella tutkijoilla ja tutkimusaineistojen parissa työskenteleville.

**MDC 2015b - Making data count - Kratz, J. & Strasser, C. (2015) - 5 s.**

Edelliseen raporttiin liittyen julkaistu erillinen kyselytutkimuksen pohjalta muodostettu tuotos, jossa paneudutaan erityisesti tutkijoiden kokemuksiin tutkimusaineistojen jatkokäytöstä ja arvioinnista. Yhteensä kyselytutkimukseen vastasi 247 eri tutkijanuran vaiheissa toimivaa aktiivista tutkijaa.

**RDA 2016 - Identification of Reproducible Subsets for Data Citation, Sharing and Re-Use - Rauber, A., Asmi, A., van Uytvanck, D., Pröll, S. (2016) - 10 s.**

Lyhyt yhteenveto RDA:n (Research Data Alliance) osana toimivasta WGDC-työryhmän työstä (Working Group on Dynamic Data Citation). Yhteenvedossa pureudutaan tarkemmin tutkimusaineistoihin ja niiden luonteeseen nimenomaan viittauskäytäntöjen kautta, ja siihen millaisia haasteita viittauksissa saattaa ilmetä tutkimusaineistojen muuttuvasta luonteesta johtuen. Näihin liittyen raportissa on julkaistu yhteensä 14 suositusta liittyen tutkimusaineistoihin ja niiden jakamiseen sekä jatkokäyttöön.

Suosituksen perustana on useampi asiantuntijoiden lausuntokierros sekä suositusten konseptien testaus valikoiduissa tutkimusinfrastruktuureissa.

## 4. Analyysi

Metasynteisiin kuuluvan vastavuoroisen käännöksen myötä muodostuu tämän tutkielman analyysi. Siinä aineistona toimivien raporttien ja selvitysten käsitteitä, fraaseja ja näkökulmia on yhtenäistetty tarvittavin osin ja tehty luokittelu ensin laajempiin kokonaisuuksiin, sekä erikseen näiden kokonaisuuksien sisälle kuuluviin aihepiireihin.

Tämän luvun alaluvuissa käydään läpi kaikki analyysin perusteella muodostuneet tutkimusaineistojen tutkimusmetriikkaa koskevat kokonaisuudet sekä erikseen vielä niiden sisältämät aihepiirit. Kukin näistä kokonaisuuksista sekä niiden yhteydet tutkimusmetriikkaan esitellään ensin yleisellä tasolla. Kokonaisuuksien sisältämät aihepiirit käydään läpi selvityksistä ja raporteista muodostuvien erilaisten näkökulmien kautta, mutta lisäksi kunkin aihepiirin lopulla muodostetaan analyysin tuottama synteesi nimenomaan tutkimuskysymyksen asettamasta näkökulmasta.

Analyysin yhteenvedona ja eräänä merkittävistä lopputuloksista syntyi myös koko aineiston luokittelu, joka esitetään tämän luvun viimeisessä alaluvussa.

### 4.1. Tutkimusaineistojen ominaisuudet

Tutkimusaineistot eroavat muista tutkimustuotoksista huomattavasti muutamien tutkimusaineistoille tyypillisten ominaisuuksien vuoksi. Nämä ominaisuudet liittyvät muun muassa tutkimusaineistojen muotoon, elinkaareen sekä tutkimusaineistojen koontiin liittyviin tehtäviin. Toisaalta tutkimusaineistojen ominaisuuksista johtuen tutkimusaineistojen viittauksessa on havaittavissa muutamia tutkimusjulkaisuista selkeästi poikkeavia ilmiöitä. Kun tutkimusaineistot ovat tutkimusmetriikan kohteina, on näiden ominaisuuksien huomiointi vähintäänkin tärkeää, sillä ne vaikuttavat huomattavasti mahdollisiin tulkintoihin tehdyistä tutkimusmetrisistä analyyseista.

#### Rakeisuus

On tyypillistä, että tutkimusaineisto muodostuu useammasta pienemmästä osajoukosta, jotka käsittelevät samaa ilmiötä. Esimerkkejä tällaisista osajoukoista voivat olla esimerkiksi kyselyaineiston eri teemaiset kysymykset ja mittausaineistojen eri parametrit tai koetulokset. Myös tutkimusaineiston ollessa esimerkiksi kokonainen tietokanta, saattaa tutkimusaineistosta muodostua myös tutkijan näkökulmasta useammasta pienemmästä hyödynnettävästä osasta koostuva kokonaisuus. Tällöin mielenkiinto

kohdistuu siihen, millainen *rakeisuus* tutkimusaineistolla on (CODATA 2013, 35; RDA 2016, 6-7). Rakeisuus (eng. *granularity*, suom. myös granulariteetti), viittaakin nimensä mukaisesti tutkimusaineiston eri osasten muodostamien osajoukkojen määrään. Tutkimusaineisto saattaa olla vähärakeinen ja täten kohdistua tiettyyn tarkasteltavan ilmiön osaan, jolloin se muodostaa tutkimuksen sekä tutkimusmetriikan puolesta selkeän kokonaisuuden, johon viittauksen tekeminen on varsin yksiselitteistä (RDA 2016, 6-7).

Tutkimusmetriikan kannalta on kuitenkin hankalaa käsitellä tutkimusaineistoja, jotka ovat erittäin rakeisia. Tällöin esimerkiksi viittauksen tekeminen ei ole enää yksiselitteistä, koska tutkimusaineisto saattaa sisältää havaintoja tai tuloksia laajasti ilmiöön liittyen ja tutkimuksessa onkin saatettu käyttää vain todella pientä osaa koko tutkimusaineistosta. Tutkimusmetriikkaa sovellettaessa tulisi kuitenkin pystyä havaitsemaan myös näitä pienempiä tutkimusaineistojen osia, sillä muutoin tuloksena saattaa olla pienten osasten tarkoituksenmukainen yhdistely laajoiksi tutkimusaineistojen kokonaisuuksiksi esimerkiksi viittauksia tavoiteltaessa (CODATA 2013, 35; KE 2013b, 14).

## Versiointi

Tutkimusaineistoilla on tyypillistä muuttua elinkaarensa aikana useitakin kertoja, jolloin tärkeäksi nousee tutkimusaineistojen *versiointi* ja sen huomiointi tutkimusmetriikan kannalta. Versioinnilla tarkoitetaan tilannetta, jossa esimerkiksi tutkimusaineistosta saattaa löytyä puutteita sen julkaisemisen jälkeen jolloin tutkimusaineistojen tekijät voivat käydä korjaamassa virheitä tutkimusaineistoon. Tällöin korjattu tutkimusaineisto muodostaa tutkimusaineistosta uuden version (CODATA 2013, 35; RDA 2016, 6). Versiointia saatetaan hyödyntää myös tietokantojen tapauksessa, jolloin esimerkiksi vuositasolla tarkasteltavan ilmiön lukemia käydään aina vuosittain täydentämässä ja täten tutkimusaineistosta muodostuu aina uusi versio (JISC 2016, 4).

Tutkimusmetriikan kannalta on tärkeää pystyä seuraamaan mitä versiota tutkimusaineistoista tarkoitetaan, sillä esimerkiksi viittaukset on saatettu tehdä huomattavasti erilaisiin versioihin näennäisesti samasta tutkimusaineistosta. Eri versioiden myötä tutkimusaineistojen tekoon osallistuneiden tutkijoiden määrä on saattanut myös lisääntyä tai vähentyä riippuen siitä, millaisista muutoksista tutkimusaineistojen eri versioiden välillä on kyse (CODATA 2013, 9-10).

## Roolien määrittely

Siinä missä yksittäisen tutkimusjulkaisun kirjoittajiksi saatetaan luetteloida useampikin kymmentä eri tutkijaa, saattaa tutkimusaineistojen kokoamiseen, analysointiin ja valmisteluun osallistua sadoittain henkilöitä aina käytännön aineistonkeruun toteuttajasta aineiston tallentamiseen ja julkaisemiseen valmistelevasta kuratoijasta lähtien. Tutkimusaineistojen tapauksessa *roolien määrittely* on siksi tärkeää, sillä osallistumisen aste tutkimusaineiston muodostamisessa on voinut vaihdella hyvinkin paljon eri henkilöiden välillä. Toki vastaavaa vaihtelua on myös tutkimusjulkaisujen tekoon osallistumisessa hieman tieteenalan käytännöistä riippuen (Puuska & Miettinen 2008), mutta tutkimusaineistojen kohdalla eri roolien määrittelyyn tulee kiinnittää erityistä huomiota sillä tutkimusaineistojen koontiin osallistuu tavallisesti laajempi joukko ihmisiä erilaisilla työpanoksilla (KE 2013, 7-8). Esimerkiksi CODATA ehdottaa raportissaan (CODATA 2013, 35), että tutkimusaineistojen tallennuksessa ja julkaistaessa tulisi voida määritellä erilaisia tutkimusaineistolle rooleja sen mukaan, miten eri henkilöt ovat osallistuneet tutkimusaineiston koontiin.

Tutkimusmetriikan näkökulmasta eri roolien arvottaminen voi osoittautua hankalaksi, sillä bibliometriikan käyttämien viittausanalyysien lähtökohtana on ollut tasapäistää jokainen kirjoittaja analyyseissa. Tutkimusaineistojen tapauksessa erilaisten roolien työmäärän vaihtelu tulisi huomioida tutkimusmetriikassa, sillä työpanos eri roolien välillä voi olla huomattava.

## Uudet tutkimusaineistomuodot

Vaikka tätä työtä varten tutkimusaineistoja käsitellään tutkimusmetriikan näkökulmasta varsin homogeenisena joukkona tutkimustuotoksia, ovat ne tosiasiaassa joukko täysin erityyppisiä niin sisällöltään, kohteeltaan kuin muodoltaankin. Raporteissa on kuitenkin keskitytty vielä erityisesti pohtimaan esimerkiksi sosiaalisen median, kauppojen asiakastietokantojen sekä matkapuhelinten sijaintitietojen muodostamia aineistoja, jotka tarjoavat valtavasti materiaalia esimerkiksi tiedonlouhinnan, käyttäytymistieteiden ja tulevaisuuden tutkimuksen näkökulmasta (CODATA 2013, 8-9). Nämä *uudet tutkimusaineistomuodot* asettavat haasteita siihen, kuinka yksittäistä tutkimusaineistoa voidaan ylipäänsä lähestyä yksittäisenä entiteettinä.

Tällaiset aineistot muodostavat tutkimusmetriikan kannalta erityisiä haasteita, sillä esimerkiksi sosiaalisen median muodostamisen tutkimusaineistojen tapauksessa on



vaikeaa määritellä, mitä osaa koko aineistosta on käytetty ja kuka sen on kerännyt ja miten. Toisaalta nämä tutkimusaineistot ovat myös luonteeltaan erittäin dynaamisia ja aineisto saattaa muuttua ja sen osia puuttua hyvin nopeasti tutkimusaineiston julkaisemisen jälkeen jolloin esimerkiksi viittauksen tekeminen hankaloituu (CODATA 2013, 9).

## Viittauksen merkitys

Akateeminen käytäntö viittaamisessa juontaa juurensa tutkimusjulkaisuihin, ja on periytynyt digitaaliseen muotoon ajalta jolloin tutkimusartikkelit ja kirjat oli saatavissa ainoastaan paperisina versioina. Kun samaa viittauskäytäntöä sovelletaan myös tutkimusaineistoihin, on syytä pohtia mikä aidosti on *viittauksen merkitys* tutkimusaineistojen tapauksessa.

Kun tutkija viittaa tutkimusjulkaisussaan toiseen tutkimusjulkaisuun, on useimmiten perusteena käsitellä toisen tutkijan kirjoittamaa johonkin ilmiöön liittyen. Viittaus ei kuitenkaan kerro, onko tutkija ammentanut toisen tutkijan teksteistä jotakin tai vahvistanut jotakin löydöstä, sillä yhtä hyvin viittauksessa saatetaan ottaa kriittisesti kantaa toisen tutkijan kirjoitukseen ja ollakin siis täysin eri mieltä. Viittaus on kuitenkin tehty ja tutkimusmetriikka ei tällä hetkellä pysty mitenkään erottamaan, onko viittaus tehty ilmiöön liittyen yhteisymmärryksen merkiksi vai ovatko tutkijat olleet asioista eri mieltä (DCC 2015, 3-6).

Tutkimusaineistojen kohdalla viittauksen merkitys korostuu, sillä tutkimusjulkaisun tekemä viittaus tutkimusaineistoon saattaa olla huomattavan tärkeä esimerkiksi tapauksessa, jossa viitattu tutkimusaineisto muodostaa koko tutkimusjulkaisun analyysin pohjan ja toimii täten uuden tutkimuksen pohjana. Toisaalta tutkimusaineistoon saatetaan viitata myös tapauksessa, jossa se vain läheisesti liittyy ilmiön tutkimiseen eikä tutkija ole viitattua tutkimusaineistoa edes välttämättä avannut tai ladannut käyttöönsä (DCC 2015, 3-6; JISC 2016, 2-3; MDC 2015b, 1-2). KE:n raportti (KE 2013b, 28) toivoisikin viittauskäytäntöihin muutosta niin, että viittauksesta voisi selvittää miksi viittaus on ylipäänsä tehty. Tällaisia voisivat olla esimerkiksi viittaukset, joissa tutkija on testannut tai kalibroinut tutkimusaineiston pohjalta jotakin tai joissa tutkimusaineiston sisältämiä tuloksia on verifioitu. Tutkimusmetriikan näkökulmasta analyysit muuttuisivat tietenkin monimutkaisemmiksi, mutta niiden tuottamat tulokset antaisivat aidomman näkymän tutkijoiden väliseen tutkimusaineistojen käyttöön.

## Viittausten suunnat

Tutkimusaineistojen tapauksessa viittausten lisäksi huomiota voidaan kiinnittää myös siihen, mitkä ovat tehtyjen *viittausten suunnat* ja kuinka viittauksia esimerkiksi tutkimusaineistojen välillä voidaan tehdä. Kuten edellä on jo todettu, tutkimusjulkaisuissa on käytössä vakiintunut käytäntö viittausten tekemiseen, jolloin viittaukset löytyvät joko tekstin joukosta tai erikseen julkaisun lopusta. Tutkimusaineistoja voidaan kuitenkin julkaista sekä niitä varten oleviin julkaisusarjoihin (*data journal*) että tutkimusaineistoille tarkoitettuihin arkistoihin (*data archive*). Julkaistaessa tutkimusaineistoja osana julkaisusarjoja, viittauksia toisiin tutkimusjulkaisuihin ja -aineistoihin on mahdollista tehdä ja ne voidaan nykyisin sovelluksin myös kerätä tutkimusmetriikan kannalta keskeisiin tietokantoihin.

Mikäli tutkimusaineisto kuitenkin julkaistaan niitä varten perustettuun tutkimusaineistojen arkistoon, on vaikeampaa toteuttaa tutkimusaineistosta poispäin tehtyjä viittauksia, sillä laajempaa joukkoa viittauksia ei välttämättä voi viedä esimerkiksi osaksi tutkimusaineiston kuvailevia tietoja (CODATA 2013; DCC 2015, 5; JISC 2016, 2). Tutkimusmetriikan näkökulmasta ei siis voitaisi näissä tapauksissa luottaa viittausten kertymiseen, sillä niiden lisääminen on teknisesti haastavaa tai täysin estetty hieman arkistosta riippuen.

## 4.2. Julkaiseminen

Jotta tutkimusaineistoa voidaan käsitellä osana tutkimusmetriikan analyyseja, sen tulee olla julkaistu tai tallennettu. Julkaisemisella tarkoitetaan tässä yhteydessä sitä, että tutkimusaineisto sijaitsee verkon kautta saavutettavassa paikassa julkaisemalla se joko tutkimusaineistojen arkistossa tai osana tutkimusaineistoille suunnattua julkaisusarjaa, eikä esimerkiksi tutkijan omalla tietokoneella tai muulla tallennusmedialla. Ilman julkaisemista tai tallennusta tutkimusaineiston olemassaolosta ei voida tutkimusmetriikan näkökulmasta tietää, sillä siitä ei jää jälkeä esimerkiksi eri tietokantoihin tai muihin tutkimusaineistojen tallentamisen mahdollistaviin palveluihin. Seuraavat ominaisuudet liittyvät tutkimusaineistojen julkaisemiseen ja julkaisemista seuraaviin viittauksiin sekä näistä tutkimusmetriikan näkökulmasta juontuviin haasteisiin.

## Erilaiset viittauskäytännöt

Koska tutkimusaineistojen julkaisemista tukevia palveluita on ollut saatavilla vasta viime vuosikymmeninä, vaivaa tutkimusaineistojen julkaisemista eri tieteenalojen omaksumat *erilaiset viittauskäytännöt*. Viittauskäytännöt eroavat aineistona olevien raporttien perusteella paljolti nimenomaan tieteen- tai tutkimusalan perinteiden perusteella (DCC 2015, 5; HEFCE 2015b, 101-102). Tutkimusaineistojen julkaisemisessa ja niihin viittaamisessa ei kuitenkaan ole havaittavissa juuri muutoksia. Esimerkiksi MDC:n raportissa (MDC 2015b, 3-4) viitataan tilanteeseen, jossa vuonna 2011 ainoastaan noin 17 prosentissa tutkimusjulkaisuista on viitattu johonkin tutkimusaineistoon ja tuo luku on pysynyt vuoden 1995 vertailutilanteeseen verrattuna lähes samana.

Ongelmana viittauskäytännöissä on myös se, että viittaukset tutkimusaineistoihin tehdään tutkimusjulkaisuissa yhä vaihtelevalla tavalla, koska standardia viittausmenetelmää tutkimusaineistoille ei ole (NISO 2016, 10). Tutkimusmetriikan puolesta on myös toisinaan epäselvyyttä, onko viittaus tarkoitettu tehtäväksi nimenomaan tutkimusjulkaisuun vai mahdollisesti sen pohjana olleeseen tutkimusaineistoon (CODATA 2013, 49-50).

Näitä puutteita paikkaamaan on useammassa raportissa suositeltu siirtymistä standardoituihin viittaustapoihin tutkimusaineistojen tapauksessa ja muodostettu erilaisia suosituksia viittauskäytäntöjen edistämiseksi (CODATA 2013, 6-7 ; FORCE11 2016, 3-4; KE 2013b, 12-15). Eräässä raportissa on myös huomioitu itsekriittisesti se, että viittauskäytäntöjä ja niiden suosituksia on tapana edistää pitkälti tutkimusinfrastruktuuria tukevien organisaatioiden toimesta eikä suinkaan tutkijälähtöisesti (JISC 2016, 7).

## Eroavat julkaisumallit

Tutkimusaineistojen julkaisemiseen ja niihin viittaamiseen vaikuttavat myös *eroavat julkaisumallit*. Havaittavissa on selkeä erottelu tutkimusaineistojen julkaisemisessa joko niitä varten perustettuihin tutkimusaineistojen arkistoihin tai vastaaviin palveluihin omana entiteettinä tai vastaavasti erilaisiin tutkimusaineistojen julkaiseviin julkaisusarjoihin, joissa tutkimusaineiston yhteyteen julkaistaan myös artikkelimuotoinen teksti. Jälkimmäisen käyttöön on ohjannut perinteinen tutkimusjulkaisujen julkaisumalli ja yksi merkittävä insentiivi tutkijalle on tällöin saada tutkimusaineistolleen tutkimusmetriikan

näkökulmasta paremmin havaittavissa oleva viittaus, sillä tällöin viittaus tapahtuu tutkimusjulkaisujen viittauskäytäntöjen mukaan (DCC 2015, 5; KE 2013b, 12-14).

Näiden kahden mallin lisäksi Knowledge Exchangen raportissa (KE 2013b, 14-15) kiinnitetään huomiota myös muutamaankin tutkimus- tai tieteenalakohtaiseen tapaukseen, joihin edellä esitetyt tutkimusaineistojen julkaisumallit eivät välttämättä toimi. Ensimmäisenä näistä tapauksista mainitaan sellainen tutkimusalue, jossa tuotetaan jatkuvasti paljon samankaltaista tutkimusaineistoa, esimerkiksi luonnontieteellisillä aloilla, joihin liittyy paljon mallintamista tai mittauskokeita. Tällöin tutkimusaineistoja ei ole mielekästä julkaista tutkimusaineistoille varatuissa julkaisusarjoissa tutkimukselle luonteenomaisesta tiheästä tuottamisrytmistä johtuen. Vastaava tapaus on myös vahvasti kokeiluihin perustuvissa tutkimuksissa, joissa esimerkiksi laboratoriotyön tuotteena on tutkijoiden tukena toimivien henkilöiden tutkimusaineistoja. Nämä tutkimusaineistot valmistellaan julkaisua varten usein jonkin muun henkilön kuin tutkijan toimesta ja valittu julkaisumalli saattaa tällöin olla merkittävässä roolissa. Kolmantena merkittävänä haasteena ovat alat, joilla tehdään tutkimusta semanttisten tutkimusaineistojen avulla. Tällöin tutkimusaineistoille valittujen julkaisumallien tulee tukea suuresta määrästä pieniä paloja koostuvia tutkimusaineistoja, joka voi asettaa rajoituksia julkaisemisen suhteen.

## Vanhojen julkaisutapojen huomiointi

Kun tutkimusaineistojen julkaisemiseen on tarjolla useampi eri julkaisumalli eikä vakiintunutta käytäntöä ole, on tutkimusmetriikan näkökulmasta aiheellista pohtia miten *vanhojen julkaisutapojen huomiointi* tulisi hoitaa mikäli tulevaisuudessa siirrytään vakiintuneempiin julkaisumalleihin tutkimusaineistojen tapauksessa (CODATA 2013). JISC:in raportissa (JISC 2016, 8) todetaan, että tutkimusaineistojen suhteen tulee väistämättä olemaan pidempi kausi, jolloin tutkimusmetriikan näkökulmasta tietoja tutkimusaineistoista kyllä on saatavissa, mutta nämä tiedot eivät ole yhteneväisiä tai luotettavia. Myös Knowledge Exchangen raportissa (KE 2013b, 12-14) kyseenalaistetaan tämän siirtymäkauden aikana julkaistujen tutkimusaineistojen mahdollisuudet tulla huomioituiksi tulevaisuudessa mahdollisissa tutkimusmetriikan analyyseissa, sillä tiedot tutkimusaineistoista ovat vääjäämättä puutteellisia.

## Hajaantuminen eri alustoille

Vakiintuneiden julkaisumallien puutteesta johtuu myös tutkimusaineistojen julkaisemisen ja tallentamisen *hajaantuminen eri alustoille*. Vaikka tutkimusjulkaisuja on yhtä lailla mahdollista tallentaa ja julkaista eri julkaisusarjoissa ja näiden julkaisusarjojen palveluissa, on tutkimusaineistojen tallentamiseen tarkoitettut arkistot ja toisaalta tutkimusaineistoja julkaisevat julkaisusarjat teknisiltä toteutuksiltaan kauempana toisistaan, kuin tutkimusjulkaisujen vastaavat palvelut (HEFCE 2015a, 136-140). Osaltaan tämä johtuu siitä, että tutkimusjulkaisut ovat kuitenkin lopulta vain eri tavoin muotoiltua tekstiä, joiden julkaisemiseen ei ole mahdollistakaan tuottaa toisistaan radikaalisti erilaisia palveluita. Toisaalta tutkimusjulkaisujen julkaiseminen käytäntönä vakiintuneempaa ja tämän vakiintumisen seurauksena myös niiden ympärille kehitetyt julkaisemisen palvelut sekä julkaisuarkistot muistuttavat väistämättä toisiaan (MDC 2015a, 2-4; 2015b, 2).

Tutkimusaineistojen osalta eri julkaisu- ja tallentamisalustojen kirjo on laaja ja niiden tekniset toteutukset voivat erota huomattavasti toisistaan, minkä seurauksena esimerkiksi tutkimusaineistoista saatavilla olevan tiedon määrä voi vaihdella (HEFCE 2015a, 137). Hajaantumisella on siis vaikutuksensa tutkimusmetriikkaan, sillä tutkimusaineistoista saatavilla olevan tiedon tulisi olla mahdollisimman yhteneväistä ja suuria aukkoja tiedoissa ei tulisi olla. Disruptoitunut tutkimusaineistojen julkaiseminen eri alustoille saattaa siis osaltaan vaarantaa tällaisen tiedon saatavuuden jatkossa, mikäli niitä ei saada ominaisuuksiltaan yhtenäisemmiksi.

## Salassapidettävät tutkimusaineistot

Tutkimusaineistojen osalta on myös tilanteita esimerkiksi yritys yhteistyön tai tutkimuksen kohteen vuoksi, jolloin tutkimusjulkaisua ja sen pohjana toiminutta tutkimusaineistoa ei välttämättä voi juridisten, sopimuksellisten, tekijänoikeudellisten tai muiden syiden vuoksi julkaista. Ongelmaksi muodostuu tällöin se, että tutkimusaineistoon ei välttämättä ole pääsyä, sillä sitä ei ole julkaistu tai tallennettu sellaiseen arkistoon tai palvelun joka mahdollistaisi tutkimusaineiston tai sen kuvailevan tiedon louhimista esimerkiksi tutkimusmetriikan käyttämien sovellusten avulla. *Salassapidettävät tutkimusaineistot* voivatkin synnyttää tilanteen, jossa tutkimusaineisto ei päädy tutkimusmetriikan kannalta olennaisiin tietokantoihin (KE 2013b, 13). Vapaata pääsyä tutkimusaineistoille ja niiden kuvailevaan tietoon pidetään esimerkiksi CODATA:n raportissa (CODATA 2013, 9-10)

erittäin oleellisena toimivan ja avoimen viittauskäytännön toimintaan saattamisessa ja sen omaksumisessa koko tiedeyhteisöön.

## Arvioinnin puute

Useassa raportissa nostetaan esiin huoli myös tutkimusaineistojen julkaisemisen suhteen siitä, että niitä vaivaa usein *arvioinnin puute*. Siinä missä tutkimusjulkaisujen kohdalla vertaisarvioinnin läpikäyneen julkaisun voi katsoa täyttäneen tieteellisen tutkimuksen kriteerit, ei vastaavaa vertaisarviointia hyödynnetä tutkimusaineistojen kohdalla ainakaan kattavasti. Tutkimusaineiston on esimerkiksi mahdollista julkaista arkistoon, ilman että se on käynyt vertaisarviointia tai muuta vastaavaa laadullista arviointia (HEFCE 2015a, viii). Knowledge Exchangen raportti (KE 2013b, 13) nostaa tärkeäksi suositukseksi nimenomaan tutkimusaineistojen laadunarvioinnin parantamisen ja sen, että tutkimusaineistojen julkaisupalvelut tai tallentamiseen tarkoitetut arkistot mahdollistavat ja jatkossa edellyttävät vakiintuneita laaduntarkastuksen prosesseja. Toisessa raportissaan (KE 2013a, 5-6) he myös korostavat, että tutkimusaineistojen laadusta vastaa lopulta tutkimusyhteisön jäsenet, sillä heillä on paras tietämys arvioida tutkimusaineistoja ja niiden sisältöä.

Tutkimusmetriikan näkökulmasta arvioinnin puute hankaloittaa tutkimusmetriikan analyysiin soveltuvien tutkimusaineistojen valintaa. Mikäli tutkimusaineistojen kohdalla ei julkaisuvaiheessa sovelleta laadunarviointia tai muuta, esimerkiksi vertaisarviointia, tulee tutkimusmetriikan tietokantoihin mahdollisesti myös tietoja tutkimusaineistoista, jotka eivät täytä tieteellisen tutkimuksen kriteereitä.

## 4.3. Tekniset valmiudet

Tutkimusaineistot käsittelyyn, tallentamiseen ja julkaisemiseen käytettävät työkalut ja palvelut eroavat huomattavasti tutkimusjulkaisujen vastaavista. Ero johtuu pitkälti näiden tutkimustuotosten eroavaisuuksista käytännössä. Siinä missä perinteinen tutkimusjulkaisu on pitkälti tekstiä mahdollisine kuvineen, taulukkoineen, viittauksineen sekä kuvailutietoineen on tutkimusaineistojen sisältö huomattavasti heterogeenisempää. Tästä syystä tutkimusaineistojen kohdalla on raporteissa nostettu esiin muutamia teknisiä valmiuksia, jotka tulee tiedostaa ja huomioida jotta tutkimusmetriikan näkökulmasta tutkimusaineistoja voitaisiin analysoida vakuuttavammin.

## Tunnisteiden merkitys

Lähes kaikissa raporteissa nostettiin esiin pysyvien *tunnisteiden merkitys* tutkimusaineistoille ja näiden hyödyntäminen tutkimusaineistojen tunnistamisessa. Tunnisteella tarkoitetaan yksittäiselle tutkimusaineistolle asetettua, hieman tunnisteiden standardista riippuen, merkkisarjaa jonka avulla tutkimusaineisto voidaan tunnistaa huomattavasti paremmalla tarkkuudella verrattuna esimerkiksi tilanteeseen, jossa tunnistus perustuu tutkimusaineiston nimeen tai sen tekijöihin (CODATA 2013, 33-34; DCC 2015, 5; KE 2013b, 14; NISO 2016, 12). Raporttien perusteella vallitsevaksi standardiksi on muodostumassa DOI (*digital object identifier*), jossa viittaus tehdään esimerkiksi seuraavanlaisella merkkisarjalla 10.1000/xyz123 (NISO 2016, 12). Siinä missä DOI vastaa tarpeeseen nimenomaan yksittäisten tutkimustuotosten tunnistamisessa, on yhtä lailla tutkijoiden tunnistamiseen muodostumassa käytäntö, joka perustuu ORCID-tutkijatunnisteeseen (*Open Researcher and Contributor ID*) käyttöön (NISO 2016, 12). Sekä tutkimustuotos- että tutkijakohtaisten tunnisteiden käytön edistämässä keskeisessä osassa ovat myös sen käyttöä tukevat työkalut ja palvelut, jolloin tunnisteiden hyödynnettävyys paranee ja niiden käyttö entisestään lisääntyy (DCC 2015, 5; NISO 2016, 12).

Syy miksi tunnisteet ovat erittäin keskeisiä myös tutkimusmetriikan näkökulmasta perustuu nimenomaan tunnisteiden kautta saavutettuun tunnistamiseen ja pysyvyyteen. Eräs tutkimusmetriikkaa ja bibliometriikkaa pitkään vaivannut haaste on liittynyt nimenomaan viittausten täsmällisyyteen ja tutkimusjulkaisujen tunnistamiseen. Mikäli tutkimusjulkaisussa on viitattu toiseen tutkimusjulkaisuun huomattavasti tavallisesta poikkeavalla viittaustavalla ja viitatus tutkimusjulkaisun nimi-, tekijä- tai muut tiedot ovat puutteellisia, on tämä viittaus saattanut jäädä huomiotta tutkimusmetriikan tietokannoista. Kun viittauksen tueksi tuodaan myös yksilöivä merkkisarja, ei viittauksen liity enää vastaavaa epävarmuutta. Pysyvien tunnisteiden hyödyntäminen lisää myös näiden viittausten koneluettavuutta ja tietokantojen välistä yhteentoimivuutta (KE 2013b, 14; NISO 2016, 12).

## Viitatus osan tunnistaminen

Eräs RDA:n ja CODATA:n raporteissa (CODATA 2013, 35; RDA 2016, 6-7) huomioitu tekninen seikka korostaa sitä, että tutkimusaineistojen *viitatus osan tunnistaminen* tulisi pystyä huomioimaan sekä tutkimusaineistojen arkistoissa ja muissa palveluissa että viittauksissa. Tällöin ajattelu yhdestä tutkimusaineistolle asetetusta pysyvästä

tunnisteesta viedään askeleen pidemmälle, jolloin tutkijan tai muun palvelun olisi mahdollista viitata tutkimusaineiston yksittäiseen osa-alueeseen (RDA 2016, 6-7). Tämä on vahvasti yhteydessä aiemmin esitettyyn tutkimusaineistojen ominaisuuteen niiden rakeisuudesta. Jotta yksittäisiä tutkimusaineistojen osia voitaisiin luotettavasti tunnistaa, vaadittaisiin työkalujen ja palveluiden kannalta uudenlaista dynamiikkaa, jossa tunnisteet yhdistettäisiin esimerkiksi koneiden välisiin kyselyihin (*query*) ja tätä kautta käytetty tutkimusaineistojen osa-alue tarkentuisi (RDA 2016, 6-7). Tutkimusmetriikan tulisi tällöin myös huomioida nämä viittaukset tutkimusaineistoon, vaikka koko tutkimusaineistoa ei olisikaan käytetty. Se, miten tällaista voitaisiin aidosti implementoida tutkimusmetriikkaan ja sen analyysiin, ei kuitenkaan raporteissa käsitelty.

## Teknisten järjestelmien toimivuus

Jotta tutkimusaineistojen voitaisiin tutkimusmetriikan näkökulmasta tarkastella luotettavasti, vaaditaan raporteissa myös sitä, että *teknisten järjestelmien toimivuus* on tarpeeksi hyvällä tasolla. Tällä viitataan hieman raportista riippuen eri aspekteihin, mutta keskiössä on kuitenkin eri palveluiden, työkalujen ja järjestelmien välinen yhteentoimivuus. Yhteentoimivuuden varmistamiseksi tutkimusaineistojen ympärille rakennettujen järjestelmien tulisi olla mahdollisimman pitkälle standardisoituja niin niiden käyttämien formaattien, indikaattoreiden kuin semantiikankin osalta, jotta järjestelmätasolla ei synny erilaisia määrytyksiä samoille ilmiöille tai ominaisuuksille (EC 2017, 15) Tämä korostuu erityisesti siitä syystä, että tutkimusaineistojen ympärille rakentuneet palvelut ja järjestelmät ovat verrattain tuoreita ja kuten jo käsitelty, tutkimusaineistot ovat disrupoituneet erilaisille alustoille (HEFCE 2015a, 136-140).

Toinen tekniseltä kannalta esiin nostettu haaste liittyy tutkimusaineistojen tunnistamiseen järjestelmätasolla. Kun tutkimusaineistoja tallennetaan ja julkaistaan useamman julkaisumallin mukaisesti, on vaarana esimerkiksi tutkimusaineistojen julkaisemisessa osana julkaisusarjoja se, että tutkimusmetriikka ei pysty erottamaan tutkimusaineistoa tutkimusjulkaisusta (JISC 2016, 1). Jos tutkimusaineistojen julkaisumallit ja viittauskäytännöt lähestyvät tutkimusjulkaisuista tuttua viittauskäytäntöä, ei viittauksen perusteella voida välttämättä tehdä päätelmiä siitä, millaiseen tutkimustuotokseen on viitattu. Tutkimusmetriikankin näkökulmasta on oleellista tietää, mistä tutkimustuotoksesta on kyse, mikäli halutaan esimerkiksi tarkastella vain ja ainoastaan tutkimusaineistoja tai tutkimusjulkaisuja (JISC 2016, 1-2).



## Tiedonlouhinnan mahdollisuus

NISO:n raportti (NISO 2016, 9) esittää myös tärkeäksi ominaisuudeksi sen, että järjestelmiin rakennetaan *tiedonlouhinnan mahdollisuus* jo varhaisessa vaiheessa. Tällä tarkoitetaan esimerkiksi sitä, että tutkimusaineistoille tarkoitetuista arkistoista ja muista palveluista on saatavilla koneluettavassa muodossa ja vapaasti esimerkiksi viittauksiin, kyselyihin ja käyttömääriin liittyvää tietoa. Näin tuetaan järjestelmien välistä tietojen hyödyntämistä ja tutkimusmetriikankin kannalta olennaisten tietojen siirtymistä.

### 4.4. Tutkijan toiminta

Jotta tutkimusaineistojen julkaisemista ja tallentamista ylipäänsä tapahtuisi, tarvitaan siihen tutkijoiden työtä. Tutkimusaineistojen julkaiseminen osana muuta tutkimustoimintaa ei tällä hetkellä ole lainkaan niin yleistä, kuin tutkimusjulkaisujen julkaiseminen. Osin syyt johtuvat puutteellisista työkaluista, käytänteiden puuttumisesta tai juridisista esteistä, mutta monesti julkaisematta jättämisen syynä on yhtä lailla tutkijan päätös jättää tämä vaihe tutkimustyöstä tekemättä. Raportteihin perustuen olen seuraavassa luokitellut näistä tutkimustyön kannalta keskeisimpiä ominaisuuksia, jotka vaikuttavat osaltaan myös tutkimusaineistojen tutkimusmetriikkaan.

#### Ei mahdollista meritoitumista

Keskeisimpänä ja yleisimpänä syynä tutkimusaineistojen julkaisematta jättämiseen on raporteissa mainittu se, että toistaiseksi tämä toiminta *ei mahdollista meritoitumista* tutkijan kannalta. Yleisimpänä tutkimustoiminnan laadun, vaikuttavuuden ja rahoituksen mittarina toimii tutkimustuoksista yhä tutkimusjulkaisu, joita mitataan sekä lukumäärältään että bibliometriikan analyysien kautta (JISC 2016, 3; KE 2013b, 3). Kansalliset rahoitusmallit kansainvälisesti sekä Suomessa jättävät tutkimusaineistojen julkaisemisen pitkälti huomiotta ja onkin täten täysin tutkimusryhmän, laitoksen tai tutkimusorganisaation sisäisen päätöksen mukaista, mikäli tutkimusaineistojen julkaisemisen perusteella arvotetaan tutkijan tutkimustoimintaa (CODATA 2013, 50-51). Tutkimusrahoittajien suhteen tilanne on hieman parempi ja esimerkiksi Suomessa Suomen Akatemia edellyttää rahoitettavilta tutkimuksilta myös tutkimusaineistojen julkaisemista.

Meritoitumisen puute vaikuttaa tutkimusmetriikkaan huomattavasti, sillä ilman tutkimusaineistojen julkaisemista tutkimusmetriikan käyttämiin tietokantoihin ei päädy tarpeeksi materiaalia tutkimusaineistoihin kohdistuvaa tutkimusmetristä analysointia

varten (CODATA 2013, 51). Knowledge Exchangen raportissa (KE 2013b, 27) tilanne on käännetty myös toisin päin ja todettu että mikäli tutkimusaineistoja julkaistaan edes vähäisissä määrin, niihin kohdistetut tutkimusmetriset analyysit korostavat näiden tutkimusaineistojen tärkeyttä ja tuovat tutkijoiden tutkimusaineistoille enemmän näkyvyyttä.

## Vaatii lisäresursseja

Tiiviisti tutkijan työhön liittyvä ominaisuus on myös se, että tutkimusaineistojen julkaiseminen ja tallentaminen *vaatii lisäresursseja*. Suurin näistä lisäresursseista on luonnollisesti tutkijan aika, joka kuluu tutkimusaineiston valmisteluun julkaisua varten. Tämä vaihe sisältää muun muassa tutkimusaineiston dokumentointia ja kuvailua, rakenteen, formaattien ja tiedostojen valitsemista sekä valitun julkaisumallin mukaista toimintaa, esimerkiksi tutkimusaineistojen arkistointia sille soveltuvaan arkistoon. Pelkän ajan käytön lisäksi tutkija saattaa tarvita lisärahoitusta esimerkiksi tapauksessa, jossa tutkimusaineiston käsittelyssä tarvitaan tutkijan lisäksi muita avustavia henkilöitä valmistelemaan ja kuratoimaan tutkimusaineistoja, tarkastamaan sen eheyttä tai muita julkaisemiseen liittyvien valmistelevien töiden hoitoon. (KE 2013b, 14)

## Tutkimusalojen eroavat julkaisukäytännöt

Tutkimusaineistojen julkaisemista määrittelevät myös *tutkimusalojen eroavat julkaisukäytännöt*. Useammassa raportissa on nostettu esiin se, että tutkimusaineistojen julkaisemista ohjaavat tutkimus- tai tieteenalan käytännöt ja perinteet (CODATA 2013, 51-52; KE 2013b, 3). Käytäntöjen osalta tutkimusaineistojen julkaiseminen saattaakin esimerkiksi keskittyä pääasiassa tutkimusaineistojen julkaisemista varten perustettuihin julkaisusarjoihin (KE 2013b, 12-14).

Tällä on vaikutuksia myös tutkimusmetriikkaan, sillä tehtävissä analyyseissa täytyy osata huomioida tutkimus- tai tieteenalojen julkaisukäytäntöjen eroista johtuvat poikkeavuudet tuloksissa (HEFCE 2013a; 2013b). Toisaalta tutkimusmetriikka ja sen analyysien pohjalla olevat tietokannat ovat myös sidoksissa siihen kuinka laajalle julkaisukäytännöt leviävät tutkimus- tai tieteenaloilla. Kuten Knowledge Exchangen raportissa (KE 2013b, 3) todetaankin, tutkimusmetriikan kehitys on riippuvaista tutkijoiden tutkimusaineistojen julkaisukäytännöistä, sillä juuri julkaisemisen perusteella analyyseja voidaan ylipäänsä suorittaa.

## Aineistohallinnan merkitys

Tutkimusaineistojen julkaisemisessa ja niiden huomiointi tutkimusmetriikassa on sidoksissa tutkijan työhön myös aineistohallinnan näkökulmasta. *Aineistohallinnan merkitys* korostuu nimenomaan siinä, että aineistohallinnalla voidaan pitkälti varmistaa, että tutkimusaineisto tulee myös tutkimustyön loputtua julkaistua tai tallennettua sille soveltuvassa arkistossa tai vastaavassa palvelussa. Erityisesti FORCE11:a raportti (FORCE11 2016) korostaa aineistohallinnan keskeistä asemaa tutkimusaineistojen julkaisemisessa ja sen varmistamisessa, että tutkimusaineistot pysyvät aidosti hyödynnettävissä ja käyttökelpoisina. Tutkimusmetriikan näkökulmasta aineistohallinnan edistämällä mahdollistetaan isommassa mittakaavassa tutkimusaineistojen hallitumpi julkaisemis ja tallentamisprosessi, jossa tutkimusaineistot nivoutuvat jo aikaisemmassa vaiheessa osaksi muuta tutkijan toimintaa.

## Julkaisematta jättäminen epätieteellistä

Muutama raportti korostaa sitä, että tutkimusaineistojen *julkaisematta jättäminen on epätieteellistä*. Tällöin on esimerkiksi korostettu sitä, että tutkimusaineistot ovat merkittävässä osassa, kun tutkimustoiminnassa päädytään johonkin tutkimustulokseen. Tutkijan tehtäviin kuuluisi tällöin läpinäkyvyyden ja toistettavuuden varmistamiseksi toimia avoimesti myös tutkimusaineistojen julkaisemisen osalta, sillä juuri ne ovat keskeisessä osassa monessa tutkimustuloksessa. Knowledge Exchangen raportissa (KE 2013a, 1) viitataan tutkimusaineiston julkaisematta jättävän tutkijan toimivan tällöin periaatteessa epätieteellisten käytäntöjen (*scientific malpractice*) mukaisesti. DCC:n raportissa (DCC 2015, 2-3) lähestytään samaa ilmiötä toisesta suunnasta, ja korostetaankin nimenomaan tutkimusaineistojen julkaisemisen lisäävän paitsi tutkimuksen laatua ja läpinäkyvyyttä, myös varmistavan myös sen jatkokäyttö. On kuitenkin tosiasia, että tutkimusmetriikan näkökulmasta julkaisematta jätetty tutkimusaineisto ei voi päätyä tutkimusmetriisiin analyysihin, sillä tietoa tällaisesta tutkimusaineistosta ei voida saavuttaa eikä sitä toisaalta pystytä muiden tutkijoiden toimesta laajamittaisesti käyttämään.

## Ilmiöön liittyvät epäluulot

Muutamissa raporteissa on tutkimusaineistoja ja niiden julkaisemista kohtaan huomattu liittyvän myös tiettyjä oletuksia tai tutkimusjulkaisuja koskevasta tutkimusmetriikasta peräisin olevia väärinymmärryksiä. Nämä *ilmiöön liittyvät epäluulot* ovat sikäli merkittäviä, että niiden myötä tutkijoiden tutkimusaineistojen julkaisuaktiivisuus saattaa

vaihdella suurestikin. Esimerkiksi tutkimusaineistojen varastaminen (*scooping*) on toisinaan esille nostettu tutkimusaineistoja koskeva ilmiö, jossa tutkijan julkaisemista saattaa hidastaa tai kokonaan estää se, että pelätään mahdollisesti kilpailevan tutkijan tai tutkimusryhmän varastavan aineiston ja julkaisevan siihen liittyviä tuloksia. Esimerkiksi CODATA:n raportissa (CODATA 2013, 52-53) kuitenkin korostetaan sitä, että kyseinen toiminta oikeastaan hankaloituu sitä mukaan, kun tutkimusaineistoista muodostuu aidosti tutkimusjulkaisun veroinen tutkimustuotos. Tämän pisteen saavuttamiseksi tarvitaan kuitenkin koko tutkimusjärjestelmän tasolla standardoidut tavat julkaista tutkimusaineistoja ja tehdä niihin viittauksia. Pelko tutkimusaineistojen varastamisesta saattaa myös kohdistua esimerkiksi immateriaalioikeuksien menettämiseen tutkimusorganisaation näkökulmasta tai siihen, että tutkimusaineistojen keruuseen käytetyt resurssit valuvat tutkimusaineiston myötä esimerkiksi kaupalliseen toimintaan, joka hyödyntää tutkimusaineistoja toimintansa pyörittämisessä (KE 2013b, 14).

Suoraan tutkimusmetriikkaan liittyvä epäluulo kohdistuu nimenomaan siihen, että tutkimusmetriikan sovellutusten ja analyysien myötä myös tutkimusaineistojen kohdalla pätevät samat järjestelmän pelaamisen mahdollisuudet, kuin esimerkiksi bibliometriikassa on havaittavissa. Näitä ilmiöitä ovat esimerkiksi tutkimusaineistojen pilkkominen mahdollisimman pieniksi osiksi julkaisemista varten, viittausten tekeminen sovitusti tutkijoiden kesken sekä tutkimusaineistojen absoluuttisen määrän korostuminen niiden laadun kustannuksella.

## 4.5. Arvottaminen

Kaikista keskeisintä tutkimusmetriikan näkökulmasta on se, kuinka tutkimusaineistoja tulisi arvottaa ja millaisia indikaattoreita valita niiden käsittelyyn erilaisissa tutkimusmetrisissä analyyseissa. Koska tutkimusaineistoja ei juurikaan huomioida esimerkiksi kansallisissa rahoitusmalleissa tai organisaatioiden sisäisissä arvioinneissa, on tutkimusaineistojen rooli jäänyt huomattavasti tutkimusjulkaisujen asemaa vähäisemmäksi. Raporteissa on otettu kantaa muun muassa siihen, miten nykyistä tutkimusmetriikkaa ja sen käyttämiä indikaattoreita voitaisiin hyödyntää tutkimusaineistojen kohdalla, millainen suhde tutkimusaineistoilla on arvottamisen kannalta tutkimusjulkaisuihin ja mitä mahdollisia uusia indikaattoreita tutkimusaineistojen analysointiin tutkimusmetrisestä näkökulmasta voitaisiin asettaa ja millaisin reunaehdoin.

## Nykyisen tutkimusmetriikan hyödyntäminen

Aineiston perusteella *nykyisen tutkimusmetriikan hyödyntäminen* voidaan jakaa kahteen eri suuntaukseen, kun tarkastelun kohteena ovat tutkimusaineistot. Toisessa suuntauksessa pohditaan sitä, millaisia mahdollisuuksia nykyinen bibliometrinen tarkastelu tarjoaa tutkimusaineistojen kohdalla. Tässä tarkastelussa korostuvat nykyiset bibliometriset tietokannat sekä käytänteet, joiden voidaan katsoa auttavasti soveltuvan myös tutkimusaineistojen tarkasteluun (DCC 2015, 12; MDC 2015b, 1; NISO 2016, 14-15). Toinen raporteissa huomioitu tapa käsitellä tutkimusaineistoja on altmetriikka ja sen hyödyntäminen esimerkiksi tutkimusaineistojen latausmäärien, eri alustoilla jakamisten tai muiden altmetristen indikaattoreiden kautta (DCC 2015, 6; EC 2017, 9-10; NISO 2016, 16).

NISO:n raportissa (NISO 2016, 14-15) korostetaan näkemystä, jossa tutkimusaineistot tulisi nähdä osana tutkimusmetriikkaa niin, että erityisiä tutkimusaineistoille suunnattuja menetelmiä tai indikaattoreita ei tulisi valita. Tällöin samat käytännöt joita sovelletaan tutkimusjulkaisuille, tulisi hyödyntää myös tutkimusaineistojen tapauksessa. Myös MDC:n raportti (MDC 2015b, 1-4) toistaa saman näkemyksen ja tarkentaa erityisesti että tutkimusaineistot, joiden julkaiseminen tapahtuu nimenomaan tutkimusaineistojen julkaisusarjoissa, voidaan jo tälläkin hetkellä sisällyttää melko helposti tutkimusmetriisiin analyyseihin. Tämä johtuu siitä, että julkaisusarjoihin keskittynyt julkaisumalli ei juuri eroa tutkimusjulkaisujen vastaavasta, sillä viittauskäytäntö on tutkimusmetriikan näkökulmasta sama. Aivan näin vaivatonta nykyisten tutkimusmetristen käytäntöjen soveltaminen tutkimusaineistoille ei kuitenkaan ole silloin, kun tutkimusaineistot julkaistaan arkistoihin suunnatun julkaisumallin mukaisesti. DCC:n raportti (DCC 2015, 12) huomauttaa että nykyiset tutkimusmetriikan keskeiset tietokannat, Scopus ja Web of Science, laajasta aggregoinnistaan huolimatta eivät ole lähellekään yhteneväiset nimenomaan tutkimusaineistojen osalta. Tähän on keskeisenä syynä se, että tutkimusaineistoille tarkoitetut arkistot ja muut julkaisemiseen liittyvät palvelut kehittyvät ja muuttuvat niin nopeasti, että tietokannat eivät välttämättä pysy ajantasaisina. Tietokantojen haasteena on myös se, että arkistoista ja palveluista mahdollisesti mitattavien asioiden kirjo on laaja, eikä yhteisiä käytäntöjä vielä ole olemassa.

Myös altmetriikan soveltuvuutta tutkimusaineistoille on perusteltu sillä, että altmetriikan näkökulmasta ei ole juurikaan väliä, onko käsiteltävä tutkimustuotos tutkimusaineisto vai tutkimusjulkaisu, sillä samat lainalaisuudet pätevät sekä teoriassa että teknisestä näkökulmasta. Mikäli altmetriikassa tarkastellaan esimerkiksi sosiaalisen median jakamisia tai yhteistyöalustojen kontribuutioita, ei varsinaisella tutkimustuotoksella pitäisi

olla väliä, sillä toimenpiteet näissä pysyvät tuotoksesta riippumatta samoina (NISO 2016, 16). Euroopan komission asettaman työryhmän raportissa (EC 2017, 9-10) kuitenkin varoitetaan altmetriikan äkillisestä vanhenemisestä sekä kerättävien tietojen että keruun kohteena olevien palveluiden ja alustojen osalta. Samalla tosin mainitaan, että nimenomaan tutkimusaineistojen palveluiden nopeasta muutoksesta johtuen altmetriikkaa tarjoaa luultavasti nopeammin indikaattoreita ja tietoa tutkimusaineistoihin liittyen, kuin esimerkiksi bibliometriikasta tutkimusjulkaisujen tarkasteluun soveltuvat käytännöt pystyvät toistaiseksi tarjoamaan. DCC:n raportissa (DCC 2015, 6) korostetaan myös altmetriikan hyödyntämistä nimenomaan tutkimusaineistojen vaikuttavuuden tarkastelussa. Keskimäärin yksittäinen tutkimusaineisto saattaa saavuttaa tiedeyhteisön ulkopuolella laajempaa käyttöä, kuin yksittäinen tutkimusjulkaisu. Tällöin altmetriikan kautta tehty vaikuttavuuden tarkastelu soveltuisi paremmin nimenomaan tutkimusaineistoille.

## Suhde muihin tutkimustuotoksiin

Kun tutkimusaineistoja hyödynnetään osana tutkimusmetriikkaa, on tarpeellista huomioida millainen *suhde muihin tutkimustuotoksiin* tutkimusaineistoilla on. Tutkimusaineiston keruu, sen muodostaminen ja lopulta julkaiseminen saattavat toisinaan työllistää yksittäistä tutkimusjulkaisua enemmän, joten tulisko se arvottaa tutkimusmetriikan näkökulmasta tutkimusjulkaisua korkeammalle? Knowledge Exchange raportti (KE 2013b, 32) ehdottaa, että tutkimusaineisto asetettaisiin samanarvoiseksi tutkimusjulkaisuun verrattuna, jolloin edesautettaisiin tutkimusaineistojen asemaa tutkimustuotoksena tutkimusjulkaisuihin verrattuna nimenomaan tutkijan näkökulmasta ja viittausten määrät sekä käytännöt vakiintuisivat nopeammin, kun tutkimusaineistoista saatava meriitti olisi yhtäläinen tutkimusjulkaisuun verrattuna. Näkemystä tukee raportin mukaan se, että monessa tapauksessa ilman tutkimusaineistoon tehtyä viittausta tutkimusjulkaisua ei tulisi edes olla, sillä lähes aina tutkimusjulkaisut perustuvat johonkin aineistoon. Toisaalta muistutetaan, että pienen tutkimusaineiston julkaiseminen saattaa joissakin tapauksissa olla huomattavasti vähemmän työtä vaativaa, kuin laajan, uuteen näkökulmaan tai teoriaan, perustuvan tutkimusjulkaisun.

FORCE11:a raportti (FORCE11 2016, 3) korostaa näkökulmaa, jossa kaikkien tutkimustuotosten, siis myös tutkimusaineistojen, arvottaminen pitäisi perustua pikemminkin niiden laatuun ja vaikuttavuuteen, eikä suinkaan tutkimustuotoksen formaattiin. Näin ollen tutkimusmetriikan pitäisikin siirtyä arvottamisen suhteen kohti

kaikkien tutkimustuotosten laadullista arviointia, jota tarvittaessa voitaisiin hyödyntää määrällisissä analyyseissa. Siihen, kuinka tutkimusaineistojen tapauksessa tällainen laadullinen arviointi voitaisiin tehdä, ei kuitenkaan oteta kantaa yhdessäkään raportissa.

## Indikaattoreiden valinta

Tutkimusmetriikan ja sen sovellutusten näkökulmasta eräs tärkeimmistä tutkimusaineistoihin liittyvistä asioista on ehdottomasti sopivien *indikaattoreiden valinta*. Koska juuri indikaattoreiden kautta tutkimusmetriikassa tarkastellaan tutkimustuotoksiin liittyviä ilmiöitä, ja esimerkiksi bibliometriikassa vakiintuneiksi indikaattoreiksi ovat nousseet paitsi absoluuttiset viittausmäärät ja niiden keskiarvot, myös Top 10 -indikaattori sekä julkaisusarjakohtainen vaikuttavuusindikaattori JIF (*journal impact factor*). Koska tutkimusaineistokohtaisia indikaattoreita ei vielä vakiintuneessa käytössä ole, on useammassa raportissa keskitytty konseptualisoimaan ja hahmottelemaan paitsi nykyisten indikaattoreiden hyödyntämistä, myös täysin uusien indikaattoreiden mahdollisuuksia nimenomaan tutkimusaineistoille.

Kun olemassa olevia tutkimusmetriikan indikaattoreita hyödynnetään tutkimusaineistojen kohdalla, voidaan käyttö jakaa kahteen eri malliin. Toinen näistä malleista on bibliometriikasta ja tutkimusjulkaisuista tuttujen indikaattoreiden hyödyntäminen, jossa indikaattorit voidaan jakaa vielä kolmeen eri tyyppiin: kokoriippuvaisiin (*size-dependent*) joissa tarkastelu kohdistuu esimerkiksi luku- ja viittausmääriin, tuotoskohtaisiin (*direct average performance*) joissa tarkastellaan tuotoksen keskiarvo- tai mediaaniperusteisia indikaattoreita esimerkiksi Top 10 -indikaattori sekä julkaisukohtaisiin (*source-based performance*) joissa tarkastelu kohdistuu julkaisusarjaan liittyviin indikaattoreihin esimerkiksi journal impact factor (KE 2013b, 15-17). Toinen malli perustuu taas altmetriikasta peräisin oleviin indikaattoreihin, joissa tarkastelussa saattavat olla paitsi sosiaalisen median alustoilta saatavilla olevat indikaattorit, kuten katselut, suosittelut ja kommentoinnit, myös esimerkiksi akatemian käyttämien yhteisöpalveluiden (Mendeley, ResearchGate) käyttöön perustuvat indikaattorit, kuten tallennukset, pisteytykset, viittaukset ja jakamiset (HEFCE 2015b, 101-102; KE 2013b, 15-17).

Pelkästään tutkimusaineistojen tarkasteluun tarkoitettuja indikaattoreita on hahmoteltu esimerkiksi tutkimusaineistojen käyttöön liittyen (*data usage index*), joilla pyritään hahmottamaan esimerkiksi tutkimusaineistojen hyödyntämistä viittaamisen lisäksi myös latausten ja katseluiden perusteella (DCC 2015, 5-6; HEFCE 2015b, 101-102). Myös

tutkimusaineistojen viittauksiin perustuen Clarivate on pyrkinyt Web of Science tietokannan omaisesti laajentumaan myös tutkimusaineistojen puolelle lanseeraamalla oman viittausindeksinsä (*Data Citation Index*), johon on aggregoitu huomattava määrä eri tutkimusaineistoille saatavilla olevia arkistoja ja näihin kertyneitä kuvailu- ja viittaustietoja (HEFCE 2015b; NISO 2016, 10). Raporttien perusteella nimenomaan viittausmäärät ovat tutkijoiden näkökulmasta suosituimpia indikaattoreita myös tutkimusaineistojen kohdalla. Syyksi tälle esitetään sitä, että viittaukset ovat lähimpänä nykyistä meritoitumismallia, joka pohjautuu pitkälti tutkimusjulkaisuihin (HEFCE 2015a, 2015b).

Knowledge Exchangen raportissa (KE 2013a, 5-6) korostetaan, että tutkimusaineistoille valittavissa indikaattoreissa olisi hyödyllisempää ensin tarkastella millaisia indikaattoreita tutkimusaineistoihin liittyen on ylipäänsä saatavilla ja mahdollista toteuttaa, ja vasta tämän jälkeen harkita näiden vakiinnuttamista osaksi tutkimusmetriikan sovellutuksia. Näin tekemällä vältetään mahdollisesti vaikuttamasta liikaa tutkimusmetriikalla tutkijoiden ja tieteen kannalta hyviksi havaittuihin tutkimusaineistojen käytänteisiin esimerkiksi julkaisemiseen liittyen.

NISO:n raportti (NISO 2016, 9-16) nostaa tutkimusaineistoihin liittyvien indikaattoreiden osalta esiin sen seikan, että tutkimusaineistoja hyödynnetään huomattavasti tutkimusjulkaisuja enemmän myös koneiden ja palveluiden toimesta esimerkiksi yhdistelemällä, analysoimalla ja koostamalla eri tutkimusaineistoja. Näin ollen indikaattoreiden tulisi huomioida myös tällainen automaattinen koneiden välinen käyttö, eikä luottaa pelkästään ihmisperusteiseen tutkimusaineiston hyödyntämiseen.



Tutkimusmetriikan malli	Indikaattorin tyyppi	Indikaattori
Bibliometriikkaan pohjautuva	Kokoriippuvainen	lukumäärä, viittausmäärä, h-indeksi
	Tuotoskohtainen	top 10
	Julkaisukohtainen	journal impact factor
Altmetriikkaan pohjautuva	Sosiaalinen media	katselu, suosittelu, jakaminen, kommentointi
	Akateemiset yhteisöt	viittaus, jakaminen, tallennus, kommentointi, pisteytys
Hahmottumassa	Käyttöperusteinen	katselut, lataukset
	Koneellinen käyttö	haku, yhdistäminen, analysointi, koostaminen

Taulukko 1: Tutkimusmetriikan indikaattorit tutkimusaineistoille (perustuen KE 2013b, Table 3, täydennetty DCC 2015, 5-6; HEFCE 2015b, 101-102; KE 2013a, 5-6; NISO 2016, 9-16; MDC 2015a, 3-4; 2015b, 2-3).

Useampi raportti kuitenkin muistuttaa tutkimusaineistojen indikaattoreita pohdittaessa siitä, että tutkimusmetriikka ja sen määrälliset indikaattorit tarvitsevat tuekseen myös laadullista tarkastelua (KE 2013b, 15-17; MDC 2015a, 3; 2015b: 2-3). Esimerkiksi tutkimusaineistoihin liittyen on vaikeaa hahmottaa sitä, miten tutkimusaineistot aidosti vaikuttavat tutkijoihin ja tutkimusryhmiin niiden jatkokäytön näkökulmasta ja toisaalta kuinka tutkimusaineistoilla voidaan edistää esimerkiksi sovelluskehitystä tutkimuksessa ja sen ulkopuolella (DCC 2015, 3-5). HEFCE:n raportti (HEFCE 2015a, 137) korostaa myös sitä, että tutkimusaineistot ovat usein tieteenalakohtaisia ja perustuvat totuttuihin käytänteisiin, jolloin tieteenalariippumattoman tutkimusmetriikan sovellutus saattaa osoittautua erittäin haastavaksi. Taulukko 1 kokoaa yhteen eri raporteissa ja selvityksissä ilmi nousseita indikaattoreita tutkimusaineistojen tutkimusmetriikkaan liittyen.

## Läpinäkyvyys indikaattoreissa

Indikaattoreiden valinnan yhteydessä raporteissa kiinnitettiin erityishuomiota siihen, että *läpinäkyvyys indikaattoreissa* tulee olemaan tutkimusaineistojen kohdalla tärkeää. Merkittävin syy tähän on se, että tutkimusaineistoille on mahdollista valita indikaattoreita joko perustuen vakiintuneisiin tutkimusmetriikan käytäntöihin, mutta erityisesti ottaa käyttöön täysin uusia ja kehittyviin alustoihin ja palveluihin liittyviä indikaattoreita. Jotta näiden uusien indikaattoreiden vakiintumista voitaisiin edistää, on niiden oltava läpinäkyviä sekä järjestelmätasolla, jolloin saavutetaan standardeja ja yhteentoimivuutta,

mutta myös tutkijatasolla, jotta tutkimusaineistoja koskeva arvottaminen on selkeää myös niitä tuottavalle taholle (DCC 2015, 2-3; HEFCE 2015a, 137). Toinen merkittävä huomio läpinäkyvyyteen liittyy myös indikaattoreiden pohjalla oleviin tietokantoihin ja aineistoihin liittyen. Sekä bibliometriikan että erityisesti altmetriikan puolella on törmätty tilanteisiin, joissa indikaattoreiden pohjalla oleva materiaali on saatavilla ainoastaan kaupallisilta toimijoilta (EC 2017, 12-15).

## Tutkimuksen kokonaiskuvan hahmottaminen

Tutkimusaineistojen kohdalla korostuu muutamissa raporteissa myös se, että hyvin sovelletun tutkimusmetriikan kautta *tutkimuksen kokonaiskuvan hahmottaminen* saattaisi onnistua aiempaa paremmin (CODATA 2013, 53; EC 2017, 13). CODATA:n raportissa (2013, 53) viitataan tähän liittyen muun muassa siihen, että yhden tutkimusaineiston ja siihen tehtyjen viittausten kautta on mahdollista kartoittaa tutkimusaineiston ympärille liittyvää tutkimusta eri tavalla, kuin tutkimusjulkaisujen tapauksessa. Syynä tähän on se, että tutkimusjulkaisuissa on tapana viitata myös varsinaisen tutkimusalan ulkopuolisiin lähteisiin, kun taas tutkimusaineiston ollessa tutkimusjulkaisujen lähteenä, on kyseessä hyvin todennäköistä, että julkaisussa on käsitelty samaa ilmiötä, kuin tutkimusaineistossa. Toisaalta tutkimusjulkaisut eivät aina viittaa toisiinsa, vaikka käsittelevätkin lähes samaa tutkimusaihetta. Mikäli tutkimusjulkaisut kuitenkin käsittelevät samaa tutkimusaineistoa ja viittauskäytäntö on vakiintunutta, yhdistyvät nämä kaksi tutkimusmetriikan näkökulmasta irrallista tutkimusjulkaisua toisiinsa tutkimusaineiston kautta. Näin ollen tutkimusaineistojen kautta päästään paremmin perille tutkimustoiminnan verkostoitumisesta esimerkiksi tutkimusaiheen tasolla.

## 4.6. Luokittelu

Metasynteesin lopputuloksena muodostui viisi edellä esitettyä toisistaan huomattavasti eroavaa kokonaisuutta, jotka tutkielman aineistona toimivien selvitysten ja raporttien mukaan tulee ottaa huomioon tutkimusaineistojen tutkimusmetriikassa. Kokonaisuudet eroavat toisistaan merkittävästi nimenomaan niiden tulokulmalla ilmiöön. Tutkimusaineistojen ominaisuudet liittyvät tutkimusaineistojen eroavaisuuksiin muihin tutkimustuotoksiin verrattuna ja käsittelevät tutkimusaineistoja yhtenä tutkimustoiminnan entiteettinä. Julkaiseminen liittyy taas laajemmin koko tutkimusjärjestelmään, jossa tieteellistä tutkimusta tuotetaan ja tähän kytköksissä olevaan tieteellisten tuotosten julkaisemiseen. Tekniset valmiudet käsittelee tutkimusaineistoja niiden teknisten ominaisuuksien ja näiden asettamien vaatimusten kautta. Tutkijan toiminta taas avartaa

yksittäisen tutkijan ja hänen toimintansa vaikutuksia tutkimusaineistojen tutkimusmetriikkaan liittyen, sekä tätä kautta mahdollisesti muodostuviin haasteisiin tutkimusmetriikan näkökulmasta. Arvottaminen taas liittyy keskeisimmin nimenomaan tutkimusmetriikkaan; ovatko tutkimusaineistot muiden tutkimustuotosten vertaisia arvioinnin näkökulmasta?

Analyysin perusteella tehty luokittelu on esitetty Taulukossa 2, josta käy ilmi sekä kokonaisuudet että niiden sisältämät aihepiirit. Tämän lisäksi taulukosta voi havainnoida eri aihepiirien mahdollinen käsittely kussakin tutkimuksen aineiston muodostavista raporteista ja selvityksistä.

Luokittelu: Kokonaisuus / Aihepiiri	Raportti													
	CODATA 2013	DCC 2015	EC 2017	FORCE11 2016	HEFCE 2015a	HEFCE 2015b	JISC 2016	KE 2013a	KE 2013b	MDC 2015a	MDC 2015b	NISO 2016	ROA 2016	
<b>1. Tutkimusaineistojen ominaisuudet</b>														
Rakeisuus	x				x		x						x	
Versiointi	x				x								x	
Roolien määrittely	x							x						
Uudet tutkimusaineistomuodot	x													
Viittauksen merkitys	x	x			x	x				x				
Viittausten suunnat	x	x			x									
<b>2. Julkaiseminen</b>														
Erilaiset viittauskäytännöt	x	x	x		x		x		x				x	
Eroavat julkaisumallit	x	x			x		x							
Vanhojen julkaisutapojen huomiointi	x				x		x							
Hajaantuminen eri alustoille					x				x	x				
Salassapidettävät tutkimusaineistot	x													
Arvioinnin puute							x	x						
<b>3. Tekniset valmiudet</b>														
Tunnisteiden merkitys	x	x	x	x	x		x	x	x				x	x
Viitatun osan tunnistaminen	x													x
Teknisten järjestelmien toimivuus				x	x	x	x							
Tiedonlouhinnan mahdollisuus														x
<b>4. Tutkijan toiminta</b>														
Ei mahdollista meritoitumista	x						x	x	x					
Vaatii lisäresursseja									x					
Tutkimusalojen eroavat julkaisukäytännöt	x				x	x		x	x				x	
Aineistonhallinnan merkitys					x									
Julkaisematta jättäminen epätieteellistä		x						x						
Ilmiöön liittyvät epäluulot	x	x							x					
<b>5. Arvottaminen</b>														
Nykyisen tutkimusmetriikan hyödyntäminen	x	x					x	x	x	x	x	x	x	x
Suhde muihin tutkimustuotoksiin				x	x				x					
Indikaattoreiden valinta	x				x	x		x	x	x	x	x	x	x
Läpinäkyvyys indikaattoreissa	x	x			x									x
Tutkimuksen kokonais kuvan hahmottaminen	x		x											

Taulukko 2: Aineiston luokittelu ja mainintojen jakautuminen raporteilla. Raportit joissa kyseistä ominaisuutta käsitellään on merkitty taulukkoon merkillä "x".

Kokonaisuuksien sisällä eri raporteissa kiinnitettiin vaihtelevasti huomiota eri aihepiireihin, mutta harvassa raportissa huomioitiin kaikkia kunkin kokonaisuuden aihepiirejä. Toisaalta yksikään raportti ei myöskään huomioinut laajasti aihepiirejä kustakin kokonaisuudesta. Luokittelun myötä voidaankin huomata, että metasynteesin tarjoama horisontaalinen näkökulma ilmiöön on tarpeen, sillä eri

asiantuntijaorganisaatioissa ja -työryhmissä keskitytään toisinaan hyvinkin kapeaan osaan tutkimusmetriikan ja sen soveltamisen osa-alueista.

Luokittelun perusteella voidaan myös todeta, että raportit ja selvitykset eivät olleet läheskään yhtenäisiä tutkimusaineistojen tutkimusmetriikan käsittelyä koskien, vaan eroja sekä käsiteltyjen aihepiirien määrässä että niiden sisällössä oli huomattavasti. Osassa raporteista kiinnitettiin huomiota vain muutamiin aihepiireihin, ja ainoastaan viidessä selvityksessä käsiteltiin jollain tasolla kuhunkin aihepiiriin liittyviä käsitteitä ja ilmiöitä. Tämä eittämättä liittyy paitsi selvitysten laajuuteen, myös siihen, millainen taustaorganisaatiot selvitystä oli ollut laatimassa ja millaisia tahoja valmistelussa oli ollut mukana. Erityisesti CODATA:n ja JISC:n raportit (CODATA 2013; JISC 2016) keskittyivät tutkimusaineistojen ominaisuuksiin ja julkaisemiseen, kun taas esimerkiksi HEFCE:n molemmat raportit (HEFCE 2015a, 2015b) huomioivat tutkimusaineistoja pitkälti vain niiden arvottamisen ja indikaattoreiden kautta. Osasyynä tähän on varmasti se, että raporttien ja selvitysten taustalla on vaihtelevasti sekä tutkimusinfrastruktuurista vastaavia toimijoita, että esimerkiksi tutkimuksen rahoittajien edustajia. Näin ollen ryhmän loppuraportti tai -selvitys voi olla laadittu vastaamaan näiden eri toimijoiden tarpeisiin tutkimusaineistojen tutkimusmetriikkaa koskien.

## 5. Yhteenveto

Tutkielman analyysin pohjalta muodostuu selkeitä kokonaisuuksia ja aihepiirejä, joiden kautta tutkimusaineistojen tutkimusmetriikkaa voidaan tarkastella sekä pohtia tutkimusaineistoihin liittyviä ilmiöitä ja niiden seurauksia tutkimusmetriikan näkökulmasta. Tämä luvun ensimmäisessä alaluvussa esitän tutkielman analyysin yhteenvetona ja vastaan tutkimuskysymykseen. Käyn myös läpi tuloksellisuuden mittaamisen dysfunktioita, jotka ovat läsnä tutkimusaineistojen tutkimusmetriikassa ja pohdin niiden todennäköisyyttä. Toisessa alaluvussa esitän tutkielman pohjalta tehdyt johtopäätökset ja tarkastelen mahdollisia jatkotutkimuksen suuntia aihepiiriin liittyen.

### 5.1. Tulosten tarkastelu ja tutkimusmetriikan dysfunktiot

Tämän tutkielman tutkimuskysymykseen “Mitkä tutkimusaineistojen ominaisuudet tai piirteet vaikuttavat tutkimusaineistojen tutkimusmetriikkaan?” vastauksen muodostavat metasynteesin pohjalta havaitut viisi eri kokonaisuutta ja niiden sisältämät eri aihepiirit, joiden kautta tutkimusaineistojen tutkimusmetriikkaa voidaan lähestyä ja vaikutuksia tarkastella.

Ensimmäisen kokonaisuuden muodostavat tutkimusaineistot itsessään ja niihin liittyvät, muista tutkimustuotoksista huomattavasti eroavat, ominaisuudet. Suurin osa tämän kokonaisuuden aihepiireistä liittyvät nimenomaan tutkimusaineistojen muotoon digitaalisena entiteettinä, jolla on tutkimusjulkaisusta poikkeavia ominaisuuksia esimerkiksi rakeisuuden ja tiedon muodon kautta. Koska tämän kokonaisuuden aihepiirejä ei ole tarvinnut ratkaista tutkimusjulkaisujen ollessa tutkimusmetriikan kohteina, ovat ne tutkimusaineistoja koskevan tutkimusmetriikan näkökulmasta selvitetävä ennen laajamittaisempaa metriikan hyödyntämistä.

Toinen kokonaisuus koskee tutkimusaineistojen julkaisemista ja siinä etenkin erilaista julkaisutavoista ja -malleista kumpuavia haasteita tutkimusmetriikan näkökulmasta. Myös näissä on kyse pitkälti tutkimusaineistojen ominaisuuksista, jotka tuovat tutkimusaineistojen julkaisemiseen uusia mahdollisuuksia ja ulottuvuuksia. Toisaalta julkaisemisessa ei ole vielä täysin ratkaistu myöskään tutkimusaineistojen arviointia, eli sitä millaisen arviointiprosessin tutkimusaineisto tarvitsee ollakseen osa muita tieteellisiä tutkimustuotoksia.

Tutkimusaineistojen ja niitä varten perustettujen palveluiden ja tallentamisen teknisten ratkaisujen kehitys ei myöskään ole siinä vaiheessa, että laajamittainen tutkimusmetriikka olisi vielä mahdollista. Tämän kolmannen kokonaisuuden merkittävien aihepiiri on tunnisteet ja niiden merkitys. Vaikka myös tutkimusjulkaisuille tunnisteiden käyttö on merkittävä kehitysaskel kohti luotettavampaa tunnistusta, ovat tunnisteet kuitenkin ensisijaisen tärkeitä tutkimusaineistojen tapauksessa, jossa vasta hahmotellaan soveltuvia viittauskäytäntöjä ja malleja.

Neljäs merkittävä kokonaisuus on tutkimuksesta vastaavan tahon, tutkijan, toiminta liittyen tutkimusaineistoihin. Vaikka tämän kokonaisuuden aihepiirit, esimerkiksi meritoituminen, resursointi sekä epäluulot, eivät suoraan liity tutkimusmetriikkaan, viitattiin niihin kuitenkin useassa raportissa ja selvityksessä käsiteltäessä tutkimusmetriikkaa. Nämä aihepiirit ovat keskeisessä roolissa siinä mielessä, että ne osaltaan vaikuttavat tutkimusaineistojen asemaan osana muita tutkimustuotoksia. Voidaan myös nähdä, että mikäli tutkijoiden käytännöt tutkimusaineistojen kohdalla eivät merkittävästi muutu, on tutkimusmetriikan hyödyntäminen tällöin erittäin haasteellista ja analyysiin vaikuttavat huomattavasti erilaiset käytännöt kansainvälisellä, kansallisella ja eri tutkimusorganisaatioiden tasolla.

Analyysin pohjalta tärkeimmäksi ja käsitellyimmäksi kokonaisuudeksi muodostui kuitenkin tutkimusaineistojen arvottaminen, eli se kuinka tutkimusaineistot nähdään osana muita tutkimustuotoksia ja millaisia indikaattoreita niiden arvioinnissa tulisi käyttää. Raporteissa ja selvityksissä pyrittiin monissa kohdin rakentamaan tutkimusaineistojen tutkimusmetriikkaa nykyisten, usein tutkimusjulkaisujen tarkastelusta periytyvien, indikaattoreiden ja analyysien mukaisesti. Näistä bibliometriikasta ja altmetriikasta peräisin olevien sovellutusten hyödyntäminen ei kaikilta osin kuitenkaan toimi tutkimusaineistojen tapauksissa ja indikaattoreiden valinnassa oli erilaisia näkemyksiä aineiston eri raporteissa. Yhtenevää näkemystä siitä, minkä arvoisia tutkimusaineistot tutkimustuotoksena ovat tai mitä indikaattoreita tulisi käyttää, ei kuitenkaan aineiston pohjalta muodostunut.

Päällimmäinen löydös analyysin pohjalta on kuitenkin se, että tutkimusaineistoja koskeva tutkimusmetriikka on vielä varhaisessa vaiheessa muodostaakseen merkittävä osaa tutkimuksen arvioinneissa. Koska selvityksissä ja raporteissa kiinnitettiin huomioita varsin perustaviin kysymyksiin tutkimusaineistoihin liittyen esimerkiksi niiden

arvottamisen näkökulmasta, voidaan tulkita, että valmiuksia tutkimusmetriikan laajamittaiseen hyödyntämiseen ei vielä ole olemassa.

On huomioitavaa, että eri selvityksissä keskityttiin melko vähäisissä määrin eri tieteenalojen välisiin eroihin. Vaikka eroa eri tieteenaloilla todettiin olevan, nähtiin eri aihepiirien olevan siinä mielessä yleistettävissä, että niitä harvoin pohdittiin tieteenalakohtaisesti. Tämä eroaa jossakin mielessä siitä, että etenkin tutkimusjulkaisujen tapauksessa käsitellään usein nimenomaan eroja julkaisemiskäyttäytymisessä eri aloilla ja niiden muodostamien käytänteiden kautta. Toki käsittelyyn voi vaikuttaa se, että tutkimusaineistojen julkaiseminen on yhä ylipäänsä harvintaista ja käytännöt vaihtelevat tieteenalojenkin sisällä huomattavasti, jolloin eri tieteenalojen välisiin eroihin on vielä vaikeaa ottaa kantaa.

Jos tuloksia sekä luokittelua katsotaan osana tutkimusjärjestelmää, huomataan että kokonaisuuksissa keskitytään vaihtelevasti eri tason toimijoihin. Tutkimusaineistojen julkaiseminen sekä tutkijan toiminta ovat sidoksissa nimenomaan tutkijoihin, eli tutkimusta toteuttavaan tasoon. Sen sijaan tutkimusaineistojen ominaisuuksista sekä teknisistä valmiuksista vastaavat pääasiassa tutkimusinfrastruktuurista huolehtivat tasot sekä muut toimijat, jotka tarjoavat tutkimusaineistopalveluita. Tällaisia ovat esimerkiksi eri tutkimusorganisaatiot. Merkittävin kokonaisuus, arvottaminen, asettuu kuitenkin koko tutkimusjärjestelmää ohjaavalle tasolle sekä tutkimusta rahoittaville toimijoille. Tutkimusaineistojen kohdalla on tarpeen määritellä niiden arvo suhteessa muihin tutkimustuotoksiin, sillä tämä väistämättä vaikuttaa myös tutkimusmetriikan soveltamiseen. Arvottaminen on myös vahvasti sidoksissa siihen, mihin suuntaan indikaattoreita tai muita tutkimusmetriikan soveltamistapoja tullaan viemään, kun tutkimusmetriikan kohteena ovat nimenomaan tutkimusaineistot.

Tämän tutkielman analyysin perusteella tutkimusaineistojen tutkimusmetriikkaan liittyy monia mahdollisia mittaamisen dysfunktioita, joissa tutkimusmetriikkaa joko ohjaa tai kaventaa tutkimustoimintaa ja sen eri ulottuvuuksia, kun tutkimusmetriikan kohteina ovat tutkimusaineistot.

Todennäköisimpiä tutkimusaineistojen tutkimusmetriikan dysfunktioita ovat seitsemästä eri dysfunktiosta nimenomaan tunnelinäkymys sekä osaoptimointi. Tunnelinäkömyksen todennäköisyyttä lisää se, että tutkimusmetriikassa tukeudutaan liiaksi tutkimusjulkaisusta perittyyn arviointimalliin. Tämä on ilmeistä esimerkiksi siinä, että tutkimusaineistojen erilaisia viittauskäytäntöjä tai tutkimusaineistoihin liittyviä vaihtuvia



tekijyyden rooleja ei päästä näin huomioimaan. Myös pelkästään tutkimusaineistojen ominaisuudet eroavat esimerkiksi versioinnin ja rakeisuuden näkökulmasta tutkimusjulkaisujen ominaisuuksista, jolloin tuloksellisuuden mittaamisessa huomio saattaa helposti kiinnittyä tutkimusaineistojen näkökulmasta epäolennaisiin seikkoihin. Osoptimoinnin vaaraa tutkimusaineistojen tapauksessa lisää se, että toistaiseksi tutkimusjärjestelmän tasolla ei ole olemassa kunnollisia mittareita tai indikaattoreita tutkimusaineistojen huomiointiin. Kun tällainen mittari tai indikaattori jollakin aikataululla otetaan käyttöön, saattaa se ohjata tutkijoita julkaisemaan tutkimusaineistonsa tuon mittarin tai indikaattorin näkökulmasta toteutettuun järjestelmään tai palveluun. Tällaisen osoptimoinnin seurauksena voi olla tutkimustoiminnan ja erityisesti tutkimusaineistojen julkaisemisen keskittyminen siihen suuntaan, johon tutkimusmetriikka on sitä ohjannut.

Mahdollisia, joskaan ei niin todennäköisiä, tutkimusmetriikan dysfunktioita tutkimusaineistojen tapauksessa ovat lyhytnäköisyys, konvergenssi, tylsistyminen sekä pelaaminen. Lyhytnäköisyyden mahdollisuus on väistämättä läsnä myös tutkimusaineistojen tutkimusmetriikassa, sillä monet tutkimusaineistoille saatavilla olevista palvelut sekä tallentamiseen että julkaisemiseen ovat uusia eikä niiden jatkuvuudesta tai tulevasta kehityssuunnasta ole varmuutta. Näin ollen tutkimusmetriikassa on vaarana perustaa tuloksellisuuden mittaamista sellaisiin indikaattoreihin tai mittareihin, joita ei nopean kehityksen myötä tulevaisuudessa ole enää viisasta tai edes mahdollista hyödyntää.

Konvergenssin mahdollisuus liittyy osin edellä esitettyyn tilanteeseen osoptimoinnin näkökulmasta, jossa tutkijat ohjaavat toimintaansa asetettujen indikaattoreiden tai mittareiden suuntaan. Konvergenssin riskiä lisää kuitenkin erityisesti se, että tutkimusaineistojen tutkimusmetriikassa saatetaan ensi vaiheessa keskittyä joihinkin eri tieteenaloilla hyväksi koettuihin käytäntöihin tutkimusaineistoihin liittyen. Tällöin vaarana on se, että tutkimusaineistojen tutkimusmetriikka voidaan soveltaa näiden tiettyjen tieteenalojen kohdalla muita tieteenaloja luotettavammin. Konvergenssi

Tylsistymisen dysfunktioita saattaa esiintyä myös tutkimusaineistojen tutkimusmetriikassa. Eräänä analyysin aihepiirinä olivat nimenomaan uudet tutkimusaineistojen muodot, esimerkiksi sosiaalisen median suuret aineistot tai semanttiset aineistot, joiden tarkastelu tutkimusmetriikan näkökulmasta tulee olemaan haastavaa. Mikäli tutkimusmetriikka ei kuitenkaan kykene huomioimaan tutkimusaineistojen kohdalla myös innovatiivisempia tai huomattavasti tavallisesta

poikkeavia aineistoja, on mahdollista että tutkimuksen tuloksellisuuden mittaamisesta seuraa tylsistymistä.

Myös pelaaminen on syytä huomioida eräänä tutkimusaineistojen tutkimusmetriikan riskeistä, sillä se on osin läsnä myös muiden tutkimusjulkaisujen tutkimusmetriikan kohdalla. Tutkimusaineistojen tapauksessa mittaamisessa on syytä huomioida pelaamisen kannalta esimerkiksi se että tutkimusaineistoja, hieman tieteenalasta riippuen, on mahdollista julkaista huomattavasti tiheämpään ja pienemmissä osissa kuin tutkimusjulkaisuja. Tieteen ja yleisen hyvän näkökulmasta olisi kuitenkin hyödyllisempää, mikäli tutkimusaineistojen kohdalla ei ilmenisi tällaista pienimmän mahdollisen julkaisuyksikön tilannetta.

Epätodennäköistä tutkimusaineistojen tutkimusmetriikan kohdalla on luovan laskentatoimen dysfunktio. Mahdolliset raportoinnin tarkoituksenmukaiset vääristelyt tutkimusaineistojen, kuten myös tutkimusjulkaisujen, kohdalla tulisi olla suhteellisen helppo havaita tutkimusjärjestelmän eri tasoilla. Analyysin pohjana olleiden selvitysten ja raporttien mukaan on kuitenkin tärkeää ohjata tutkimusmetriikkaa kohti mahdollisimman läpinäkyviä indikaattoreita ja mittareita, jotta mahdolliset arvioinnin epäkohdat on mahdollista tarkistaa niin tiedeyhteisön, kuin muidenkin tutkimuksen toimijoiden taholta.

## 5.2. Johtopäätökset

Tutkimusmetriikan osalta tutkimusaineistot ovat vielä liian varhaisessa vaiheessa esimerkiksi julkaisemisen, teknisten valmiuksien ja tutkijan toiminnan näkökulmista. Lisäksi tutkimusaineistot itsessään eroavat merkittävältä osin ominaisuuksiltaan muista tutkimustuotoksista.

Merkittävimpänä ratkaisemattomana asiana tutkimusaineistojen tutkimusmetriikan näkökulmasta on kuitenkin se, miten tutkimusaineistot arvioidaan osana muuta tutkimustoimintaa. Tämän tutkielman aineistona toimineiden raporttien ja selvitysten pohjalta vaikuttaa siltä, että julkaisemisen, teknisten valmiuksien ja tutkijan toiminnan kokonaisuudet tulee olla ensin ratkaistu kaikilla tutkimusjärjestelmien eri tasoilla, ennen kuin yhtenäistä näkemystä tutkimusaineistojen arvottamisesta tai valituista indikaattoreista voidaan tehdä. Vaikka useissa raporteissa arvottamiseen liittyviä asioita on jo alustavasti hahmoteltu, tiedostetaan kuitenkin ajallinen haaste niiden toteuttamisessa. Ei voida aidosti asettaa indikaattoreita tai arvottaa tutkimusaineistoja,

jos ei vielä tiedetä, mitä oikeasti mitataan tai mitä edes tulisi mitata. Eräästä raportista vapaasti käännetty toteamus siitä, että “ei kannata muodostaa karttaa, jos ei tunne maastoa” pätee varsin hyvin tutkimusaineistojen tutkimusmetriikkaan.

Tutkimusaineistojen tutkimusmetriikan kehityksessä on myös syytä huomioida tuloksellisuuden mittaamisen dysfunktioista erityisesti tunnelinäkemys ja osaoptimointi, joiden todennäköisyys tutkimusaineistojen tapauksessa on verrattain suuri. Näiden kahden dysfunktion riskiä on mahdollista keskeisesti vähentää sillä, että tutkimusaineistoille ei oteta suoraan käyttöön nykyistä, pääosin tutkimusjulkaisujen näkökulmasta rakennettua, tutkimusmetriikkaa ja sen käytäntöjä. On ensin syytä tarkoin pohtia ja tarkastella niitä ominaisuuksia ja lähtökohtia, joita tutkimusaineistoilla eräänä tutkimustuotoksena on, ja vasta tämän jälkeen siirtyä indikaattoreiden ja mittareiden laatimiseen.

Tutkimusaineistojen nykytila muistuttaa jossain määrin tutkimusjulkaisujen vastaavaa tilannetta 2000-luvun taitteessa, jolloin digitaalisesti tutkimusjulkaisuista saatavilla olevan tiedon määrä ja kattavuus olivat vahvassa kasvussa. Kansallisesti tai kansainvälisesti ei kuitenkaan oltu vielä asetettu arvottamisen kannalta keskeisiä mittareita tai indikaattoreita tutkimusjulkaisuille, ja esimerkiksi rahoituksen suhde tutkimusjulkaisuihin oli vielä vaillinaisen tai perustui lähinnä tutkimusjulkaisujen määrään. Tällöin myös tutkimuksen kokonaisarvioinneissa ei osattu hyödyntää kaikkea sitä tietoa, jota tutkimusjulkaisuista oli saatavilla ja kansainväliset vertailut ja ranking -listat olivat vasta muotoutumassa nykyiseen, tutkimusjulkaisuja merkittävästi hyödyntävään, muotoonsa.

Aihepiiriin liittyen jatkotutkimukselle on huomattava tilaus. Jo tämän tutkielman tuloksena muodostuneiden kokonaisuuksien tarkastelu tarkemmin esimerkiksi tapaustutkimusten kautta olisi tarpeen, jotta eri aihepiirien sisältämiä käytäntöjä voitaisiin paremmin huomioida osana tutkimusmetriikkaa. Merkittävää ja hyödyllistä tietoa tutkimusmetriikan näkökulmasta voitaisiin saada perehtymällä tarkemmin julkaisemisen käytäntöihin tutkimustoiminnassa nyt ja mahdollisesti tulevaisuudessa, jos ja kun eri käytännöt muodostuvat vakiintuneeksi osaksi tutkijoiden työtä. Jatkotutkimuksen näkökulmasta olisi tarpeen selvittää myös nykytilaa sen suhteen miten tutkimusaineistoja huomioidaan tällä hetkellä tutkimustuotoksena erilaisissa tutkimuksen arvioinneissa ja onko tässä tapahtunut kehitystä viime vuosina, kun esimerkiksi tutkimusorganisaatiot ovat tehneet linjauksia tutkimusaineistojen julkaisemisen suhteen.

Jatkotutkimuksen näkökulmasta on myös syytä huomata, että tutkimusaineistot yhtenä tutkimustuotoksen muodoista kattaa koko tutkimusjärjestelmän aina yksittäisestä tutkijasta kansallisiin tutkimusta ohjaaviin tasoihin ja kansainvälisiin tutkimuksen rahoittajiin. Näin ollen tutkimusaineistojen tarkastelu osana tutkimuksen arviointia jokaisella näistä tutkimusjärjestelmän tasoista olisi tarpeen, sillä tutkimuskirjallisuutta aiheeseen liittyen ilmiön tuoreudesta johtuen ei juuri ole.

# Lähteet

- Aagaard K. (2015). *How incentives trickle down: Local use of a national bibliometric indicator system*, Science and Public Policy, Volume 42, Issue 5, Pages 725–737. <https://doi.org/10.1093/scipol/scu087>
- Aalto-yliopisto (2009). *Striving for Excellence - Aalto RAE 2009 and Bibliometric Analysis 2003-2007 -raportti*. <http://arts.aalto.fi/fi/research/researchevaluation/>
- Andrés, A. (2009). *Measuring academic research: How to undertake a bibliometric study*. Oxford: Chandos Publishing.
- Auranen, O., & Nuutinen, A. (2016). *Bibliometriikka on hyvä renki mutta huono isäntä*. Tieteessä Tapahtuu, 34(3). <https://journal.fi/tt/article/view/57061>
- Auranen, O., & Nieminen, M. (2010). *University research funding and publication performance: An international comparison*. Research Policy, 39(6), 834;822;-834. 10.1016/j.respol.2010.03.003
- Avoin tiede ja tutkimus -hanke (2017). *Avoin tiede ja tutkimus -hanke 2014–2017 -Toiminta*. <https://avointiede.fi/toiminta> haettu 27.11.2017.
- Avoin tiede ja tutkimus -hanke (2015a). *ATT-vaikuttavuusselvitysryhmän raportti*. <http://urn.fi/URN:NBN:fi-fe2016122731715>
- Avoin tiede ja tutkimus -hanke (2016). *Evaluation of Openness in the Activities of Research Organisations and Research Funding Organisations in 2016*. <http://urn.fi/URN:NBN:fi-fe2016111829246>
- Avoin tiede ja tutkimus -hanke (2015b). *Avoimen tieteen käsikirja tutkijoille ja tutkimusorganisaatioille*. Haettu 4.1.2018. <https://avointiede.fi/www-kasikirja>
- Ball, A., & Duke, M. (2015). 'How to Track the Impact of Research Data with Metrics'. DCC How-to Guides. Edinburgh: Digital Curation Centre. <http://www.dcc.ac.uk/resources/how-guides>
- Becher, T. (1989). *Academic tribes and territories: Intellectual enquiry and the cultures of disciplines*. Milton Keynes: Open University Press.
- Becher, T., & Trowler, P. R. (2001). *Academic tribes and territories: Intellectual enquiry and the cultures of disciplines* (2nd ed.). Ballmoor: Society for Research into Higher Education.
- Beck C. (2002). *Mothering multiples: a meta- synthesis of qualitative research*. Maternal and Child Nursing 27(4), 214–221.
- Bilder, G., Lin, J., Neylon, C., (2015). *Principles for Open Scholarly Infrastructure*, <http://dx.doi.org/10.6084/m9.figshare.1314859>
- Clemmens D. (2003) *Adolescent motherhood: a meta- synthesis of qualitative studies*. American Journal of Maternal and Child Nursing 28(2), 93–99.

Costas, R., Meijer, I., Zahedi, Z. and Wouters, P. (2013). *The Value of Research Data - Metrics for datasets from a cultural and technical point of view*. A Knowledge Exchange Report, [www.knowledge-exchange.info/datametrics](http://www.knowledge-exchange.info/datametrics)

Data-asiain kansalliskomitea (2017). *Tracing data. First Draft for a Data Citation Roadmap for Finland*. <http://urn.fi/URN:NBN:fi-fe201709198656>

van Drooge, L., & Deuten, J. (2016). *Joint evaluation for joint governance of challenge-oriented research*. <http://hdl.handle.net/20.500.11755/55b2a48a-4205-4681-bc6e-6ca8f5093554>

van Drooge, L., Leonie, Jong, d., Stefan, Faber, M., & Westerheijden, D. F. (2013). *Twenty years of research evaluation*. Rathenau Instituut. <http://hdl.handle.net/20.500.11755/cc9ef717-0aba-4563-9f70-a4f20960955e>

Garfield, E. (2009). *From the science of science to scientometrics visualizing the history of science with HistCite software*. Journal of Informetrics, 3(3), 173-179. 10.1016/j.joi.2009.03.009

Glänzel, W. (2003). *Bibliometrics as a research field : a course on theory and application of bibliometric indicators*

Glänzel, W. (2001). *National characteristics in international scientific co-authorship relations*. Scientometrics, 51(1), 69-115. 10.1023/A:1010512628145

Ellegaard, O., & Wallin, J. A. (2015). *The bibliometric analysis of scholarly production: How great is the impact?* Scientometrics, 105(3), 1809-1831. 10.1007/s11192-015-1645-z

Englund, J., & Forsman, M. (2013). *Altmetriikka: Bibliometriikan uusi suuntaus*. Signum, 2013(6), p. 13.

Euroopan komissio (2012). *Commission recommendation on access to and preservation of scientific information*. C(2012) 4890 final. [http://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/recommendation-access-and-preservation-scientific-information\\_en.pdf](http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf)

Euroopan Komissio (2011). *Digital agenda: Turning government data into gold*. [Press release]. [http://europa.eu/rapid/press-release\\_IP-11-1524\\_en.htm](http://europa.eu/rapid/press-release_IP-11-1524_en.htm)

Euroopan komissio (2016). *Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020*. Version 3.2. [http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination\\_en.htm](http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm)

Euroopan komissio (2010). *Riding the wave: How Europe can gain from the rising tide of scientific data*. Final report of the High level Expert Group on Scientific Data A submission to the European Commission [http://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=707](http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=707)

Forsman, M. (2016). *Julkaisut ja tieteen mittaaminen: Bibliometriikan käännekohtia*. Tampere: Enostone.

- Halevi, G. and Colledge, L. (2014). *Standardizing research metrics and indicators- perspectives and approaches*. Research Trends. 39. [www.researchtrends.com/issue-39-december-2014/standardizing-research-metrics-and-indicators/](http://www.researchtrends.com/issue-39-december-2014/standardizing-research-metrics-and-indicators/). Haettu 13.9.2017.
- Helsingin Sanomat (2018a). 17.1.2018. Mieli-pide: *Professori Hannu Heikkinen Vieraskynä-kirjoituksessa: Tiedepolitiikka näivettää suomen kielen*. Hannu L.T. Heikkinen.
- Helsingin Sanomat (2018b). 29.1.2018. Mieli-pide: *Julkaisufoorumi kannustaa tutkijoita rinnakkaiskielisyteen*. Janne Pölönen
- Helsingin Sanomat (2018c). 26.2.2018. Mieli-pide: *"Mielivaltaista peliä" ja "suorastaan rikollista toimintaa" – professorivalinnat vaikuttavat pienen piirin suhmuroinnilta, syyttää yliopistoväki*. Vesa Harmaakorpi.
- Helsingin Sanomat (2018d). 1.3.2018. Mieli-pide: *Väite Julkaisu-foorumin korruptiosta on perätön*.
- Helsingin yliopisto (2012). *International Evaluation of Research and Doctoral Training at the University of Helsinki 2005–2010*. University of Helsinki, Administrative Publications 81. <http://www.helsinki.fi/arviointi2010-2012/>
- Helsingin yliopisto (2015). *Helsingin yliopiston tutkimusdatapolitiikka*. <http://www.helsinki.fi/kirjasto/fi/avuksi/tutkimusdatan-hallinta/tutkimusdatapolitiikka/>
- Hicks, D., ym. (2015). *Bibliometrics: The leiden manifesto for research metrics*. Nature, 520(7548), 429-431. <http://dx.doi.org/10.1038/520429a>
- Hicks, D. (2012). *Performance-based university research funding systems*. Research Policy, 41(2), 251. 10.1016/j.respol.2011.09.007
- Ilva, J. (2014). *Julkaisuarkistot ja tutkimustietojärjestelmät*, Johdantoesitys Kirjastoverkkopäivien työpajaan 23.10.2014. <http://urn.fi/URN:NBN:fi-fe2014102445601>
- Jyväskylän yliopisto (2011). *Evaluation of Research Activities 2005-2009*. <https://www.jyu.fi/yliopistopalvelut/research-and-innovation/research-assessment-exercise>
- Karvonen, E., Kortelainen, T., & Saarti, J. (2014). *Julkaise tai tuhoutu: Johdatus tieteelliseen viestintään*. Tampere: Vastapaino.
- Knowledge Exchange (2013). *Making data count - Research Data and Research Assessment*. Knowledge Exchange.
- Koski L. (2002). *Yliopiston sisäinen toimintalogiikka 1900-luvun lopun muutoksissa*. Teoksessa Ahonen, S. ja Välimaa, J. (toim.), Heimoja, hengenviljelyä ja hallintoa. Korkeakoulutuskimpuksen vuosikirja.Koulutuksen tutkimuslaitos. ss. 33-51.
- Kratz, J. & Strasser, C. (2015). *Making data count*. Scientific Data 2, article number 150039. <http://dx.doi.org/10.1038/sdata.2015.39>
- Kuusniemi, E. & Salmi, A. (2017). *Project MILDRED: Charting ground for research data management services at the University of Helsinki*. JEAHIL Vol. 13 No. 2 June 2017— Open Science 2: Research Data. <http://eahil.eu/wp-content/uploads/2017/06/journal-2-2017-web-1.pdf>

Lee, J. (2010). *10 Year Retrospect on Stage Models of e-Government: A Qualitative Meta-Synthesis*. *Government Information Quarterly* 27: 3, 220–230.

Lilja, J. (2013). *Tieteellinen julkaiseminen eilen, tänään, huomenna*. *Tieteessä Tapahtuu*, 31(2). <https://journal.fi/tt/article/view/7852>

Lähdesmäki, K. (2003). *New Public Management ja julkisen sektorin uudistaminen*. *Acta Wasaensia* No 113, *Hallintotiede* 7. Vaasa: Universitas Wasaensis

Markkanen, M. (2006) *Yliopistojen kolmas tehtävä ja suomen kieli*. *Tieteessä tapahtuu* 3/2006: 37–38)

Muhonen, R.; Eskola, O.; Leino, Y.; Pölönen, J. (2016). *Tutkimuksen monitieteisyys ja laatu*. *Opetus- ja kulttuuriministeriön julkaisuja* 2016:12. Opetus- ja kulttuuriministeriö. <http://urn.fi/URN:ISBN:978-952-263-397-2>

Muhonen, R. & Puuska, H. (2014). *Kansallista tiedettä tekemässä*. Teoksessa *Tutkimuksen kansallinen tehtävä*. Tampere: Vastapaino. pp.11-34

Mustajoki, A. (2014). *Arviointeja, arviointeja, arviointeja: Jättääkö mitaamisen hurmos tilaa luovalle ajattelulle?*. Teoksessa *Tutkimuksen kansallinen tehtävä*. Tampere: Vastapaino. pp.11-34

Mustajoki, A. (2010). *Yliopistojen rankingit – paljon melua tyhjästä?*. *Tieteessä tapahtuu*, [S.I.], v. 28, n. 8, jouluku. 2010. ISSN 1239-6540. <https://journal.fi/tt/article/view/3795>

Neylon, C. (2016). *Jisc Briefing Document on Data Citations*. <http://repository.jisc.ac.uk/id/eprint/6399>

Neylon, C. ed. (2017) *Knowledge Exchange Approach towards Open Scholarship*. <http://dx.doi.org/10.5281/zenodo.826643>

National Information Standards Organization (2016). *NISO recommended practice: Outputs of the alternative assessment metrics project*.

Nykyri, S. (2017). *Research Data Metrics – Where are we now at the University of Helsinki?* (Version 1). figshare. <https://doi.org/10.6084/m9.figshare.5624776.v1>

Opetus- ja kulttuuriministeriö (2015). *Yliopistojen rahoitusmalli 2017*. Julkaistu artikkelin “Yliopistojen rahoitusmalli uudistuu” yhteydessä. [http://minedu.fi/artikkeli/-/asset\\_publisher/yliopistojen-rahoitusmalli-uudistuu](http://minedu.fi/artikkeli/-/asset_publisher/yliopistojen-rahoitusmalli-uudistuu)

Opetus- ja kulttuuriministeriö (2014). *Tutkimuksen avoimuudella yllättäviä löytöjä ja luovaa oivaltamista: Avoimen tieteen ja tutkimuksen tiekartta 2014–2017*. <http://urn.fi/URN:ISBN:978-952-263-317-0>

Opetus- ja kulttuuriministeriö (2017). *Tutkimustietovaranto avaa tutkimustiedot käyttöön yhdestä paikasta*. [http://minedu.fi/artikkeli/-/asset\\_publisher/tutkimustietovaranto-avaa-tutkimustiedot-kayttoon-yhdesta-paikasta](http://minedu.fi/artikkeli/-/asset_publisher/tutkimustietovaranto-avaa-tutkimustiedot-kayttoon-yhdesta-paikasta)

Opetus- ja kulttuuriministeriö (2013). *WoS vai Scopus? Suomalaisen tutkimuksen tila 2010-luvun alussa kansainvälisten viiteaineistojen mukaan*. Opetus- ja kulttuuriministeriön julkaisuja 2013:17



Oulun yliopisto (2013). *Research Assessment Exercise Report 2014 - International Evaluation of Research at the University of Oulu*.

<http://www oulu.fi/yliopisto/node/35272>

Paterson B., Thorne S. & Dewis M. (1998). *Adapting to and managing diabetes*. *Image Journal of Nursing Scholarship* 30(1), 57–62.

Peters, I. et al. (2016). *Research data explored: An extended analysis of citations and altmetrics*. *Scientometrics*, 107(2), 723-744. <http://dx.doi.org/10.1007/s11192-016-1887-4>

Pielstick C. (1998). *The transforming leader: a meta- ethnography analysis*. *Community College Review* Winter, 1–14.

Piwovar, H. A., & Vision, T. J. (2013). *Data reuse and the open data citation advantage*. *Peerj*, 1, e175. <http://dx.doi.org/10.7717/peerj.175>

Puuska, H. & Miettinen, M. (2008). *Julkaisukäytännöt eri tieteenaloilla*. Helsinki: Opetusministeriö. <http://urn.fi/URN:ISBN:978-952-485-566-2>

Puuska, H. (2014). *Scholarly publishing patterns in Finland: A comparison of disciplinary groups*. Väitöskirja. Tampere: Tampere University Press.

<http://urn.fi/URN:ISBN:978-951-44-9480-2>

Pölönen, J. & Wahlfors, L. (2016). *Bibliometrics and Research Policy, Nordic Workshop* (2016): NWB'2016 Posters. <http://dx.doi.org/10.6084/m9.figshare.c.3581330.v4>

Pölönen, J. & Auranen, O. (2017). *Julkaisupaine suomalaisessa tiedeyhteisössä*. *Informaatiotutkimus*, Vol 36 Nro 2. <http://dx.doi.org/10.23978/inf.65186>

Rauber, A., Asmi, A., van Uytvanck, D., Pröll, S. (2016). *Identification of Reproducible Subsets for Data Citation, Sharing and Re-Use*. *Bulletin of the IEEE Technical Committee on Digital Libraries*, 12:1, 2016

Salmi, A., Ojanen, M., & Kuusniemi, M. E.. (2016). *Project MILDRED Research Data Repository Survey*, University of Helsinki (Version 4). figshare.

<https://doi.org/10.6084/m9.figshare.3806394.v4>

Salminen, A. (2003). *New Public Management and Finnish Public Sector Organisations: The Case of Universities*. In A. Amaral, Meek V.L. & I. M. Larsen (Eds.), *The Higher Education Managerial Revolution?* (pp. 55–69).

Salminen, A. (2011). *Mikä kirjallisuuskatsaus?: Johdatus kirjallisuuskatsauksen tyyppeihin ja hallintotieteellisiin sovelluksiin*. Vaasa: Vaasan yliopisto.

Sandgren, T. (2016). *Kansallinen bibliometriikkaseminaari 2016*. *Informaatiotutkimus*, Vol 36 Nro 1 <https://doi.org/10.23978/inf.63191>

Sherpa Juliet (2017). *Data Archiving Policies*. Haettu 15.10.2017 [http://v2.sherpa.ac.uk/view/funder\\_visualisations/1.html](http://v2.sherpa.ac.uk/view/funder_visualisations/1.html)

Smith, P. (1996). *A Framework for Analysing the Measurement of Outcome*. Teoksessa Smith Peter (toim.) *Measuring outcome in the Public Sector*. 1–19. London; Taylor & Francis.

- Socha, Y. ed. (2013) *Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data*. Data Science Journal. 12. <http://dx.doi.org/10.2481/dsj.OSOM13-043>
- Suomen Akatemia (2017). Syyskuun hakuilmoitus 2017. Haettu 4.1.2018. <http://www.aka.fi/fi/rahoitus/aiemmat-haut/>
- Suomen Akatemia (2016). Tieteen tila 2016. <http://www.aka.fi/fi/tiedepoliittinen-toiminta/tieteen-tila/>
- Suomen Akatemia (2014). Tieteen tila 2014. <http://www.aka.fi/fi/tiedepoliittinen-toiminta/tieteen-tila/tieteen-tila/>
- Suomen yliopistokirjastojen verkosto (2016). *Leidenin tutkimusmetriikkamanifesti*. SYNin tutkimuksen tuen verkoston ohjausryhmä. <http://yliopistokirjastot.fi/verkostot/tutkimuksen-tuen-verkosto/leidenin-manifesti/>
- Strasser, C., Kratz, J., Lin, J. (2015). *Make Data Count - Unit 1 Final Report*. figshare. <https://doi.org/10.6084/m9.figshare.1328291.v4>
- Taavitsainen, I. (2004): *Englanti tieteen kielenä*. Tieteessä tapahtuu 5/2004: 13–18.
- Tammi, T. (2014). *Piittaamattomuus ja vilppi tutkimusjulkaisemisessa*. Yhteiskuntapolitiikka / ISSN 1455-6901. - 79 (2014) : 6, s. 603-604.
- Tampereen teknillinen yliopisto (2017). *Tampere University of Technology Research Assessment Exercise 2017*. <http://www.tut.fi/fi/tutkimus/tutkimuksen-painopisteet/tutkimuksen-arviointi/>
- Teelken, C. (2012). *Compliance or pragmatism: how do academics deal with managerialism in higher education? A comparative study in three countries*. Studies of Higher Education Vol. 37, No. 3, 271–290.
- Tolle, K. M., Tansley, D. S. W., & Hey, A. J. G. (2011). *The fourth paradigm: Data-intensive scientific discovery*. Proceedings of the IEEE, 99(8), 1334-1337. 10.1109/JPROC.2011.2155130
- TUHA-tietomallityöryhmä (2017). *TUHA - sanaston valmiit käsitteet ja määritelmät 6/2017*. Haettu 20.10.2017. <https://confluence.csc.fi/pages/viewpage.action?pageId=64236640>
- Turun yliopisto (2015). *Bibliometric Analysis For The University Of Turku 2010-2013: Part 1 & Part 2*. <http://www.utu.fi/fi/Tutkimus/tutkimuksen-kokonaisarviointi/arviointiaineisto/bibliometrinen-analyysi/>
- Teichler, U. (2003). *The future of higher education and the future of higher education research*, Tertiary Education and Management 9(3), 171–185.
- Tenopir, C. ym. (2011). *Data sharing by scientists: Practices and perceptions*. PloS One, 6(6), e21101. <http://dx.doi.org/10.1371/journal.pone.0021101>
- Yhteiskuntatieteellinen tietoarkisto (2018). *Aineistonhallinnan käsikirja*. Haettu 4.1.2018. <http://www.fsd.uta.fi/aineistonhallinta/fi/>

Ylijoki, O., Lyytinen, A., & Marttila, L. (2011). *Different research markets: A disciplinary perspective*. Higher Education, 62(6), 721-740.  
<http://dx.doi.org/helios.uta.fi/10.1007/s10734-011-9414-2>

Vakkuri, J., & Meklin, P. (2006). *Ambiguity in performance measurement: A theoretical approach to organisational uses of performance measurement*. Financial Accountability & Management, 22(3), 235-250. 10.1111/j.0267-4424.2006.00401.x

Vakkuri, J., & Meklin, P. (2003). *The impact of culture on the use of performance measurement information in the university setting*. Management Decision, 41(8), 751-759. 10.1108/00251740310496260

Vakkuri, J. (1998). *Tehokkuuden rajoilla DEA-menetelmä tulostittauksessa. Esimerkkinä yliopistojen ainelaikokset*. Acta Universitatis Tamperensis 635. Tampere.

Wallin, J. A. (2005). *Bibliometric methods: Pitfalls and possibilities*. Basic & Clinical Pharmacology & Toxicology, 97(5), 261-275. 10.1111/j.1742-7843.2005.pto\_139.x

Walsh, D., & Downe, S. (2005). *Meta- synthesis method for qualitative research: A literature review*. Journal of Advanced Nursing, 50(2), 204-211. 10.1111/j.1365-2648.2005.03380.x

Wilkinson, M. D. et al. (2016). *The FAIR Guiding Principles for scientific data management and stewardship*. Sci. Data 3:160018  
<http://dx.doi.org/10.1038/sdata.2016.18>

Wilsdon, J., ym. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*.  
<http://dx.doi.org/10.13140/RG.2.1.4929.1363>

Wilsdon, J. ym. (2017). *Next-generation metrics: Responsible metrics and evaluation for open science* European Commission. <http://dx.doi.org/10.2777/337729>

Wouters, P. (1999). *The Citation Culture*. Doctoral Thesis, University of Amsterdam.  
<http://garfield.library.upenn.edu/wouters/wouters.pdf>

Wouters, P. ym. (2015). *The Metric Tide: Literature Review (Supplementary Report I to the Independent Review of the Role of Metrics in Research Assessment and Management)*. HEFCE. <http://dx.doi.org/10.13140/RG.2.1.5066.3520>

# Liitteet

## Liite 1 - Aineistosta tutkimusaineistoksi

Tässä tutkielmassa tutkimusaineistoja käsitellään suhteellisen yhtenäisenä joukkona tutkimustuotoksia. Tämä kuitenkin edellyttää sitä, että tutkimusaineistot ovat aineistonhallinnan näkökulmasta hyvin järjestetty sekä tallennettu asianmukaisesti siten, että esimerkiksi viittaaminen niihin on ylipäänsä mahdollista. Näin niiden huomiointi esimerkiksi tutkimusmetriikan näkökulmasta mahdollistuu ja ne on siirrettävissä analyysien pohjana oleviin tietokantoihin. On kuitenkin huomioitava, että tällaisia tutkimusaineistoja on yhä varsin harvassa sillä tutkimusaineiston muodostaminen edellyttää tutkijalta varsin huomattavaa työtä ja käytännöt vaihtelevat huomattavasti eri tieteenaloilla. Tämän liitteen tarkoituksena on lyhyesti kuvailla sitä, mitä vaiheita tutkimusaineiston muodostaminen vaatii, jotta siitä muodostuu tässä tutkielmassa määritelty tutkimusaineisto. Sen laatimisessa on hyödynnetty osin Yhteiskuntatieteellisen tietoarkiston aineistonhallinnan käsikirjaa ja Avoin tiede ja tutkimus -hankkeen Avoimuuden käsikirjaa (Avoin tiede ja tutkimus -hanke 2015b, Yhteiskuntatieteellinen tietoarkisto 2018).

### Aineisto

Kun tutkija ensimmäistä kertaa hahmottelee uudelle tutkimukselleen tutkimuskysymystä, hyödynnetään lähes tieteenalasta riippumatta tutkimuskysymykseen vastatessa erilaisia aineistoja. Näitä aineistoja voivat olla esimerkiksi arkistosta löytyvät sanomalehtiartikkelit, biopankin varastoimat ihmisperäiset biologiset näytteet, ilmaperän happipitoisuutta tarkkailevan mittarin lukemat tai ryhmähaastatteluista kertyneet nauhoitukset. Karkeasti aineistot voidaan jakaa vielä määrällisiin ja laadullisiin aineistoihin. Jotta edellisen kaltaisia aineistoja, joihin voidaan viitata myös raakadatan tai tutkimusdatana, voidaan hyödyntää, on tutkijan kuitenkin ensin analysoitava näitä aineistoja sillä vanhat sanomalehtiartikkelit tai ilmakehän happipitoisuus eivät voi suoraan vastata yhteenkään tutkimuskysymykseen.

Analyysi edellyttää siis aineiston käsittelyä. Tällöin tutkija, tai esimerkiksi laboratoriotyöntekijä, käy aineistoa läpi ja pyrkii siten vastaamaan asettamaansa tutkimuskysymykseen. Tämä voi olla esimerkiksi haastatteluaineiston litterointia ja läpikäyntiä, regressioanalyysin muodostamista rekisteriaineistosta tai genomidatan sekvensointia. Vasta käsittelyn jälkeen aineistosta voidaan tehdä johtopäätöksiä ja mahdollistetaan täten tutkimuskysymykseen vastaaminen. Tämä käsittely tehdään usein tutkijan omalla tietokoneella tai jollakin muulla tutkimusinfrastruktuuriin kuuluvalla laitteella. Usein lopputuloksena on uusi, aikaisemmasta aineistosta täydennetty tai laajennettu, aineisto, johon tutkijan tekemä käsittely on tuonut olennaisesti uutta tietoa. Jos tutkija ei koe tarpeelliseksi käsittelyn jälkeisen aineiston julkaisemista eikä esimerkiksi tutkimuksen rahoittaja vaadi muuta, voidaan aineiston elinkaaren katsoa päättyneen, sillä paikallisesti muodostettuihin aineistoihin ei pääse muut tutkijat käsiksi. On myös aineistoja, joita ei ole merkityksellistä säilyttää tai jatkokäyttää, esimerkiksi niille asetettujen käyttörajoitusten, kustannusten tai käytettävyyden näkökulmasta.

Jotta aineistosta voidaan muodostaa aidosti hyödynnettävä tutkimusaineisto jota muut tutkijat voivat hyödyntää omassa tutkimuksessaan, edellyttää se aineiston valmistelua myös julkaisemista varten. Seuraavassa on esitelty niitä seikkoja, joita tutkijan tulee huomioida, jotta aineisto olisi aidosti jatkokäytettävissä sekä hyödynnettävissä myös muiden tutkijoiden toimesta.

### Kuvailu

Jatkokäytön näkökulmasta on erittäin tärkeää tietää, mitä aineistolle on mahdollisesti tehty, mitä teknisiä ratkaisuja sille on tehty ja kuinka ne vaikuttavat aineiston laatuun. Toisaalta jatkokäytön näkökulmasta on syytä tietää perustietoja muun muassa aineiston ajallisesta rajauksesta, sen tekijöistä tai käytetyistä muuttujista. Tätä kutsutaan aineiston kuvailevaksi tiedoksi (*metadata*), joka on syytä säilyttää osana aineistoa.

### **Oikeudet ja tietosuojaja**

Aineiston julkaisemista voi rajoittaa tai sen voi jopa estää esimerkiksi aineistonkeruun ja käsittelyn toteuttaminen osana yrityksen kehitystyötä. Toisaalta aineistoon voivat vaikuttaa esimerkiksi tekijänoikeuslaki sekä muut sopimukselliset seikat. Myös tietosuojaan tulee kiinnittää huomiota niissä tapauksissa, kun tutkimusaineisto koskee henkilöitä jolloin se saattaa sisältää sensitiivistä aineistoa.

### **Tallennus ja saatavuus**

Jotta aineisto voidaan julkaista, tulee se olla tallennettu sellaiseen paikkaan tai palveluun, josta se on saatavilla digitaalisessa muodossa. Tällaisia tallennuspaikkoja voivat olla perinteisemmät arkistot (esimerkiksi Yhteiskuntatieteellinen tietoarkisto tai UK Data Archive) tai erilaiset tutkimusaineistoille suunnatut verkkopalvelut, joihin voidaan tallentaa tutkimusaineistoja (esimerkiksi Opetus- ja kulttuuriministeriön tarjoama IDA-palvelu tai kansainvälinen Zenodo). Tallennuspaikasta riippuen aineistoilta edellytetään erilaisia kuvailutietoja, mutta tärkeimpänä ne mahdollistavat aineistoon pääsyn myös tutkimuksen ulkopuolisilta tutkijoilta.

Kun aineistoa tallennetaan, tulee huomioida esimerkiksi eri tiedostoformaattit ja ohjelmistot, joiden uusiutumisen sykli on usein nopea. Tallennuksessa pitää siis huomioida se, että aineisto on sellaisessa formaatissa, jossa se on saatavilla myös tulevaisuudessa mahdollisimman suurella todennäköisyydellä. Tallennus saattaa myös paperisten aineistojen tai esimerkiksi ihmisperäisten näytteiden tapauksessa edellyttää digitointia, jossa siitä luodaan digitaalisessa muodossa ymmärrettävää materiaalia.

### **Julkaiseminen**

Aineistosta muodostuu vasta edellä mainittujen vaiheiden kautta tutkimusaineisto, joka on aidosti jatkokäytettävissä tai hyödynnettävissä. Ilman kuvailua julkaistua aineistoa on lähes mahdoton tulkita, sillä ilman sitä voi kyseessä olla vain joukko eri formaatin tiedostoja tai epämääräisiä merkkijonoja tallennettuna johonkin verkkopalveluun. Toisaalta asiallisesti kuvailtua ja tietosuojan näkökulmasta huolehdittua aineistoa ei voi hyödyntää, mikäli sitä ei ole tallennettu siten, että aineistoa voisi tarkastella tai ladata omaan käyttöön. Tutkimusaineistojen julkaisemisessa on toistaiseksi käytössä pääasiassa kaksi eri julkaisumallia, joista toisessa tutkimusaineistoista julkaistaan lyhyt kuvaileva artikkeli osana tutkimusaineistojen julkaisemiseen keskittyntä julkaisusarjaa (*data journal*) ja toisessa julkaisu tapahtuu suoraan jonkin tutkimusaineistoille suuntautuneen arkiston tai palvelun kautta (*data repository*).

## Lite 2 - Aineisto taulukkona

Julkaisun nimi	Vuosi	Organisaatio	Metriikka	Aineisto	Taso	Linkki
Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data	2013	CODATA	datametriikka	tutkimuskirjallisuus, työpajat, työryhmä	empiria, teoria	<a href="https://drive.google.com/file/d/17BeZm2BjSMv-8ZBQ6ldw21A65ol_zqzrv/view?usp=sharing">https://drive.google.com/file/d/17BeZm2BjSMv-8ZBQ6ldw21A65ol_zqzrv/view?usp=sharing</a>
How to Track the Impact of Research Data with Metrics	2015	DCC (JISC)	datametriikka	työryhmä	empiria	<a href="https://drive.google.com/file/d/12Y2Rq9EXtN_e003HRIZTS6qKkm11uUjMc/view?usp=sharing">https://drive.google.com/file/d/12Y2Rq9EXtN_e003HRIZTS6qKkm11uUjMc/view?usp=sharing</a>
Next-generation metrics: Responsible metrics and evaluation for open science	2017	EC	altmetriikka	kyselevaineisto, tutkimuskirjallisuus, työryhmä	empiria, teoria	<a href="https://drive.google.com/file/d/0B91KDW3wvTYRbYvDpKIEWIBvU0E/view?usp=sharing">https://drive.google.com/file/d/0B91KDW3wvTYRbYvDpKIEWIBvU0E/view?usp=sharing</a>
The FAIR Guiding Principles for scientific data management and stewardship	2016	FORCE11	tutkimusmetriikka	työryhmätyö	empiria	<a href="https://drive.google.com/file/d/1MzSxXEWGzC8pys1DqV8UQ-iBxyq03FEk/view?usp=sharing">https://drive.google.com/file/d/1MzSxXEWGzC8pys1DqV8UQ-iBxyq03FEk/view?usp=sharing</a>
The Metric Tide	2015	HEFCE	tutkimusmetriikka	työpajat, työryhmä	empiria	<a href="https://drive.google.com/file/d/0B91KDW3wvTYRbU5xIThniTmdqX3c/view?usp=sharing">https://drive.google.com/file/d/0B91KDW3wvTYRbU5xIThniTmdqX3c/view?usp=sharing</a>
The Metric Tide Literature Review Supplementary Report I	2015	HEFCE	tutkimusmetriikka	tutkimuskirjallisuus	teoria	<a href="https://drive.google.com/file/d/1OYp_W1RlRlUZTTiBLBL0bHPx8t2OOL0WC/view?usp=sharing">https://drive.google.com/file/d/1OYp_W1RlRlUZTTiBLBL0bHPx8t2OOL0WC/view?usp=sharing</a>
Jisc Briefing Document on Data Citations	2016	JISC	datametriikka	tutkimuskirjallisuus	empiria	<a href="https://drive.google.com/file/d/1ML_BMO56k4KqC4XtdIjqWlU1x1CdWzCdC/view?usp=sharing">https://drive.google.com/file/d/1ML_BMO56k4KqC4XtdIjqWlU1x1CdWzCdC/view?usp=sharing</a>
Making data count	2013	KE	datametriikka	työpajat	empiria	<a href="https://drive.google.com/file/d/1Hfn9Kz2HSVWVlCBxR_bJmeRAVlUyM3Pndp/view?usp=sharing">https://drive.google.com/file/d/1Hfn9Kz2HSVWVlCBxR_bJmeRAVlUyM3Pndp/view?usp=sharing</a>
The Value of Research Data	2013	KE	tutkimusmetriikka	kyselevaineisto, tutkimuskirjallisuus	empiria, teoria	<a href="https://drive.google.com/file/d/1-uPe3z93kKaGz0BuazqICw_SYC3H64L/view?usp=sharing">https://drive.google.com/file/d/1-uPe3z93kKaGz0BuazqICw_SYC3H64L/view?usp=sharing</a>
Make Data Count - Unit 1 Final Report	2015	PLoS, DataONE, CDL	datametriikka	kyselevaineisto	empiria	<a href="https://drive.google.com/file/d/1Rd5u0L_BVITD2Q1C0NhtjxSsYOUS91KG3Y/view?usp=sharing">https://drive.google.com/file/d/1Rd5u0L_BVITD2Q1C0NhtjxSsYOUS91KG3Y/view?usp=sharing</a>
Making data count	2015	PLoS, DataONE, CDL	datametriikka	kyselevaineisto	empiria	<a href="https://drive.google.com/file/d/1bWIRZv3UJKuud2Wl_-5c65xeJ3FExs5wU/view?usp=sharing">https://drive.google.com/file/d/1bWIRZv3UJKuud2Wl_-5c65xeJ3FExs5wU/view?usp=sharing</a>
Outputs of the NISO Alternative Assessment Metrics Project	2016	NISO	datametriikka	tutkimuskirjallisuus, työpajat	empiria, teoria	<a href="https://drive.google.com/file/d/1z7Tkk5021AUle9SwDHMMonJBM_LA88ae5h/view?usp=sharing">https://drive.google.com/file/d/1z7Tkk5021AUle9SwDHMMonJBM_LA88ae5h/view?usp=sharing</a>
Identification of Reproducible Subsets for Data Citation, Sharing and Re-Use	2016	RDA	datametriikka	työpajat, työryhmä	empiria	<a href="https://drive.google.com/file/d/1UOXESuY9deZTW6iOZAq_rffkIkw/SLDA6V/view?usp=sharing">https://drive.google.com/file/d/1UOXESuY9deZTW6iOZAq_rffkIkw/SLDA6V/view?usp=sharing</a>