

Tieteen laatu ja avoimuus

Näkökulmat Julkaistu 04.11.2016 klo 15:48 Kirjoittaja Pekka Nygren

Keskustelu tieteen avoimuudesta on voimakkaasti keskittynyt siihen, kuinka julkisesti rahoitetun tieteen tulee olla kaikkien luettavissa. Usein käsitellään vain tieteellisten artikkeleiden vapaata käytettävyyttä, joka Suomessakin tunnetaan ehkä paremmin englanninkielisenä käsitteenä open access. Tieteen avoimuudella on kuitenkin erittäin tärkeä tehtävä myös tutkimuksen laadun takaajana ja keskitynkin tässä esseessä ennen kaikkea siihen, kuinka avoimuus voi edistää tutkimuksen laadun tarkkailua ja kehittämistä.

Kaikille tieteellisessä jatkokoulutuksessa oleville opetetaan, kuinka ihannetilanteessa tutkimus raportoidaan niin, että toinen asiantunteva tutkija pystyy toistamaan sen. Kuinka hyvin tämä ihannetilanne sitten toteutuu käytännössä? Biolääketieteellinen yhteisö on joutunut pohtimaan asiaa, kun tuli ilmi, että lääkeyhtiö [Amgen pystyi toistamaan vain kuusi 53 syöpätutkimuksesta ja Bayer onnistui toistamaan 14 testaamastaan 67 biolääketieteellisestä tutkimuksesta](#). Toisella tieteenalalla, [psykologiassa, on päästy 39 % toistettavuuteen](#). Tilanteeseen ei liene yhtä ainoaa syytä, mutta huolimaton tapa tehdä ja raportoida tutkimusta (sloppy science) ja avoimuuden puute ovat kaksi merkittävää tekijää.

Menetelmien avoimuudesta

Kun numeroiden murskaus nykyisillä työkaluilla on helppoa, tutkija voi analysoida dataansa monilla erilaisilla tilastollisilla testeillä. Mahdollisuus tarkastella dataa eri tavoin on sinänsä erinomainen asia, mutta kiusaus raportoida vain ennako-odotuksia tukevat tulokset on suuri. Numeroiden murskauksen helppous johtaa myös helposti sellaisten testien käyttöön, jotka eivät sovikaan käsillä olevaan dataan esim. jakaumaoletusten takia. Näin saadut tulokset eivät tietenkään ole päteviä, vaikka ne tukisivatkin tutkijan ennako-odotuksia. Ensimmäinen tutkimuksen laadunvarmistusta tukeva avoimuusvaatimus onkin menetelmien avoimuus.

Yksinkertaisimmillaan kyse on siitä, että [seurataan tilastotieteen peruskurssin opetuksia](#): kirjoitetaan auki miten otoskoko on arvioitu, miten koe on satunnaistettu, tiesikö laboratorio- tai kenttähenkilökunta mihin käsittelyyn kukin tutkimusyksikkö kuului ja miten data on saatettu testattavaan muotoon. Datan käsittelyssä erittäin tärkeää on kirjoittaa onko joitakin datapisteitä poistettu ja miksi. Menetelmien avoimuus edellyttää kuitenkin myös sitä, että [tehdyt analyysit ja käytetyt tilasto-ohjelmien koodit tallennetaan tunnettuun arkistoon](#), jossa ne ovat muiden tutkijoiden saatavilla. Tutkimusartikkelissa on oltava selkeät viitteet menetelmäarkistoon ja arkistoon tallennetussa aineistossa taas selkeät viitteet artikkeliin, missä niitä on käytetty.



Esseen kirjoittajan Pekka Nygrenin (Suomen Metsätieteellinen Seura) oma tutkimusala on metsäekologia ja hän on tutkinut symbioottisen typensidonnan merkitystä ekosysteemien toiminnassa. Kuvassa oleva tervaleppä muodostaa tyypeä sitovan symbioosin aktinobakteereiden kanssa. Kuva: Tommi Kutilainen CC BY-NC-SA

Avoim data tutkimuksen laadunvarmistuksessa

Datan avoimuus on nousemassa myös suomalaiseen keskusteluun. Senkin perusteluna esitetään uudelleenkäyttämömahdollisuus muissa tutkimuksissa, mutta yhtä tärkeää datan avoimuus on tutkimuksen laadunvarmistuksen kannalta.

Omalla tutkimusalallani, metsäekologiassa, suurin osa tutkimuksista on sellaisia, joita ei voida toistaa suoraan eli toistaa koejärjestelyä ja aineiston keruuta täsmälleen samanlaisissa oloissa kuin alkuperäinen tutkimus on tehty.

Esimerkiksi sopii [tutkimus metsärajan muutoksista Pohjois-Suomessa 25 vuoden aikana](#). Suora toisto vaatisi uuden 25 vuoden tutkimusjakson, mutta sen aikana ei ole täsmälleen samanlaisia ilmasto-oloja kuin alkuperäisen tutkimuksen aikana ja metsän kehittyessä sen sisäiset ympäristöolot ja niiden myötä puupopulaation muodostavien puiden keskinäiset vuorovaikutukset ovat muuttuneet; voipa käydä niinkin, että jonkin kaarnakuoriaisen tai taudinaiheuttajasienen massaesiintyminen heikentää jotakin tutkimusmetsikköä merkittävästi. Konseptuaalinen toistaminen eli saman koejärjestelyn tutkiminen joissakin muissa kuin alkuperäisen tutkimuksen metsiköissä on periaatteessa mahdollista, mutta sekin vaatisi tuon uuden 25 vuoden tutkimusjakson.

Avoimuus on ainoa tapa varmistaa tieteen laatu ja itsensäkorjaavuus

Ainoastaan avoin aineisto antaa mahdollisuuden arvioida alkuperäisen työn laatua. Toiset tutkijat voivat toistaa tehdyt analyysit ja mahdollisesti analyysimenetelmien kehittyessä tehdä uusia analyysseja tai he voivat liittää datan esim. laajempaan mallinnustutkimukseen.

Asteikko avoimuuden edistämiseksi

Yhdysvaltalainen Center for Open Science, joka aloitti Virginian yliopiston osana, mutta itsenäistyi siitä voittoavoittelemattomaksi yritykseksi, on laatinut [kolmiportaisen asteikon avoimuuden edistämiseksi](#) tieteellisissä julkaisuissa.

Avoimuuskohteita on kahdeksan: dataan viittaamisen käytännöt, datan avoimuus, analyysiohjelmien ja -koodin avoimuus, tutkimusaineiston avoimuus, koesuunnittelun ja analyysin avoimuus, tutkimussuunnitelmien etukäteisrekisteröinti, analyysisuunnitelmien etukäteisrekisteröinti ja tutkimusten toistaminen. Näistä tutkimus- ja analyysisuunnitelmien etukäteisrekisteröinti ovat jonkin verran käytössä kokeellisessa psykologiassa ja biolääketieteissä, mutta ne eivät soveltune kaikille tieteenaloille, etenkin sellaisiin, jotka perustuvat kenttäaineiston keruuseen. Tutkimusten suora toistaminenkin on monilla aloilla mahdotonta. Tieteellisten julkaisusarjojen tulisi kuitenkin olla valmiita julkaisemaan myös aiempien tutkimusten suoria tai konseptuaalisia toistoja, koska vain niiden avulla ensimmäisten läpimurtojen tulokset voidaan varmentaa.

Tutkijoiden ja tiedejulkaisijoiden olisi kuitenkin hyvä nähdä avoimuus osana tieteen perusolemusta, josta on hiljaisesti etäännytty

Center for Open Sciencen ykköstaro vaatii lähinnä sitä, että julkaisusarjan kirjoittajien ohjeissa on selkeät ohjeet avoimuuden soveltamisesta, mutta sarja ei vaadi niiden noudattamista julkaisemisen ehtona. Kakkostarolla hyvät avoimuuskäytänteet ovat ehtona sille, että käsikirjoitus lähetetään vertaisarviointiin. Kolmostarolla mennään jo varsin korkeisiin vaatimuksiin, mm. siihen, että jonkin riippumattoman tutkimusryhmän tulisi toistaa tutkimuksen analyysit. Lyhyesti, ykköstarolla esim. suositellaan että data avataan ja annetaan ohjeet siitä miten avaaminen tapahtuu, kakkostarolla data on tallennettava avoimeen julkiseen tietokantaan ja kolmostarolla data-analyysit on toistettava. Henkilökohtaisesti näen kakkostaroon tavoiteltavana myös pienten julkaisijoiden lehdissä ja se on kohtuullinen rahoittajien vaatimus, mutta kolmostarossa tuntuu vaikealta saavuttaa ja ainakin se lisää merkittävästi tutkijoiden, tiedejulkaisijoiden ja tarkastajien työmäärää, eikä rahoittajakaan ehkä ole tyytyväinen sen aiheuttamaan tutkimusprosessin hidastumiseen.

Julkiset tietokannat

Useaan kertaan olen viitannut julkisiin tietokantoihin. Genetiikassa on jo pitkään ollut tapana tallentaa kaikki geenisekvenssit julkisiin tietokantoihin, joista tunnetuin on yhdysvaltalainen [GenBank](#).

Hyvä esimerkki tällaisen keskitetyn tietokannan arvosta on muutama viikko sitten julkaistu [tutkimus lajirunsauden ja lajien sisäisen geneettisen monimuotoisuuden maantieteellisestä vaihtelusta](#), joka perustuu 138 000 Genbankiin tallennettuun geenisekvenssiin. Siksi julkaisijan oma tietokanta ei ole hyvä ratkaisu aineistojen, datan ja koodien tallentamiseen - tällöinhän tieto jakautuu moniin pieniin arkistoihin ja sen etsiminen vie aikaa. Lisäksi pienen julkaisijan olisi vaikeaa ylläpitää tiukat laatuvaatimukset täyttävää tietokantaa. Itse asiassa on hyvä miettiä onko

suomalainen kansallinenkin arkisto liian pieni tieteen avoimuuden ajamiseen, ainakin tiedonhaun näkökulmasta, vai pitäisikö suomalaisia tutkijoita rohkaista tallentamaan tutkimusdatansa olemassa oleviin kansainvälisiin tietokantoihin.

Suomessa tutkimusaineistojen ja -datan avoimuuden vaatimus on tullut monen tutkijan näkökulmasta ylhäältä, EU:lta, opetus- ja kulttuuriministeriöltä tai Suomen Akatemiasta. Tutkijoiden ja tiedejulkaisijoiden olisi kuitenkin hyvä nähdä avoimuus osana tieteen perusolemusta, josta on hiljaisesti etäännytty. Avoimuus on ainoa tapa varmistaa tieteen laatu ja itsensäkorjaavuus. Sekä tutkijoiden että julkaisijoiden on syytä hylätä kilpailuhenki, vaikka maan nykyinen hallitus sitä aatteellisista syistä korostaa, ja palata alkuperäiseen avoimen tieteen ideaan.