

Miksi tutkijan kannattaa jakaa dataa?

Näkökulmat Julkaistu 11.03.2015 klo 13:15 Kirjoittaja Ari Asmi

Avointa tiedettä perustellaan monilla tavoin, usein maksajan ja yhteiskunnan näkökulmasta: Tieteen taso, vaikuttavuus ja yhteiskunnallinen hyöty paranevat tietoaisteiden jaon myötä. Rahoittajan tulee saada kaikki siitä, mistä maksaa. Samoin tuodaan esille tieteenfilosofisia näkökulmia tutkimusten toistettavuudesta, tarkistettavuudesta, uudelleenkäytöstä ja laadun varmistamisesta.

Yhteiskunnallinen vastuu ja tieteenfilosofia ovat kuitenkin usein aika kaukana tutkijan päivätyöstä. Vaikka jollain korkealentoisella tasolla tutkija voi pitää näitä asioita tärkeinä, aika on tiukalla – pitäisi tehdä tutkimusta, opettaa, kirjoittaa julkaisuja ja hakemuksia. Muiden ongelmat ja potentiaalinen tutkimusdatan jatkokäyttö eivät pysty suoraan kilpailemaan tutkijan ajasta, ja tutkijan urakehitystä ohjailevat H-indeksit, julkaisuluettelon pituus ja hankitut tutkimusapurahat. Ainoastaan hakemusten *impact*-osuutta kirjoittaessa saattaa avoin tiede ehtiä ajatuksiin kadotakseen sieltä taas hankkeen työn varsinaisesti alkaessa.



Mitä iloa datan jakamisesta sitten on tutkijalle? Miksi sitä kannattaa tehdä?

Aineiston kuvaaminen on aikaavievää ja motivaatio voi olla äkkiä hukassa, jos ainoana syynä on ulkoinen paine. Ensimmäinen ja tärkein aineiston käyttäjä on kuitenkin usein sen alkuperäinen kerääjä ja senkin vuoksi aineisto on syytä kuvailla huolella, jotta voisi itse palata siihen myöhemmin. Aineiston kuvailu onkin ennen kaikkea oikean työtavan oppimista, ja sen hyödyntäjä on tutkija itse.

Itse opin kuvailun tärkeyden – niinkuin monet muutkin tutkijat – kantapään kautta. Parissa kuukaudessa niin selkeältä tuntunut koodinpätkä oli muuttunut täysin käsittämättömäksi. Samalla tavalla on käynyt jo monelle vanhalle aineistolleni ennen kuin aloin ymmärtää oman etuni – joka on hyvin lähellä kaikkien muidenkin etua.

Data jakaminen on myös erittäin hyvä tapa edistää omaa uraa, ainakin pitkällä tähtäimellä. Tämä voi alkuun kuulostaa oudolta: eihän dataan tunnetusti oikein voi viitata tällä hetkellä, eikä kukaan vielä tiedä, miten se huomioonotetaan virkojen täytössä.

H-indeksit ja julkaisumäärät eivät kuitenkaan ole mikään itsetarkoitus. Ne ovat vain heikkoja tapoja yrittää mallintaa olennaista tutkijan kokemusta ja tutkijan työn merkittävyyttä, joita yleensä kuitenkin kuvataan tutkijan maineella. Miksi tutkijan kannattaa pitää esitelmiä konferensseissa? Tai suostua toimittajan haastatteluun? Näihin ei yleensä ole mitään ongelmaa löytää motivaatiota: Tietäähän jokainen, että tunnetulla tutkijalla on paremmat mahdollisuudet löytää uusia yhteistyökumppaneita, saada tutkimuksilleen viitteitä ja loppujen lopuksi rahoitusta ja sitä kautta uutta kiinnostavaa tietoa.

Vaikka tutkimusdatalla ei vielä ole suoraa näkyvyyttä perinteisissä tutkijoiden arviointimenetelmissä, ihmettelen suuresti, jos sama pätee vielä kymmenen vuoden kuluttua. Nyt julkaisemalla datasi olet muita kymmenen vuotta edellä – asia, josta ei voi olla kuin hyötyä pitkän aikavälin menestyksellesi.