

KOHTI VASTUULLISTA TUTKIMUSKULTTUURIA

JUUSO ALA-KYYNY

Tutkimuksen arvioinnissa välineet ovat kehittyneet nopeammin kuin arvioinnin tekijöiden osaaminen. Helposti ja halvalla tuotettavaa bibliometristä dataa hyödynnetään rekryointi- ja rahoituspäätöksissä usein käyttötarkoituksensa vastaisesti. Bibliometriikan asiantuntijoille ongelma on tuttu. Kirjastot toimivat vastuullisessa metrikassa viestinvälittäjinä, mutta tutkimuskulttuurin muuttamiseen tarvitaan koko tiedeyhteisön tukea.

Tarve tutkimuksen arvioinnin pelisääntöjen määrittelyyn tiedeyhteisössä on lisääntynyt sitä mukaa kuin menetelmät bibliometrisen datan tuottamiseen ovat kehittyneet. Erilaiset verkkotyökalut, kuten vuosituhannen alkupuolella kehitetyt InCites (Web of Science -tietokanta) ja SciVal (Scopus), ovat tehneet julkaisemista koskevan numerotiedon tuottamisesta teknisessä mielessä helppoa.

Yhä helpommin saatava bibliometrinen data on johtanut erilaisten tutkimusta koskevien indikaattorien kehittelyyn ja käyttöön. Tunnetuimpia lienevät jo 1970-luvulla tieteellisten julkaisujen vertailuun luotu vaikuttavuuskerroin (*impact factor*, IF, tai *journal impact factor*, JIF) ja erilaiset tutkijantyötä mittaavat indeksit, kuten tutkijan saamia viittauksia laskeva h-indeksi. Ja kun lukuja on yhä enemmän tarjolla, mittaaminen ei rajoitu vain julkaisuihin, tutkimukseen ja tutkijoihin. Myös tutkimusorganisaatioita vertaillaan, ja erilaisia yliopistorankingeja, kuten Shanghai-ranking ja Times Higher Educationin lista, on noussut 2000-luvulla kuin sieniä sateella.

Samalla kun metriikkaan pohjautuva tutkimuksen arviointi on lisääntynyt, tarve välineiden käytön kriittiseen tarkasteluun on ilmeinen. Tutkimuksen tuen ja hallinnon verkosto (TUHA) järjesti yhdessä Tieteellisten seurain valtuuskunnan (TSV) kanssa Helsingin yliopistossa elokuun lopussa vastuullisen

metriikan kick-off-tilaisuuden. Sen tarkoituksena oli avata keskustelua metriikan vastuullisuudesta tutkimusorganisaatioissa ja yhteisten metriikan käytön pelisääntöjen tarpeesta Suomessa. Tässä katsausartikkelissa käydään läpi tilaisuudessa esiin nostettuja teemoja alustuspuheenvuorojen, tutkija-haastattelun (Niilo Kauppi) ja kirjallisuuden kautta.

Mitä ongelmia mittareiden käytössä on?

Tutkimuksen arvioinnin kritiikissä nostetaan toistuvasti esiin tietyt teemat. Eräs keskeisimmistä ongelmista liittyy metriikkatyökalujen helppokäyttöisyyteen. Koska dataa on suhteellisen helposti ja halvalla saatavilla, käyttökynnys on matala. Tällöin metriikkadataa käyttävät sellaisetkin tahot, joilla ei ole lukujen tulkitsemiseen riittävää osaamista. Jos bibliometristä dataa päätöksissään hyödyntävät tahot eivät tunne datan tai indikaattorien luonnetta, rajoituksia tai hyviä käytäntöjä, päätöksiä saatetaan tehdä väärin perustein rekrytoinneissa tai tutkimuksen rahoituksen ohjaamisessa. ”Asiaan perehtyneille on selvää, mitkä ovat bibliometriikan ja altmetriikan rajoitukset, mutta kaikille, esimerkiksi yliopiston johdolle, se ei välttämättä ole ihan niin selkeätä”, tutkimuksen kehittämispalveluiden asiantuntija Laura Himanen Tampereen teknillisen yliopistosta (TTY) toteaa.

Mittareiden väärinkäyttö voi saada myös vaikiintuneita muotoja. Tunnetuin esimerkki lienee edellä mainitun vaikuttavuuskertoimen hyödyntäminen tutkimuksen arvioinnissa: alun perin julkaisujen vaikuttavuuden arviointiin kehitettyä indikaattoria käytetään käyttötarkoituksensa vastaisesti tutkimuksen laadun arvioimiseen. Vaikuttavuuskerroin on myös altis manipuloinnille, mikä vähentää sen luotettavuutta.

Suomessa keskustelua on käyty TSV:n yhteydessä toimivan Julkaisufoorumi-luokituksen

(JUFO) käytöstä. JUFO on kehitetty tieteellisten julkaisujen laadun indikaattoriksi yliopistojen rahoitusmallissa, julkaisutoiminnan vertailussa ja julkaisupaikan valinnassa. Suunnittelupäällikkö Janne Pölonen TSV:stä kuitenkin toteaa, että JUFO:n käyttö ei aina vastaa sen käyttötarkoitusta:

Tarkoitus on ollut lisätä tietoisuutta julkaisukanavien vaikuttavuudesta ja arvostuksesta. Luokitusta ei ole tarkoitettu kuitenkaan yksittäisten tutkijoiden arviointiin. Ja tässä on selvästi hieman sekaannusta käyttötarkoituksesta, koska viesti, mikä tiedeyhteisöltä ja myös arviointipaneelilta tulee, on, että luokitusta käytetään joskus myös yksilötasolla.

Yliopistorankingien kohdalla nostetaan usein esiin niiden taustalla olevat puutteelliset tai kokonaan pimentoon jäävät indikaattorit. Läpinäkyvyyden puute, ”mustat laatikot” ja salaiset algoritmit syövät indikaattorien uskottavuutta, mutta eivät välttämättä vaikuta niiden vastaanottoon. Esimerkiksi juuri rankingeja hyödynnetään mielellään markkinoinnissa, jos luvut ovat suotuisia, oli lukujen takana mitä tahansa.

Perusongelmana nykyisessä arviointikulttuurissa pidetään sitä, että määrällinen metriikkadata on saanut liian hallitsevan roolin suhteessa laadulliseen vertaisarviointiin. Kvantitatiivisen datan ylivalta voi olla yhteydessä resurssitekijöihin, esimerkiksi aikaan ja numeroiden tuottamisen helpouteen ja halpuuteen, mutta lukujen avulla voidaan tavoitella myös objektiivisuutta. Tällöin ei kuitenkaan aina tiedosteta lukujen tuottamiseen liittyviä valintoja. Käytetyt indikaattorit voivat vinouttaa tutkimuksen tavoitteiden määrittelyä, ja välineestä voi tulla itsetarkoitus. Erityisen ongelmallista vinoutuminen on silloin, kun käytetään vain harvoja indikaattoreita.

Ratkaisuna paremmat käytännöt

Edellä kuvatut ongelmat tunnetaan varsin hyvin bibliometriikan asiantuntijoiden keskuudessa. On ratkaisevaa, että metriikkaa päätöstensä tukena hyödyntävät tiedostaisivat ongelmat ja osaisivat toimia vastuullisella tavalla. Tätä varten on kehitetty erilaisia vastuullisen metriikan julkilausumia, jotka kokoavat yhteen hyviä käytäntöjä.

Julkilausumista tunnetuin on Leidenin manifesti (2014), joka on nimetty syntypaikkansa mukaan. Manifesti on käännetty 19 kielelle, suomeksi vuonna 2016. Bibliometriikan asiantuntijoiden

laatima manifestin taustalla oli erityisesti huoli kvantitatiivisen arvioinnin ylivallassa suhteessa laadulliseen arviointiin. Kymmenen periaatteen lista kokoaa yhteen tutkimuksen arvioinnin hyvät käytännöt:

1. Määrällisen arvioinnin pitäisi tukea laadullista asiantuntija-arviointia.
2. Tarkastele suoritusta suhteessa instituution, ryhmän tai tutkijan omaan tutkimusmissioon.
3. Ota huomioon, että paikallisesti merkittävä tutkimus voi olla huippututkimusta.
4. Huolehdi, että aineiston kerääminen ja analyysiprosessit ovat avoimia, läpinäkyviä ja yksinkertaisia.
5. Anna arvioitaville mahdollisuus tarkistaa data ja analyysit.
6. Ota huomioon tieteenalojen erot julkaisemisessa ja viittauskäytännöissä.
7. Perusta yksittäisten tutkijoiden arviointi heidän portfolionsa laadulliseen tarkasteluun.
8. Vältä väärin kohdistettua konkretiaa ja epäluotettavaa tarkkuutta.
9. Myönnä järjestelmään kuuluvat arvioinnin ja mittarien seuraukset.
10. Tutki ja päivitä mittareita säännöllisesti.

Manifestin perimmäisenä tavoitteena on, että tutkimuksen arviointi palvelisi sekä tieteellistä tutkimusta että tutkimuksen vuorovaikutusta ympäröivän yhteiskunnan kanssa. Vaasan tiedekirjasto Tritonian johtaja Anne Lehto kiinnittääkin huomiota reflektiivisyyteen mittareiden käytössä: ”Pitää myöntää, että arviointi vaikuttaa toimintaan. Sen olemme huomanneet Suomessakin, yliopistojen JUFO-keskustelussa.”

Leidenin manifestin ohella toinen keskeinen julkilausuma on DORA (*San Francisco Declaration on Research Assessment*). Yhdysvaltain solubiologien yhdistyksen alulle panema julistuksen on allekirjoittanut yli 500 organisaatiota ja yli 12 500 henkilöä. Suomessa DORA-julistukseen sitoutuminen on kirjattu Suomen yliopistojen rehtorineuvoston (UNIFI) keväällä julkaisemaan Avoin tiede ja data -toimenpideohjelmaan yhtenä kahdeksasta toimenpiteestä. DORA-suomennos ilmestyy *Informaatiotutkimus*-lehdessä 4/2018.

DORA pyrkii vaikuttamaan siihen, miten arvioidaan ja mitä arvioidaan. Lähtökohtana ovat olleet vertaisarvioidut tutkimusartikkelit, mutta

julkilausumassa tutkimustuotoksen käsite ymmärretään laajasti (mm. tutkimusdata, ohjelmistot). Muodosta riippumatta olennaista on arvioida tieteellistä laatua ja vaikuttavuutta tuotoksesta itsestään käsin (*on its own merits*). Riippumattomuus julkaisupaikasta taas on nostettu DORAn yleisperiaatteeksi ennen vastaanottajakohtaisia suosituksia: ”Älä käytä lehtiperustaista metriikkaa, kuten lehtien vaikuttavuuskertoimia, yksittäisen tutkimusartikkelin laadun mittaamisen korvikkeena arvioidaksesi yksittäisen tutkijan tieteellisiä ansioita, tai rekrytointi-, ylennys- ja rahoituspäätöksissä.”

Kolmas merkittävä vastuullista metriikkaa linjaava paperi on *The Metric Tide* -raportti (2015). Isossa-Britanniassa toteutettuun laajaan selvitykseen osallistui metriikan, tutkimusrahoituksen ja -politiikan, julkaisemisen ja yliopistohallinnon edustajia. Selvitys nostaa vastuullisen metriikan pääperiaatteiksi kerätyn datan kattavuuden (*robustness*), menetelmien rajoitukset ymmärtävän nöyryyden (*humility*), läpinäkyvyyden (*transparency*), tieteenalojen julkaisukulttuurien erot huomioidun monimuotoisuuden (*diversity*) ja refleksiivisyyden (*reflexivity*).

”Ehkä vertaisarvioinnista voisi palkita”

Yhteiskuntatieteiden tutkija Niilo Kauppi tuntee tutkimuksen arvioinnin paitsi arvioijana ja arviotavana myös tutkimuskohteena. Kauppi on vastikään kirjoittanut yliopistorankingien vaikutuksesta tutkimukseen artikkelin, joka julkaistaan *Studies in Higher Education* -lehdessä.

Yliopistorankingit ovat hyvä esimerkki siitä, miten valta – päätösten ja rahavirtojen muodossa – kytkeytyy ja kätkeytyy metriikkaan. Kaupin mukaan perusasioihin kuuluu tietää, mitä tarkoitusta varten mittarit ovat olemassa. ”Ongelma on se, että indikaattorit otetaan usein annettuna. Se, miten ne on rakennettu ja mihin data perustuu, unohtuu. Onko Shanghain indikaattori hyvä kuvaamaan yliopiston tasoa vai ei? Tätä ei kauheasti pohdita. Ja tätä kautta tullaan vastuulliseen metriikkaan. Minusta sitä ei paljon ole tällä hetkellä”, Kauppi sanoo.

Kauppi toimii tutkimusjohtajana CNRS-tutkimusorganisaatiossa Ranskassa mutta työskentelee parhaillaan valtio-opin vieraillevana FiDiPro-professorina Jyväskylän yliopistossa. Hän on huo-

mannut, miten tutkimuksen arviointi yliopistoissa on muuttunut koko ajan mittarivetoisemmaksi. ”Väittelin vuonna 1991, ja tänä aikana on tapahtunut suuria muutoksia siinä, miten arviointia tehdään. Tendenssi on sellainen, että arviointi on entistä muodollisempaa. Ihmiset eivät välttämättä lue tekstejä, kun he arvioivat. Riittää, kun he näkevät kustantajan tai lehden. Tällaista pika-arviointia on entistä enemmän.”

Kaupin mukaan pika-arviointiin liittyy paitsi laadullisia myös muita ongelmia: ”Arviointi ei tuo esiin mitään uutta. Se, mitä tieteen pitäisi olla, uuden keksimistä tai yllättävän yhteyden esiintuomista – tällaisen merkitys minimoituu muodollisessa arvioinnissa. Jos ajattelee professorihakua, vertaisarviointi olisi loistava tilaisuus laittaa omat tiedot ajan tasalle. Jos sinulla ei ole aikaa lukea, menetät itse siinä. Asiaa pitäisi miettiä myös tältä kantilta.”

Kaupin mukaan laadullisen vertaisarvioinnin roolia pitäisi vahvistaa: ”Se on välttämätöntä. Mutta ihmiset ovat hyvin kiireisiä. Ehkä vertaisarvioinnista voisi palkita, jos se nähdään tärkeänä työnä eikä vain tällaisena sivuhommana. Tiedän, että kollegoillani ei ole aikaa. Ja jos heillä on aikaa, he käyttävät sen omaan tutkimukseen, eivät vertaisarviointiin.”

Leidenin manifestiin ja vastaaviin julkilausumiin Kauppi suhtautuu periaatteessa myönteisesti. ”Niiden hyvyys riippuu kuitenkin lopulta siitä, noudatetaanko niitä vai ei. Se on se ongelma. Vaikka yliopisto päättäisi, että nämä ovat hyviä julistuksia, noudatetaanko niitä jossain laitoksilla? Kyse on aina inhimillisestä toiminnasta. Kaikilla yksittäisillä tutkijoilla ja opettajilla on vastuu oman työnsä kautta, kun he toimivat arviointsijoina. Ja siitä ylöspäin opetusministeriöön.”

Vastuullinen metriikka käytännössä?

Julkilausumien kokoamien hyvien käytäntöjen toteuttaminen käytännössä on, kuten Kauppi toteasi, ydinkysymys. Kirjastoilla on perinteisesti keskeinen asema bibliometriikassa, ja niillä on tärkeä rooli myös vastuullisen metriikan viemisessä käytäntöön. Lehto mainitsee, miten Kööpenhaminan yliopistossa kirjasto hyödyntää simppeleitä tuoteselostetyypistä tapaa informoidessaan asiakkaita Leidenin manifestin periaatteiden toteutumises-

ta: ”Kun bibliometrisen analyysin tulokset toimitetaan tilaajalle, esitetään taulukko, jossa on lueteltu manifestin periaatteet ja kuvattu hymynaamoilla, miten ne ovat toteutuneet prosessissa.”

Euroopan tieteellisten kirjastojen yhdistys LIBER on perustanut metriikkatyöryhmän, jonka tehtävänä on laatia suositukset yliopistokirjastoille metriikan käyttöön. Suosituksissa pyritään huomioimaan kirjastojen erilaisuus, mikä on Lehdon mukaan tärkeää: ”Työryhmä on todennut, että tieteellisten kirjastojen on tärkeätä tuntea periaatteet jo osaamisen kehittämisen kannalta, mutta niitä ei välttämättä pystytä hyödyntämään ihan resurssisyistä. Eivätkä läheskään kaikki periaatteet ole kirjaston käsissä.”

Vaikka kirjasto on vastuullisessa metriikassa luonteva viestinvälittäjä, kulttuurillinen muutos vaatii koko tiedeyhteisön tukea. ”Olen vakuuttunut siitä, että yliopistojen kirjastoissa tehdään tällä hetkellä metriikkaa vastuullisesti. Ongelma on enemmän metriikan hyödyntämisessä”, Himanen toteaa. Eikä kyse ole pelkästään asenteista. Esimerkiksi Leidenin manifestin toimeenpano vaatii resursointia: se vaatii aikaa, rahaa ja metriikkaan liittyvää osaamista. Toistaiseksi Leidenin manifesti on otettu käyttöön vain Loughboroughin yliopistossa, jossa se on osa *building excellence* -strategiaa. Lehdon mukaan muutama muukin yliopisto on omaksunut manifestin. Vaikka Suomessa yliopistot eivät ole muodollisesti sitoutuneet manifestiin, ainakin Helsingin yliopisto on käytännössä toteuttanut Leidenin manifestia meneillään olevassa tutkimuksen arvioinnissa.

Miten suomalaista tutkimuskulttuuria pitäisi kehittää vastuullisempaan suuntaan? Himanen nostaa esiin kolme kohtaa. Ensinnäkin, yliopistojen johdon on sitouduttava vastuullisen metriikan käyttöön. Toiseksi, on tunnistettava kohdat, joissa metriikkaa käytetään. Ja tässä on huomattava, että vastuullisesti tuotettu metriikka ei varmista vastuullisesti käytettyä metriikkaa. Kolmantena kohtana Himanen painottaa koulutuksen merkitystä. Ison-Britannian laajan *The Metric Tide* -selvityksen kaltainen raportti voitaisiin Himasen mukaan toteuttaa myös Suomessa. TSV onkin jo tarttunut tehtävään: se asetti syyskuussa työryhmän, jonka tehtävänä on laatia tutkijan vastuullista arviointia koskeva kansallinen suositus keväällä 2019.

Lähteet

- Coombs, Sarah K. ja Isabella Peters (2017). The Leiden Manifesto under review: what libraries can learn from it. *Digital Library Perspectives* 33 (4): 324–338. DOI: 10.1108/DLP-01-2017-0004.
- DORA Community (2012). San Francisco Declaration on Research Assessment (DORA). URL: <<https://sfedora.org/read/>> (luettu 4.9.2018).
- DORA Community (2018). DORA Roadmap: A two-year strategic plan for advancing global research assessment reform at the institutional, national, and funder level. URL: <<https://sfedora.org/2018/06/27/dora-roadmap-a-two-year-strategic-plan-for-advancing-global-research-assessment-reform-at-the-institutional-national-and-funder-level/>> [4.9.2018]. DORA-julistuksen suomennos julkaistaan *Informaatiotutkimus*-lehdessä, vol 37, nro 4 (2018), ja suomennos löytyy myös DORA-verkkosivuilta: <https://sfedora.org/read/fi/> (luettu 23.10.2018).
- Hicks, Diana, Paul Wouters, Ludo Waltman, Sarah de Rijcke ja Ismael Rafols (2015). The Leiden manifesto for research metrics. *Nature* 520 (7548): 429–431. DOI: 10.1038/520429a. Suomenkielinen käännös Hicksin ym. artikkelista löytyy Suomen yliopistokirjastojen neuvoston verkkosivuilta: <http://yliopistokirjastot.fi/leidenin-manifesti/> (luettu 1.9.2018).
- Julkaisufoorumi (2018). Julkaisufoorumi-luokituksen käyttöohje. Tuorein versio julkaistu 9.2.2018, http://www.julkaisufoorumi.fi/sites/julkaisufoorumi.fi/files/jufo_manual_2018vi.pdf
- Suomen yliopistojen rehtorineuvosto UNIFI ry (2018). *Avoin tiede ja data. Toimennpideohjelma suomalaiselle tiedeyhteisölle.* <http://urn.fi/URN:NBN:fi:fe2018052424593>
- Tieteellisten Seurain Valtuuskunta (2018). Työryhmä ohjeistamaan tutkijan vastuullista arviointia. Uutinen TSV:n verkkosivuilta: <https://www.tsv.fi/fi/uutiset/ty%C3%B6ryhm%C3%A4-ohjeistamaan-tutkijan-vastuullista-arviointia> (luettu 1.10.2018).
- Wilsdon, J. ym. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management.* DOI: 10.13140/RG.2.1.4929.1363.

Kirjoittaja on tietoasiiantuntija Helsingin yliopiston kirjastossa.

SKS:SSA TAPAHTUU

20.11. Lasten päivän seminaari: Lapsi sodassa ennen ja nyt

22.11. Kriisipuheesta tekoihin – kirjalala nyt!

26.11. Romanian kulttuuriperintö: arkiarvostus ja tutkimus -tutkimushankkeen päätösseminaari

SKS on mukana Tieteiden yössä 10.1.2019!

Tarkemmat tiedot ja kaikki SKS:n tapahtumat: www.finlit.fi